11TH ANNUAL EUROMEDIA CONFERENCE
2005

TOULOUSE, FRANCE

APRIL 11-13, 2005

Organized and Sponsored by:

EUROSIS

Co-Sponsored by:

ASPASIA KONWLEDGE SYSTEMS

EU-DG INFSO

BELGACOM

DE MONTFORT UNIVERSITY

GHENT UNIVERSITY

HOSTED BY

IRIT

UNIVERSITE PAUL SABATIER

EUROMEDIA'2005

FEATURING

ELEVENTH ANNUAL SCIENTIFIC CONFERENCE
ON WEB TECHNOLOGY, NEW MEDIA
COMMUNICATIONS AND TELEMATICS THEORY
METHODS, TOOLS AND APPLICATIONS

Marwan Al-Akaidi

and

Leon Rothkrantz

APRIL 11-13, 2005
TOULOUSE, FRANCE

A Publication of EUROSIS

Printed in Ghent, Belgium

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

Editor

Prof. Marwan Al-Akaidi
De Montfort University
Leicester
United Kingdom

Dr. Leon Rothkrantz
TU Delft
Delft
The Netherlands

Programme Committee

Prof. Marwan Al-Akaidi
De Montfort University
Leicester, United Kingdom

Fernando Boronat Seguif
UPV
Valencia, Spain

Dr. Uwe M. Borghoff
University of the Federal Armed Forces
Neubiberg, Germany

Prof. Dr.J.Broeckhove
RUCA-UA
Antwerp, Belgium

Dr. Juan Carlos Guerri Cebollada
UPV
Valencia, Spain

Dr. Paul Dowland
University of Plymouth
Plymouth, United Kingdom

Victor Dozortsev
JV Petrocom
Moscow, Russia

Dr. Steven Furnell
University of Plymouth
Plymouth, United Kingdom

Prof. Chris Guy
University of Reading
Reading, United Kingdom

Prof. Dr. Jan Knop
University of Düsseldorf
Düsseldorf, Germany

Ass. Prof. Qingping Lin
Nanyang Technological
University
Singapore

Rachel Moreau
LUC
Diepenbeek, Belgium

Lorenzo Motta
Ansaldo Trasporti s.p.a.
Genoa, Italy

Dr. H. Joachim Nern
Aspasia Knowlegde Systems
Dusseldorf, Germany

Dr.ir.Johan Opsommer
Belgacom
Brussels, Belgium

Dr. Carlos Enrique Palau Salvador
UPV
Valencia, Spain

Dr. Ana Pajares
UPV
Valencia, Spain

Prof. Jehan Francois Paris
University of Houston
Houston, USA

Prof. Marco Roccetti
University of Bologna
Bologna, Italy

Dr. Leon Rothkrantz
Delft University of Technology
Delft, The Netherlands

Prof. Paola Salomoni
University of Bologna
Bologna, Italy

Prof. Jeanne Schreurs
LUC
Diepenbeek, Belgium

Ph. D. Oryal Tanir
Bell Canada
Montreal QC, Canada

Ass. Prof. Vassilis Triantafillou
Techn. Education Institute
Greece

Prof. Rik van de Walle
Ghent University
Ghent, Belgium

Charles van der Mast
Delft University of Technology
Delft, The Netherlands

Dr. Matthew Warren
Deakin University Geelong
Victoria, Australia

© 2005 EUROSIS-ETI

Responsibility for the accuracy of all statements in each paper rests solely with the author(s). Statements are not necessarily representative of nor endorsed by the European Simulation Society. Permission is granted to photocopy portions of the publication for personal use and for the use of students providing credit is given to the conference and publication. Permission does not extend to other types of reproduction nor to copying for incorporation into commercial advertising nor for any other profit-making purpose. Other publications are encouraged to include 300- to 500-word abstracts or excerpts from any paper contained in this book, provided credits are given to the author and the conference.

All author contact information provided in this Proceedings falls under the European Privacy Law and may not be used in any form, written or electronic, without the explicit written permission of the author and/or the publisher.

For permission to publish a complete paper write EUROSIS-ETI, Executive Director, Ghent University, Faculty of Engineering, Dept. of Industrial Management, Technologiepark 903, Campus Ardoyen, B-9052 Ghent-Zwijnaarde, Belgium.

EUROSIS is a Division of ETI Bvba, The European Technology Institute

EUROSIS Publication

ISBN: 90-77381-17-1

EUROMEDIA´2005
PROCEEDINGS

PREFACE

As digital information is becoming more and more pervasive, in every aspect of our daily lives, and increases exponentially day by day, flooding us with terabits of text data, images, audio files etc.., efforts to manage this chaotic onslaught of information are reflected in this year's EUROMEDIA conference topics. We refer to the special workshops on semantic web applications and distributed multimedia databases, which try to address the problem.. Of course, the other historical topics of the event are not neglected and feature at the beginning of the Proceedings. One aspect, which immediately jumps out is the further convergence between the different areas of multimedia research and more so then in previous years: a trend which, in our opinion, will continue unabated in the future.

As with any conference, EUROMEDIA would not be possible without the help and support of a number of people, and we would like to begin by thanking all of the reviewers for their efforts, which have resulted in a truly interesting and varied conference programme. we are also most grateful to Dr. Hans Joachim Nern, Prof. Vincent Charvillat, Prof. Romulus Grigoras and Prof. Harald Kosch for expanding this year's event with two very topical and specific workshops, and of course to all of the paper authors and presenters for their hard work, and their willingness to share their results in the other sessions. Thanks also to the session chairs and other delegates who we are sure will guarantee us a lively and thought-provoking conference. Finally, special thanks are due to Philippe Geril, whose continued dedication and hard work as the conference organiser has enabled us to maintain the standard expected of EUROMEDIA events.

We sincerely hope that all of the delegates enjoy the conference, and that other readers of these proceedings will be encouraged to participate in EUROMEDIA events in the future. On behalf of all of EUROSIS, the International Programme Committee, and the Universite Paul Sabatier, we welcome you to this event and look forward to a successful conference.

Prof. Marwan Al-Akaidi
General Chair

Dr. Leon Rothkrantz
Program Chair

CONTENTS

Preface ..... **IX**
Scientific Programme **1**
Author Listing **263**

WEBTEC

Distributed Digital Library Architecture for Business Solutions
Marcin Helinski, Cezary Mazurek and Marcin Werla **5**

Design of a Scalable Collaborative Virtual Environment System
Lin Qingping, Zhang Liang and Irma Kusuma **10**

COMTEC

COMMUNICATION NETWORKS

Load balancing in IP networks in normal operation and with failure resilience
Gerhard Hasslinger, Stefan Schnitter and Martin Franzke **23**

CR-LDP Usage for setting up a LSP with bandwidth guaranteed in a SIP over MPLS network
C.Zhang and C.G.Guy **31**

MOBILE COMMUNICATIONS

Specification and Design of an advanced Hot Spot Controller Based on Radius Authentication Server
Panagiotis S. Alefragis, Nikolaos S. Voros, Vassilis D. Triantafillou and Konstantinos E. Tsatsaronis **37**

A News Flash System for Mobile Devices Using J2ME
Khaled Belhoul, Mohammed Jamal Zemerly and Arif Al Hammadi **42**

Location Information System: a description of a positioning middleware system
Kimmo Koskinen, Radek Spacil and Jouni Ikonen **47**

COMMUNICATIONS SECURITY

On the Security of non-RSA EMV Payment Cards
Marten Trolin **57**

CONTENTS

Securing Data by Steganography on Small Devices
Alessandro Amoroso, Marco Masotti and Michele Pio Di Pasquale.....**62**

Intrusion Detection via Behavioural Profiling on Mobile and Wireless Networked Devices
Ibrahim Zincir, Stephen M. Furnell and Andrew D. Phippen **67**

MEDIATEC

MULTIMEDIA COMMUNICATION TOOLS

Creation and Application of Metadata for Content Production in Broadcasting Stations-- A Case Study of Metadata using high school baseball games in Japan
Tomohisa Akafuji, Kaname Harumoto, Mitsuru Yasukata, Yozo Yamamoto, Masaaki Kobayashi, Keishi Kandori and Shinji Shimojo **77**

An Icon-Based Communication Tool on a PDA
S. Fitrianie and L.J.M. Rothkrantz.....**83**

VIDEO BASED SYSTEMS

Complexity analysis of Interpolation filters for Scalable Video Coding
Koen De Wolf, Yves Dhondt, Jan De Cock and Rik Van de Walle **93**

Providing Accessible and Portable Video Lecture from Content Transcoding
Silvia Mirri and Paola Salomoni **97**

RECOGNITION SYSTEMS

Machine Learning Approaches for Face Analysis
D.Datcu and L.J.M.Rothkrantz.....**105**

A Wearable Gesture Recognition System for Natural Navigation Interfaces
Elisabetta Farella, Luca Benini, Bruno Ricco and Andrea Acquaviva **110**

APTEC

A Networked Telepresence System for cars and its Transmission Issue
Kazutoshi Fujikawa, Kazumasa Yamazawa, Tomoya Ishikawa, Yuichi Butsuen, Tomokazu Sato, Yutaka Nakaramura, Hideki Sunahara and Naokazu Yokoya.....**119**

CONTENTS

Superassist: Personal Assistants for Cooperative Healthcare Treatment
Geert de Haan and Charles A.P.G. van der Mast, Olivier Blanson Henkemans and Mark A. Neerincx **124**

The Effect of Locomotion Technique on Presence, Fear and Usability in a Virtual Environment
M.J. Schuemie, B. Abel, C.A.P.G. van der Mast, M. Krijn and P.M.G. Emmelkamp **129**

Almost Blue: the Design of a Cooperative Game by integrating accessible interaction
Stefano Cacciaguerra, Silvia Mirri, Paola Salomoni and Manuel Baldassarri ... **136**

ETEC

Adaptive RFID Tag Identification for eCommerce Applications
Selwyn Piramuthu **143**

A Knowledge Based Decision Support System for E-Commerce
Stephan Kassel, .Christian-Andreas Schumann, Kay Grebenstein and Claudia Tittmann **146**

Experiences from Building an Information Security e-Learning Environment for Industry
Jorma Kajava and Rauno Varonen **151**

DISTRIBUTED MULTIMEDIA DATABASES AND MULTIMEDIA ADAPTATION

Multimedia Database Systems; Where are we now?
Harald Kosch **159**

Flexible Multimedia Query Rewriting
Georges Chaloub, Samir Saad, Richard Chbeir and Kokou Yetongnon **164**

Experiencing Audio Indexation on Broadcasted News
Messaoud Bengherabi, Abdenour Sehad and Ahmed Mostefaoui **169**

Using a Trace-Based Simulation Tool to Test the Efficiency of Erasure Codes For Wireless Multimedia Streaming
Julien Bourgeois, Adrien Henriet, Jerome Lacan and Tanguy Pérennou **173**

CONTENTS

Distributed Open Cross Media Indexing Platform
Bassem Haidar, Philippe Joly and Jean Paul Bahsoun **178**

Adaptive multimodal communication for multimedia
Cezar Pleşca, Romulus Grigoraş, Philippe Queinnec and Gerard Padiou..... **183**

Removing the MAC Retransmission Times from the RTT in TCP
Eugen Dedu, Sébastien Linck and François Spies **190**

Inside the adaptation decision-taking engines: monitoring adaptabilities by learning
Romulus Grigoraş, Pascal Dayre, Vincent Charvillat and Hadj Batatia..... **194**

Composing Multimedia Content Adaptation Services
Girma Berhe, Lionel Brunie and Jean-Marc Pierson..... **199**

SEMANTIC WEB APPLICATIONS

Web Integration for Enterprise Applications
Hristina Dascalova **205**

Existing Approaches for Modelling and Integration of Business Logics and Processes
Hristina Dascalova, Vladislava Grigorova and Tatiana Atanasova **210**

Infrawebs Semantic Web Unit for Design and Composition of Semantic Web Services
Tatiana Atanasova, Gennady Agre and Joachim Nern..... **216**

Migrating from Web Services to Semantic Web services Infrawebs Approach
Gennady Agre and Tatiana Atanasova and Joachim Nern..... **221**

Case Based Semantic Web Services Designer and Composer
Gennady Agre, Tatiana Atanasova and Joachim Nern..... **226**

Semantic Web Service Execution for WSMO based Choreographies
James Scicluna and Axel Polleres..... **230**

Distributed Decision Support in Multi-Agent Environment for Semantic Web Services
Tatiana Atanasova and Joachim Nern..... **233**

The SUA-Architecture within the Semantic Web Service Discovery and Selection Cycle
Csaba Fülöp, Laszló Kovacs and András Micsik **236**

Semantic Information Routing as a Pre-Process for Semantic Web Services Generation - SIR & OM
T. Westkaemper, J. Saarela and H. Joachim Nern **239**

Fuzzy Concept Sets (FCS) Applied to Semantic Organizational Memories as a Pre-Processing Stage within the Semantic Web Services Designing and Composing Cycle
H. Joachim Nern, J. Tacheva and A. Dziech **242**

Applying Clustering Methods to Distributed Decision Making in Semantic Web Service Designing and Composing Cycles
H.-Joachim Nern, Andrzej Dziech and Tatiana Atanasova **247**

LATE PAPER

PARIS: A MPEG-7 Geo-Referenced Personal Photograph Database
Pei-Jeng Kuo, Terumasa Aoki and Hiroshi Yasuda **255**

SCIENTIFIC PROGRAMME

WEBTEC

DISTRIBUTED DIGITAL LIBRARY ARCHITECTURE FOR BUSINESS SOLUTIONS

Marcin Heliński
Cezary Mazurek
Marcin Werla
Poznan Supercomputing and Networking Center
ul. Noskowskiego 12/14
Poznan, POLAND
{helin,mazurek,mwerla}@man.poznan.pl

KEYWORDS

Digital libraries, Distributed systems, Business applications, Information exchange, Content management

ABSTRACT

Modern companies gather more and more documents associated with their various activities. Fast and efficient access to these documents is often a crucial element, deciding about the success or failure of many enterprises. Digital libraries are becoming a mature and considerable solution to the information chaos dominating in the Internet. We intend to show that DL systems can also be used in business environments, as a basis of corporation document repositories. We believe that features like scalability, distribution and outsourcing possibilities are the key factors required to successful business usage of digital libraries. Based on our experience gained during the development and several deployments of the dLibra Digital Library Framework, we describe the modular digital library structure and present three different distributed DL architectures that can be used to create economic and effective commercial solutions. We also describe a sample digital library based service that can be used to reduce company expenses associated with documents publishing and distribution.

INTRODUCTION

Poznan Supercomputing and Networking Center (PSNC) has been involved in digital library (DL) research for the recent few years. The dLibra digital library system (http://dlibra.psnc.pl/) has been developed in Poznan since 1999 and was presented on several occasions during international conferences (Gruszczyński et al. 2002; Mazurek et al. 2003; Kosiedowski et al. 2004). So far the system has been deployed in two digital libraries: the regional Digital Library of Wielkopolska Region (http://www.wbc.poznan.pl/) in 2002 and academic Wroclaw University of Technology Digital Library (http://dlib.bg.pwr.wroc.pl/) in 2004. New dLibra-based deployments of academic and regional digital libraries in Poland are planned in the near future. During our research we treat digital libraries not only as systems to preserve cultural heritage, but also as tools that can be used to create large repositories of any type of digital resources.

Nowadays effective information storage and exchange is essential in almost all business areas. Many companies store large quantities of documentation about their present or recent activities, projects, customers etc. Sometimes it is because of different legal obligations and sometimes for internal statistical analysis or "lessons learned" type of usage. Gathered information is stored for both internal and external users. Employees need access to internal company resources, while customers are interested in materials like manuals, video tutorials or usage licenses.

When there is a small amount of information that should be stored or shared, a simple workgroup or CMS systems like free, open source eGroupware (http://egroupware.sf.net) or Exponent CMS (http://exponent.sf.net/) can be used. These systems can be used to set up an internal or external company website with some standard, but customized functionality, including basic documents storing, searching and browsing. But when there is a need to store large quantities of diverse digital assets, with additional advanced, content-type dependent search capabilities, these systems are insufficient. There are many commercial systems prepared especially to handle such tasks. One of them is, for example, the IBM DB2 Content Manager (http://www-306.ibm.com/software/data/cm/). This product is created to store, manage and deliver critical business information on demand. Additionally it can cooperate with many external systems like SAP or PeopleSoft solutions. However, it has one major flaw – it is very expensive.

We think that digital library systems can be used to create a functional and economic solution. Such systems are designed to store, manage, browse, search and access large sets of various digital resources. All gathered resources can be described with metadata and grouped in hierarchical collections. Access to those resources is strictly controlled, and the entire content can be browsed or searched. Often there are some additional semantic mechanisms that improve searching possibilities or enable different kinds of knowledge discovery (Falquet et al. 2004). The majority of modern DL systems have WWW based user interface,

which is essential for distributed large scale usage. Example DL systems are the Greenstone Digital Library (http://www.greenstone.org/; Witten et al.2002) and Fedora (http://fedora.info/; Payette and Staples 2002). Both of them are free and can be easily adapted to be the basis of a business digital library. Depending on internal DL system structure and distribution possibilities, the created system has different scalability possibilities (Pyrounakis et al. 2004). Scalability is very important, as it often decides on the cost effectiveness of enterprise solutions based on a given system. Both mentioned DL systems and majority of others have very similar, extended user functionality, but are not designed to support advanced distribution scenarios.

Modern business enterprises require sophisticated possibilities of documents storage, access and exchange. Systems used to satisfy such needs must support various types of documents (text, audio, video etc.), depending on a company profile. Due to a large number of documents, complex searching and browsing functionality is required, to allow employees to quickly find the desired information. Another crucial feature is strict access control and distributed access possibilities. There are companies with branches in different countries, and employees of such company must have access to necessary documents, but this access must be controlled to prevent critical information leaks. A system providing all mentioned functionality must also be very scalable, so that it can be used in different enterprises. A possibility to outsource the most expensive parts of such system is also required from the business point of view.

Below we first describe the dLibra Digital Library Framework architecture and functionality, and then we present a few dLibra business usage scenarios. In these scenarios we want to show how the internal library structure used in dLibra and its distributed module-based architecture both give interesting usage and scalability possibilities. We also show how our system functionality can be used to outsource the most expensive system functionality. There are separate sections for SME usage, and for basic and advanced distributed solutions. In the end we describe a sample digital library-based service that can be effectively used to reduce the documents publishing costs.

DLIBRA DIGITAL LIBRARY FRAMEWORK

The architecture of the dLibra system consists of a number of essential components. The primary element of the system structure is the dLibra server divided into some specialized modules. They are responsible for performing operations on various dLibra objects (see Table 1). The Content module is responsible for storing digital content of gathered resources. The Metadata module allows for various operations on resources metadata. The Search and Distributed Search modules are responsible for indexing and searching functionality on local and distributed resources. The Rights module is used to determine access rights to specified resources. The Events module is an internal mechanism for inter-module notifications. Other system elements are System Services (SS) and User Interface (UI). SS supports communication between server modules and other system components and helps in remote modules addresses resolving. UI is an access point for the client applications which allows users to work with documents, manage the digital library, search, browse and read the resources gathered in the library. Communication between clients and server modules is done using the Java RMI mechanism.

Table 1: dLibra Server Modules Functionality

Module	Functionality
Content [C]	Stores contents of the documents in the file system
Distributed Search [DS]	Collects metadata from remote dLibra servers and stores them in the local index
Events [E]	Sends event notifications to dLibra components
Metadata [M]	Performs various operations on documents (add, modify, delete) and their metadata
Rights [R]	Performs rights control
Search [S]	Builds index of the documents content and metadata and performs searching

The standard configuration of the dLibra server (Part A of Figure 1) assumes that all modules are located on the same machine. However, dLibra is a highly flexible system and the server modules may be spread all over the network and placed on completely different machines (Part B of Figure 1). Each machine has its own System Services that communicate with other System Services in order to get access to the remote modules. From the client applications' point of view, the dLibra system has still only one access point, which hides the real server modules location.

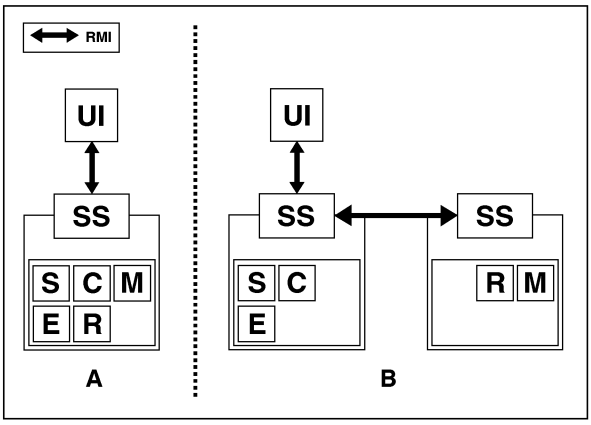


Figure 1: Basic dLibra Architecture

What is not shown in the above figure is that dLibra uses a database to store all information about objects maintained by the system. The location of the DBMS is independent of the location of the dLibra server. If the machine used for the dLibra server does not have enough resources to serve the database, another machine can be used. The documents

contents are stored in the file system of a server on which the Content Module is deployed.

dLibra, as an advanced digital library system, provides a flexible way of describing documents with a fully configurable set of attributes. If dLibra is used to create a standard digital library, the built-in Dublin Core version 1.1 compatible, set of attributes may be successfully used. However, a company may extend or modify the attributes set whenever there is a need to describe added documents in a special way. There are also specialized mechanisms like attributes values dictionaries and thesaurus to improve the metadata quality and search capabilities. Additionally, access to each resource placed in the dLibra-based DL system is precisely controlled by the advanced rights mechanism with features like multiple types of access levels, rights inheritance, and user grouping. It also supports different authorization methods from simple username/password combination to LDAP usage and IP address based authorization.

ARCHITECTURE FOR BUSINESS SOLUTIONS

dLibra In SME

Resources held in the dLibra-based digital library are organized in the hierarchical structure consisting of directories, publications, editions and file versions. Additionally, publications can be associated to a fully definable set of collections. A single document should be treated as a dLibra publication. Documents may have several versions which would correspond to dLibra editions. Having this in mind, we can imagine using such structure for storing and managing documents of a small or medium firm. The firm consists of several departments producing internal documents of all kinds. Separation of the documents belonging to different departments is provided by the dLibra system using directories. Each department has its own directory. Employees can add, modify or delete documents only in their department's directory. Documents from all directories are put into the index and can be searched by employees of all departments. There is a WWW portal that allows searching, browsing and reading documents. However, certain documents are available for reading only to the employees with appropriate access rights.

The dLibra administrator can create a collection named, for example “Technical documentation”, and every document that is qualified to the technical documentation category should be associated with it. As collections are visible for the users in the WWW portal, one might perform a search only within the “Technical documentation” collection. Associating documents to the thematic collections may be very helpful in finding the appropriate resource, as it narrows the documents search range.

Simple Distributed Solution

dLibra provides a fully distributed solution for the digital library. If we think of a larger firm with its departments located in distant parts of a city or country, dLibra comes with a suitable solution again. It is possible to install a separate instance of the dLibra system for each department (Figure 2). There is a separate database, server, documents index and WWW portal for each department. In this case the physical separation of the departments’ resources is provided. The directories structure may be completely different in each dLibra installation and may not correspond to the real company structure. However, administrators can define the same sets of collections in each department to unify the structure of every digital library instance from WWW point of view. Such a division may be helpful in distributed searches (Larson 2003).

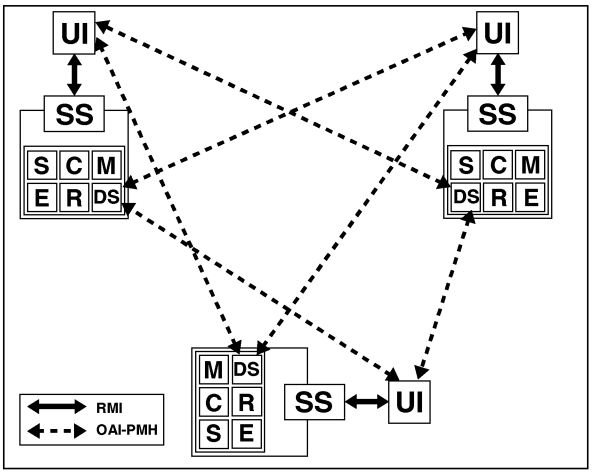


Figure 2: Distributed Digital Library Architecture

dLibra provides a Distributed Search module that supports indexing documents from the remote dLibra servers. Currently only metadata of the remote documents is indexed but indexing remote documents contents is also possible. Communication between remote dLibra servers is based on an OAI-PMH protocol (Lagoze and Van de Sompel 2001).

Each dLibra instance allows searching both local and remote documents. The actual searching takes place in the local index which contains contents and metadata of the local documents and only metadata of the remote documents. To remove unwanted search results, the employees are free to choose whether the search results should also contain remote documents. Each search result is presented together with information on its location (local or remote). Access to the remote server is required only after a document's link has been followed.

Large Scale Solution

Solutions presented in the first part of this article are directed to the firms that produce mainly text documents. The described architectures do not give users any

possibility to perform sophisticated searches within documents of other types. dLibra may be used on a larger scale, for example in a company which uses the digital library to store large video documents.

This approach uses a two-level architecture which gives much better scalability possibilities (Figure 3). The number of machines used on the lower level is almost unlimited. It is determined by the power of the higher level machine only. The lower level of the structure consists of dLibra servers located in the company departments. None of those servers contains any content module, as the large video content is stored on the higher level. The lower level servers maintain indexes of the video content metadata only. Therefore the machines used as the department servers need not be very powerful, which reduces the purchase costs.

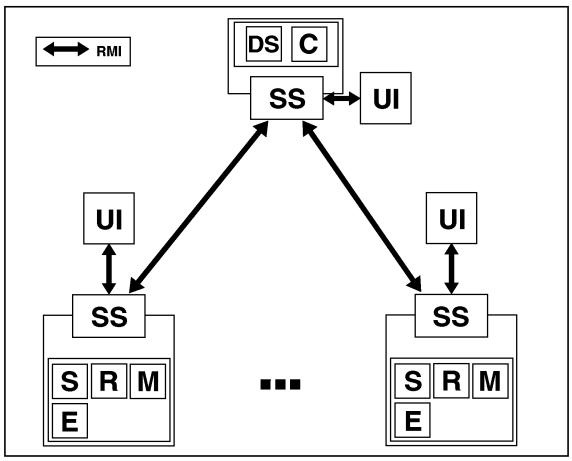


Figure 3: Two-level Architecture with Dedicated Server for Video Content

The server on the higher level stores video files and provides mechanisms to search certain parts of them within the video content index. The machine used for this purpose should be fast enough to serve all requests coming from the department servers. The size of the files that are stored in the system should also be taken into consideration while completing the main server configuration.

Users of the department dLibra servers can perform searching in the local metadata indexes without accessing the main server. Access to the main server is necessary when viewing or searching the video files. Searching the content is done by the main server's DS (Distributed Search) module. Despite having the same name, the DS module is different from the one used in the previous solution. There should be a specialized component that is able to index and search video data.

The hardware requirements of the main server are quite high and not every company can afford it. Less wealthy companies may consider outsourcing the higher level of this architecture. We can think of outsourcing the server machine (hardware) or the whole dLibra server (hardware and software). When the company outsources the hardware, it only has to do the installation and administration of the dLibra server itself. In the latter case, the whole main server management is done by the external company.

The two-level architecture may also be used for companies storing small content like text documents. The content module may be placed in each department server. Searching the content would be available at the department level separately. The main server, being the access portal for employees of the whole company, would support inter department searches. We can even extend this structure with another portal. The additional digital library access point would support searching and reading documents that the company wants to make available for its customers. In this case using firewall would be sensible. The only ports needed to get WWW access to the dLibra digital library are port 80 and 443. Together with the access rights it would ensure the safety of the internal company documents.

PUBLISHING WITH XML USAGE

Example business functionality can be achieved with the DL system supporting XML based documents. XML documents can be created with many advanced but freeware editors like XXE or Open Office suite. These documents can be placed in the company digital library, which should have a possibility to transform XML documents with many different user defined XSL style sheets.

Depending on the chosen DL system architecture we can have different XSL style sheets defined on different levels of the system structure. In the solution presented in Figure 1, all style sheets are defined in the main DL server. In the first of the distributed solutions, each company branch can have its own specific style sheets defined. In the last, two level architecture, again each company branch can have its own style sheets defined, and all style sheets can be accessible on the top level.

Such XSL style sheets can have different functionality. In basic usage we can assume that the created documents are DocBook (http://www.docbook.org/) compliant company publications such as reports, guides or even official correspondence. With the use of unified style sheets, the user does not have to care about the final look of the created documents. The designer prepares XSL files corresponding to the official company report or fax layout and all documents look exactly the same. In this scenario the digital library becomes not only a store for information but also a small publishing framework.

However, we can go one step further. XML documents do not have to be limited to typical office documents. Recently we can observe evolution of many business related XML based standards. For example, there is the SRS XML project (http://xmlbasedsrs.tigris.org/), whose main aim is to create a set of XML documents definitions useful for writing software requirements specifications as XML documents. With such projects there are not only XML documents definitions, but also XSL style sheets for generating different kinds of reports and comparisons. Adding similar style sheets to a DL system can make it an

advanced reporting tool with all other DL functionality like access control or advanced search mechanisms. This is another situation where a DL based solution is much cheaper than commercial, off the shelf product specialized in one domain, like Rational Requisite Pro and Rational SODA – tools for storing software requirements specifications and for generating various reports based on this data.

CONCLUSIONS

The business need for storage, management and exchange of large quantities of documents can be satisfied by sophisticated and specialized, but expensive commercial systems. However, there is a significantly cheaper, but not worse alternative. Digital library systems can be used for this purpose. Properly designed digital library systems are very usable in various configurations. Advanced distribution possibilities are crucial when it comes to large, enterprise systems. Using an efficiently designed digital library system we can create flexible and scalable company document repository with DL characteristic, advanced possibilities like metadata document description, content searching, collections, sophisticated access control or even knowledge discovery. Distribution gives a possibility to reduce the number of most expensive elements of such system or even outsource them. The outsourcing level is highly customizable – it can be done with the entire system or only with the most expensive parts. Various solutions presented in this article are based on different configurations of dLibra architecture components. Our future works will focus on testing and evaluation of presented configurations.

Using DL systems also gives additional benefits when connected with XML usage. Basic functionality allows unifying company documents layout. Advanced XML standards together with proper DL system configuration can give a business customer functionality offered by many commercial systems.

We believe that digital libraries can be successfully used not only in academic, educational or cultural areas. These systems, if efficiently used, can become powerful business tools giving great possibilities in enterprises of different size – from small text document stores to large audio and video repositories, giving precise access control and very high quality of service.

BIBLIOGRAPHY

Falquet, G.; Mottaz-Jiang, C.; Ziswiller, J. "Ontology Based Interfaces to Access a Library of Virtual Hyperbooks" at 8th ECDL, Bath, UK, September 2004.

Gruszczyński, P.; Mazurek, C.; Osinski S.; Swedrzynski A.; Szuber S. "dLibra Content Maintenance for Digital Libraries" in Euromedia'2002, pages 28–32, 7*th* Annual Scientific Conference, April 2002.

Kosiedowski, M.; Mazurek, C; Werla, M. – „Digital Library Grid Scenarios" in European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, 25-26.05.2004, London, U.K. Workshop Proceedings, p. 189 – 196.

Lagoze, C.; Van de Sompel, H. – "The Open Archives Initiative: Building a low-barrier interoperability framework", pages 54-62, Proceedings of the 1st ACM/IEEE-CS JCDL, Roanoke, VA, USA, June 2001.

Larson, R. R. "Distributed IR for Digital Libraries" in LNCS 2769, p. 487 – 498, 7th ECDL, Torndheim, Norway, August 2003.

Mazurek C.; Stroiński M.; Swędrzyński A.; „dLibra – Integrated Framework for Publishers and Libraries" – poster at 7th ECDL, Torndheim, Norway, August 2003.

Payette, S.; Staples, T. "The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services" in LNCS 2458, p. 406 – 421, 6th ECDL, Rome, Italy, September 2002.

Pyrounakis, G.; Saidis K.; Nikolaidou, M.; Lourdi, I. "Designing an Integrated Digital Library Framework to support Multiple Heterogeneous Collections" at 8th ECDL, Bath, UK, September 2004.

Witten, I. H.; Bainbridge, D.; Paynter, G.; Boddie, S. "Importing Documents and Metadata into Digital Libraries: Requirements Analysis and an Extensible Architecture" in LNCS 2458, p. 390 – 405, 6th ECDL, Rome, Italy, September 2002.

AUTHOR BIOGRAPHY

MARCIN HELIŃSKI, born in 1974, received his M.Sc. Eng. in Computer Science at the Poznan University of Technology (PUT) in 1998. He worked for PUT between 1998 and 2003. Since 2003 he has worked for the Poznan Supercomputing and Networking Center as a member of the dLibra Digital Library Framework team. His interests concern digital libraries, software engineering and designing GUI interfaces.

CEZARY MAZUREK, Ph.D., is the head of the Network Services Department at the Poznan Supercomputing and Networking Center. He received his PhD degree in computer science from the Poznan University of Technology in 2004. His research interests concern a wide variety of advanced network services including portal solutions, digital multimedia libraries, streaming technologies, distance learning and access to grid services. He has been the manager of numerous projects in those fields coordinated by PSNC. Some of the major ones include the Multimedia City Guide, Polish Educational Portal, Digital Library Framework: dLibra, Nabor 2003-2004, PROGRESS and interactive TV platform.

MARCIN WERLA, M.Sc. Eng., is the project leader of the dLibra Digital Library Framework project developed in Poznan Supercomputing and Networking Center. He received his M.Sc. Eng. degree in computer science from the Poznan University of Technology in 2004. His professional interests concern digital libraries, distributed services, software development and engineering and project management.

Design of A Scalable Collaborative Virtual Environment System

Lin Qingping, Zhang Liang, Irma Kusuma
Information Communication Institute of Singapore
School of Electrical & Electronic Engineering
Nanyang Technological University
iqplin@ntu.edu.sg

Abstract

The scalability of the existing Collaborative Virtual Environment (CVE) systems is limited due to the constraints in computer processing power and network bandwidth. In this paper, we propose a mobile agent-based large-scale CVE (LCVE) system to support a large number of concurrent participants in a CVE with a large amount of evolving virtual entities. The LCVE system comprise of mobile agents, each of which is responsible for an independent task for the LCVE. With the mobility and autonomy of mobile agents, the workload of the LCVE will be balanced dynamically; the management of LCVE contents will be autonomous at agent level. This paper presents the design and preliminary implementation of the proposed mobile agent based LCVE system.

1. Introduction

Collaborative Virtual Environment (CVE) systems support multiple geographical dispersed human-to-human & human-to-machine communication and real-time interaction in a shared virtual environment. CVE provides users a three-dimensional shared virtual world over the networks, where participants can interact with each other naturally. It has many promising applications, such as tele-conferencing, military and industrial team training, collaborative design and engineering, e-learning, and multiplayer games, etc. One of the main issues in CVE its scalability in supporting large number of participants geographically distributed over the Internet to interact in a common virtual world with large number of virtual entities including static, dynamic and evolving objects. Extensive research work in LCVE has been done by both academia and industry.

There exist numerous systems supporting LCVE, such as AVIARY [1], BrickNet [2], DIVE [3], MASSIVE [4], NPSNET [5], PARADISE [6], RING [7], SPLINE [8], VLNET [9] and VELVET [10]. These systems they adopt specific mechanisms or technologies for some aspects of a particular application to achieve the scalability. For example, MASSIVE adopts Awareness Management to reduce the unnecessary communication traffic; NPSNET adopts IP-Multicast technology to improve the transmission efficacy; SPLINE adopts Locale notion to confine the information exchange in a small group, etc. However, the scalability of the existing system architecture is still an issue in an unpredictable network traffic environment over the Internet or an ever-evolving VE system. In this paper, we address this scalability issue from the aspect of system architecture by proposing a mobile agent-based LCVE.

Mobile agent paradigm is becoming increasingly popular for network-centric systems. Mobile agent is an autonomous entity that can migrate from one machine to another in a heterogeneous network. Compared to traditional distributed computing schemes, mobile agents promise (at least in many cases) to cope more efficiently and elegantly with a dynamic, heterogeneous, and open environment which is characteristic for today's Internet [11] [12]. Thus, we apply mobile agent paradigm to LCVE to support a large number of concurrent participants in a VE with a large amount of evolving virtual entities.

This paper is organized as follows. Section 2 will review the related work; Section 3 will discuss the design and preliminary implementation of the proposed system; Section 4 will draw conclusions and outline the future work.

2. Related work

A number of existing CVE systems have attempted to address the issue of scalability. In this section, we review some well-known LCVE systems.

NPSNET

NPSNET [13] (Naval Postgraduate School Networked Vehicle Simulator) is a 3D networked virtual environment system developed at the Computer Science Department of the U.S. Naval Postgraduate School in 1993. It is designed to support large-scale military training and simulation exercises. NPSNET-IV is the most successful system to support LCVE. It complied with the Distributed Interactive Simulation (DIS) protocol to interoperate with other simulation system, and incorporates the dead-reckoning algorithms and IP Multicast protocol to reduce network traffic to achieve scalability.
NPSNET-IV is the first CVE system to adopt the multicast. And it is also the first successful LCVE system. In NPSNET-IV, it logically partition virtual environments by associating spatial, temporal, and functionally related entity classes with network multicast groups. An entity can belong to several groups. For example different media communication can belong to different multicast group, say, position message and audio message. Multicast can save the bandwidth, filtering different kinds of traffic in the network interface hardware and does not consume processor cycles. What's more, it makes communication architecture easy to realize.
However, NPSNET is running at dedicated hosts and networks; the requirement of NPSNET system is extremely demanding in terms of network and computational resource [14].

MASSIVE

MASSIVE [15] (The Model, Architecture and System for Spatial Interaction in Virtual Environments) was developed as an experimental CVE system by the Computer Science Department of the University of Nottingham, which aims to be a large-scale multi-user Virtual Environment with rich facilities to support user interaction and cooperation. Though, it does not fully address the scalability issue, MASSIVE propose a spatial model of interaction that is very useful for the later research. In the spatial model there are three basic notions --- aura, focus, and nimbus.

- Aura: Each object in a virtual world has an aura for each medium in which it can interact. This aura defines the extent to which interaction with other objects is possible.
- Focus: Focus describes the observer's allocation of attention.
- Nimbus: Nimbus describes the observed object's manifestation or observability.

In MASSIVE, the clients communicate through peer-to-peer connections and the spatial trader serves as an aura manager to check for aura collisions. If aura manager detect a collision it will notifies any objects concerned to set up a peer-to-peer connection and find out about each other. When the peer-to-peer connection is enabled, the communication is managed according to mutual levels of awareness which are negotiated through the use of focus and nimbus. The calculation of mutual awareness levels is the responsibility of the peer objects. This model can decrease the unnecessary communication traffic.
However, the communication between spatial trader and each peer is still a client/server mode. The spatial trader will become a bottleneck of the whole system as the number of participants increases to a certain level.

SPLINE

SPLINE [16] (Scalable Platform for Large Interactive Networked Environments), developed at the Mitsubishi Electric Research Laboratory in 1996, is a software platform suitable for implementing multi-user interactive environments. SPLINE aims to make the multi-user Virtual Environment large in spatial extent, large in number of objects, and large in numbers of users interacting with the environment.

One of the most interesting features of SPLINE is the definition of locales. The concept of locale is based on the idea that even in a very large virtual world, most of what a single user can observe at a given moment is nevertheless local in nature. Locales divide a virtual world into chunks that can be processed separately. This division is purely an implementation issue—it is not apparent to the user. A user sees several locales at once—generally the locale containing the user's point of view and those neighboring it.

Each locale is associated with a separate set of multicast addresses, which guarantees the efficiency of the communication. Each locale has its own coordinate system, which guarantees precise position and velocity information about objects that are arrayed across a large volume of space. Each locale can has arbitrary shape, size, and orientation which allows designers to create parts of a virtual environment separately and combine them together later[17]. The locales notion can make VE spatially scalable to an arbitary extend. However, similar to NPSNET, supporting heterogenous LCVE is still an issue.

DWTP

DWTP (the Distributed Worlds Transfer and communication Protocol) [18] is an application layer protocol proposed by *Broll* from GMD. It provides a flexible and scalable network architecture for LCVE, which is not based on Client/Server approach including a single daemon to provide the required service, but uses a set of different daemons to provide particular services to the participants. Each daemon is responsible for a duty. DWTP provides four different daemons,

- World daemons, to download static or dynamic files and to keep modifiable worlds persistent.
- Reliability daemons to provide reliable transfers by unreliable network protocols.
- Recovery daemons to receive lost messages.
- Unicast daemons to provide access via unicast network connections.

These daemons can either be combined in a single program to provide all the services required to realize a single shared virtual environment or they maybe split among several hosts. The basic communication architecture of DWTP is based on IP multicasting groups [19]. The heterogeneous architecture of the approach also allows non-multicast capable participants to join shared virtual worlds [20].

This architecture is flexible that the individual daemons can be distributed on several hosts to reduce the load at each host. Additionally most daemons (except the reliability daemon) can be replicated to split the load of a particular daemon between several hosts. However run-time autonomous adapting system architecture to support evolving virtual entities and participants is not possible in this approach.

3. Design of Mobile-Agent based LCVE

We attempt to enhance the existing system architectures by proposing a mobile agent based LCVE system. In our proposed approach, each mobile agent in a LCVE system will be responsible for an independent task for the LCVE, such as the management of the VE contents, maintenance of consistency and persistency of a LCVE. Mobile agents will execute in system-registered nodes, and when necessary, agents will move from time to time to other system-registered nodes. The features of mobile agents (e.g. cloning, autonomous decision making and computing, and dynamic movement in a network) will be exploited in order to achieve scalability and extensibility of LCVE. The following sections will present the system architecture and design of the mobile agent based LCVE.

3.1 Basic Concepts

In a LCVE, there are a large number of computer hosts (nodes) connected to the virtual environment. In this system, the participating nodes are divided into two types: *User Nodes* and *Service Provider Nodes*. The classification is made based on the functional roles played by each of the nodes. User Nodes are the participating nodes that logged on to the system. Service Provider Nodes are hosts owned by the LCVE owner. The complete classification of nodes is illustrated in Figure 1.

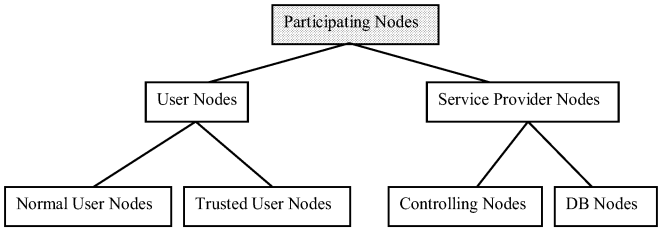


Figure 1 Classification of LCVE Nodes

Referring to Figure 1, User Nodes and Service Provider Nodes are both further divided into sub-types. User Nodes are divided into Normal User Nodes and Trusted User Nodes:

Normal User Nodes are nodes that perform only basic actions such as navigating inside the virtual environment and interacting with virtual entities and other CVE participants.

Trusted User Nodes are nodes that have a higher computing ability that can be used to assume some of the task in the LCVE management. Thus, besides performing basic actions as Normal User Nodes, these nodes also take part in system management. Because of this, it is very important for these nodes must meet the requirement set by the LCVE owner, and notify the system when they are logging off from the system such that the computing task assigned to them can be re-assigned accordingly.

Service Provider Nodes are also divided into two types of nodes:

Controlling Nodes are nodes owned by the system owner that performs some tasks for LCVE system management and maintenance.

DB Nodes are nodes owned by the system owner that provides database support for the system.

Trusted User Nodes and Controlling Nodes both are used to share the workloads in performing computing tasks for system management. Therefore, both of them are further classified together as **System Controlling Nodes (SC Nodes)**. Next section will elaborate on the architecture of the mobile agent-based LCVE.

3.2 System Architecture

The proposed mobile agent-based LCVE is a fully distributed system. It comprises of mobile agents, each mobile agent is a software component that performs one of the management, maintenance, and control task by collaborating with other agent(s). The agents will be deployed to different participating nodes and able to move and reproduce themselves when necessary. The system architecture is visualized by Figure 2.

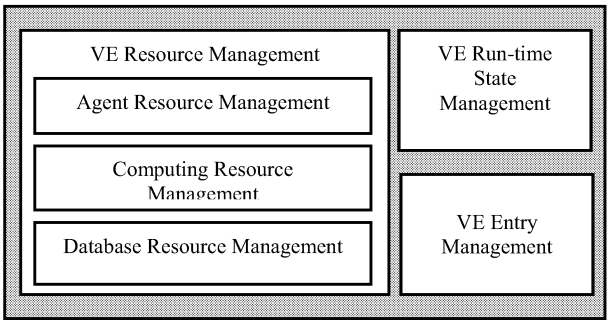


Figure 2 Mobile Agent-Based LCVE Architecture

It can be seen from Figure 2 that the system architecture consists of three parts: VE Resource Management, VE Run-Time State Management, and VE Entry Management. Mobile agents are allocated to each of the three to perform the corresponding management tasks as described in the following paragraphs.

VE Resource Management is the *resource provider* that handles all tasks related to system resource management, and it is further divided into:

Agent Resource Management (ARM): responsible for managing all mobile agents. In particular, it is responsible in the registration of the mobile agents, allocating them to members of the nodes for them to perform their tasks, and upgrading the agents with new functionalities. ARM is also responsible for allocating SC Nodes to different groups for different functionalities, monitoring them, and balancing workload among them.

Computing Resource Management (CRM): responsible for managing the computing resources (processing power, bandwidth) which are shared among the SC Nodes.

Database Resource Management (DRM): responsible for managing the database resources in the system. The database is distributed among the DB Nodes (refer to Figure 1) in order to support the large size virtual environment. DRM is responsible for registering active DB nodes, monitoring the storage and fetching capacity of the active DB Nodes, and performing load balancing among the DB Nodes. This will ensure that the database management is done efficient and effectively.

VE Run-Time State Management is the *content provider* that maintains and manages the run-time state of the virtual environment. It is responsible for sending relevant part of the virtual environment to participating nodes and maintaining the virtual world states' persistency and consistency. The run-time management is achieved by the collaboration of Region Agents, Cell Agents, Persistency Agents, and Consistency Agents (will be explained in Section 3.2.1).

VE Entry Management is the *entry service provider* that handles the joining of new participants into the system. It is responsible for user authentication and directing the user node to the VE Run-Time State Management. This is done by way of

Gateway Agents (will be explained in Section 3.2.2) to which users will be registered when first joining the system.

In this system, each of the management aspects mentioned above is achieved by one or more type(s) of mobile agents, several of them have been mentioned briefly. So mobile agents working in this system are divided into several types and each agent performs their tasks accordingly. The classification of agents is shown in Table 1.

Management Aspects		Agents
VE Resource Management	Agent Resource Management	ARM Agent
	Computing Resource Management	CRM Agent, Group Manager Agents, Node Agents
	Database Resource Management	DRM Agent, DB Agents
VE Run-time state Management		Region Agents, Cell Agents, Persistency Agents, Consistency Agents
VE Entry Management		Gateway Agents

Table 1 Classification of Mobile Agents based on Logical Aspect

In next section, each of the management aspects will be explained in more details.

3.2.1 VE Resource Management

Agent Resource Management (ARM)

ARM is achieved by ARM Agent. When LCVE first constructed, all agents need to register to ARM Agent. Likewise, when there is new agent joining the system at run time, it will need to register to the ARM Agent. After ARM Agent has received the information regarding the existing agents, it is responsible to distribute the mobile agents to different nodes based on the current states of the system. This is to achieve a distribution of tasks, load balancing, and to enable system extensibility.

Task distribution is especially important during system's initialization and evolvement. During LCVE system initialization, ARM Agent will distribute the task of managing the run-time state by allocating the agents of the VE Run-Time State Management to the nodes to do the necessary system setup. When the CVE content evolves, ARM will distribute mobile agents to manage the newly extended environment. Another case is when new SC Nodes are joined to the system, CRM can ask ARM to allocate Group Manager Agent to manage the new group of SC Nodes.

ARM also has the responsibility to upgrade the mobile agents. Examples of upgrading scenarios are when there is new management mechanism introduced to the system or when there is a newer version of mobile agent. In those cases, ARM Agent will be the one responsible to replace/upgrade the agents currently running in the system.

Computing Resource Management (CRM)

In this system, workload is distributed between Service Provider Nodes and Trusted User Nodes. Therefore, Trusted User Nodes have to be part of User Nodes that have the computing resource above certain requirement and are willing to contribute its resource for the management task of the system. Besides avoiding bottleneck, this mechanism also enhances the system's reliability. The responsibility of CRM is to manage the computing resources such that a sufficient amount of resources can be allocated to each computing task.

Each node belongs to the SC Nodes has Node Agent running in it. This Node Agent is responsible for the local computing resource management of its host. To make the system nodes more manageable, the SC Nodes are divided into groups, and each group is controlled by a Group Manager Agent which runs in a node called Group Manager Node. Group Manager Agent is responsible for monitoring and ensuring balancing of workload between nodes in its group. When there is one node whose workload has exceed its capability, Group Manager Agent has to first try to find within the group a suitable node with which the workload can be shared. If the ideal node is not found locally, Group Manager Agent will communicate with the CRM Agent to negotiate with other Group Manager Agent to search for the ideal node in other groups.

In this system, hosts belong to the same subnet are most likely to be put into the same group. This is under the assumption that those belong to one subnet tend to have a higher communication speed. Thus, they can communicate faster to facilitate task transferring. Figure 3 illustrates the management hierarchy of the agents involved in CRM.

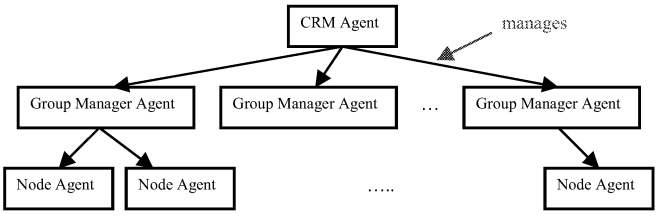


Figure 3 Structure of Agents for CRM

CRM Agent manages the node groups. When a SC Node first joins the system, it will register to CRM Agent, and CRM Agent will allocate it to one of the node groups. Node groups can be created dynamically, for example when existing node groups have reached their limit. If that need arises, CRM Agent will request ARM Agent to send a new Group Manager Agent to manage the newly formed node group.

Group Manager Agent (as explained previously) is responsible for managing the computing resource management within its group.

Node Agent monitors its host's computing load and network traffic. When the load exceeds certain limit, the Node Agent can negotiate with other Node Agents to share and balance the load.

This can be done by cloning or transferring mobile agents to other nodes.

Database Resource Management (DRM)

The database of LCVE consists of a large number of VE content and configuration data. The database is distributed among the DB Nodes, and new DB Nodes can be added to the system at run time to support extensibility. DRM has the responsibility to register the DB Nodes, monitor DB Nodes' computing and fetch capacity, and manage the data storage. To achieve this, there are two types of agents involved in DRM: DRM Agent and DB Agent, as shown in Figure 4.

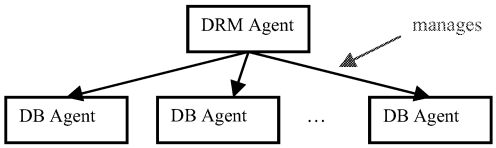


Figure 4 Structure of Agents for DRM

DRM Agent is responsible to manage all the DB Nodes in the system. All DB Nodes need to register to DB Agent, and their storage and fetch capacity will be continuously monitored by DRM Agent. DRM Agent maintains a table containing the DB Nodes addresses for each *cell* (the smallest spatial unit of VE). DRM Agent also responsible in allocating DB Nodes for extending VE content and for load sharing when there is a DB Node which has reached its limit. When a DB Node is allocated to share storage load with another DB Node, there might be times when data retrieval must be done from both of the DB Nodes.

DB Agent resides in each of DB Nodes and is responsible to continuously monitor its own node's computing and storage capacity to be reported to DRM Agent. DB Agent is also responsible for the reading and writing of data to the database portion stored in its node.

3.2.1 VE Run-Time State Management

The virtual environment is spatially divided spatially to a number of manageable regions. Each VE region can be further sub-divided into more child regions and cells if required. The child regions can also have their own child regions and cells. So the division of regions can be recursive and forms nested regions. As mentioned before, cell is the smallest spatial unit of virtual environment. It is the basic unit for virtual entity downloading, communication, persistency and consistency. Communication within a cell is achieved using multicast. Figure 5 shows the example of a VE region's composition.

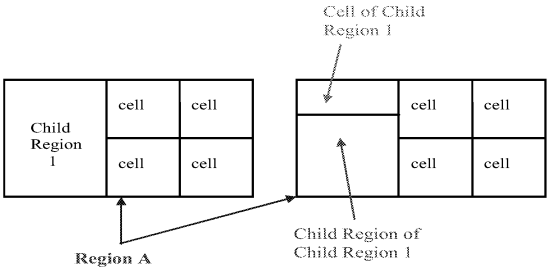


Figure 5 Example of Region Composition

Figure 3.5 shows a region, namely Region A, which consists of 1 child region and 4 cells. As explained before, a child region can be further sub-divided into child regions and cells. That is what happened to Child Region 1. As shown in the figure, Child Region 1 is sub-divided into 1 child region and 1 cell. This architecture is just like the file management system in a computer Operating System, a cell is like a file while a region is like a folder. Region is used to manage cells just as a folder is used to manage files. By having this kind of spatial division of virtual environment, the system will be easier to manage, ensuring that there will not be any region or cell that is so crowded beyond certain limit.

The VE Run-Time State Management is achieved by the collaboration of Region Agents, Cell Agents, Persistency Agents, and Consistency Agents. The relationships among those agents are shown in Figure 6.

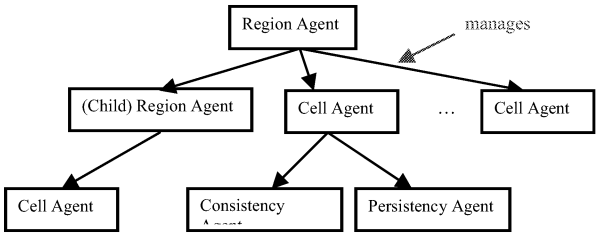


Figure 6 Structure of Agents for VE Run-Time State Management

Figure 6 shows Region Agent is in the top of the hierarchy. Region Agent can manage several other Region Agents and Cell Agents, depending on how the region is divided (refer to Figure 5). There are also Consistency Agent and Persistency Agent running in each cell and managed by the corresponding Cell Agent. Below are the functionalities of each agent:

Region Agent does not directly manages the VE content. Instead, it is responsible for managing the hierarchy of VE content. It monitors the relationship between the child regions and cells and it also stores the information regarding the network address of the (Child) Region Agents and Cell Agents respectively.

Cell Agent tasks are to download the VE content of its cell from the database, maintain the cell VE content, maintain the information of users who are active in the cell, and deliver the current cell VE content to new users.

Persistency Agent is responsible for cell persistency. It maintains the cell VE content by sending the cell history and the latest status of cell content to the database periodically.

Consistency Agent is responsible in maintaining cell consistency. It ensures that any interaction occurs in the cell are sent to those users who need to be aware of the changes (e.g. users active in the cell). Consistency Agent also ensures that there is no ownership conflicts of virtual entities among the cell members.

It is important to note that the regions and cells can be divided or merged dynamically. For example, Region Agent can merge several cells and child regions if all of them have very few active users. After the merging, Region Agent can re-divide the newly merged region into child regions and cells in a different manner according to the need.

During navigation time, users in a cell may only see a portion of the whole virtual environment at one point of time depending their viewpoint. This portion may include the current cell and its neighboring cells. So each time a User Node joins a cell, it needs to download the VE content of its own cell and also the neighboring cells. When the user moves to another cell, the Cell Agent of the new cell in which the user currently residing will contact its parent Region Agent to get the neighboring cells so that the User Node can download the new VE content correspondingly. To ensure smooth navigation, contents of the neighboring cells can be pre-fetched based on user's location and moving vector.

3.2.2 VE Entry Management

VE Entry Management is responsible in authenticating users and directing users into the VE. This management is done by the use of Gateway Agent. When users first connected to the system, they will need to provide their profile/authentication information to the nearest Gateway Agent along with the destination VE that they would like to go to. Gateway Agent then perform authentication based on the user information. If authentication succeeds, the user will be directed to the destination. Thus, Gateway Agent is also responsible in maintaining the record of region names and addresses. If a user is identified as a Trusted User Node, Gateway Agent will direct it to CRM Agent for registration and allocation to a Node Group.

When a User Node leaves the CVE, the Gateway Agent is responsible to store the user's login history in the system database. If the User Node is a Trusted User Node, before leaving, the current tasks assigned to the node needs to be transferred to other SC Nodes. The CRM Agent will also be informed of its leaving the CVE.

4. Implementation

As the existing mobile agent platforms is designed for general purpose which is not suitable for real-time multi-agent communication required in LCVE, we have implemented our own mobile agent platform with our own mobile agent communication protocol designed for LCVE application. Figure 7 illustrate a sample screen capture of mobile agent environment. We have also implemented mobile agents like Cell Agent, Consistency Agent, Persistency Agent, Gateway Agent, Database Agent etc as illustrated in Figures 8, 9, 10, 11, 12 respectively to support the LCVE development. Since there is no multicast support in our university LAN environment, we have also implemented a soft multicast agent, as shown in Figure 13 to support multicast in a cell. Figure 14 shows a sample screen capture of a LCVE developed based on our platform. It can be seen from Figure 14, the display of the virtual world itself is complemented with interface to facilitate users' collaborative interaction of adding, removing, or modifying the objects within the virtual world to evaluate the feasibility of the proposed system. There are also small display areas for monitoring the state of the entities and avatars in the currently displayed cell. Other features like chatting and file transfer between users in the cell have also been implemented in the system.

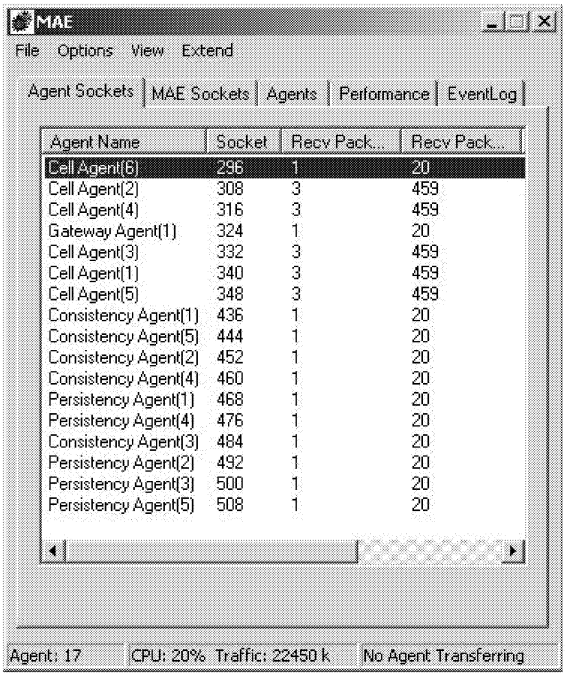


Figure 7. Screen Capture of Mobile Agent Environment

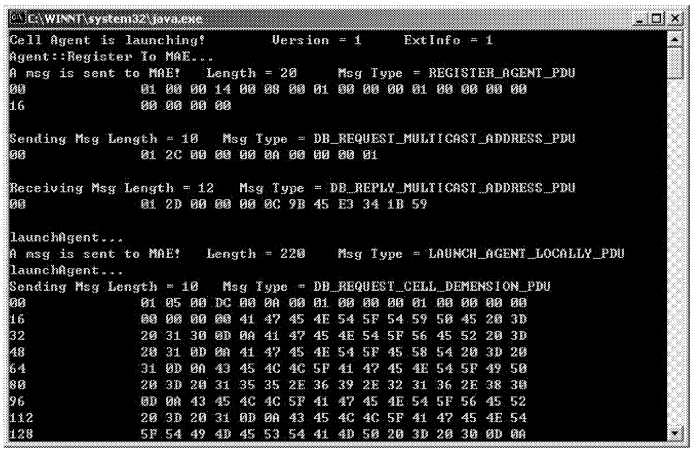


Figure 8. Sample Screen Capture of Cell Agent

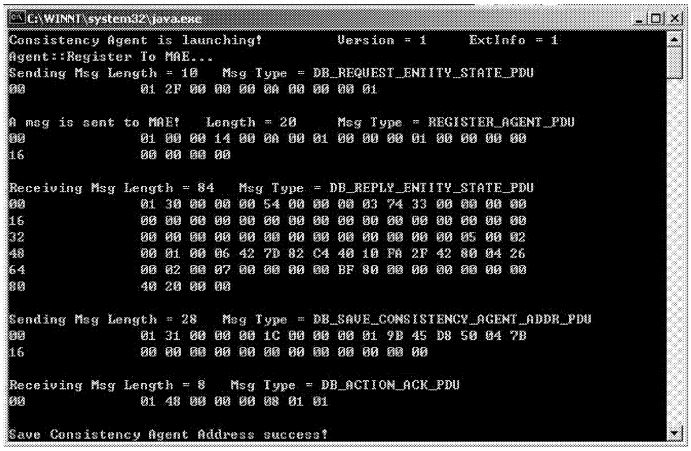


Figure 9. Sample Screen Capture of Consistency Agent

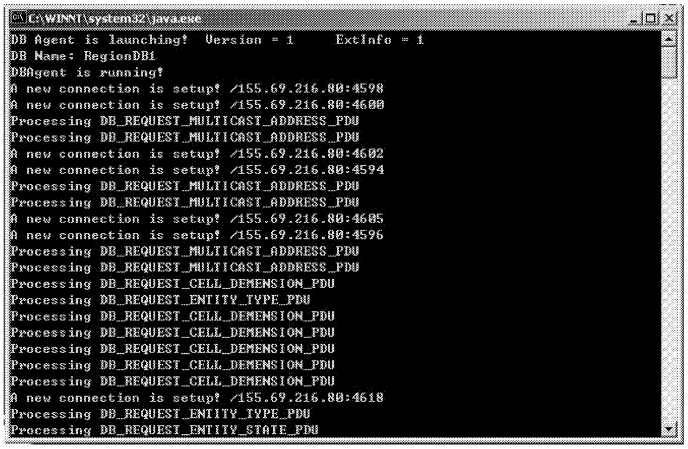


Figure 12. Sample Screen Capture of Database Agent

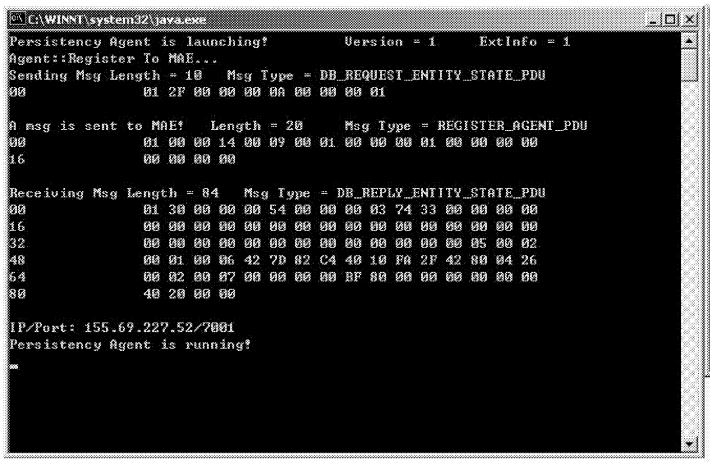


Figure 10. Sample Screen Capture of Persistency Agent

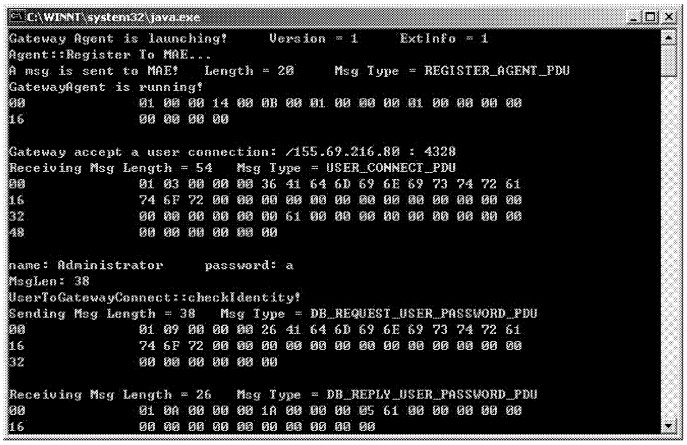


Figure 11. Sample Screen Capture of Gateway Agent

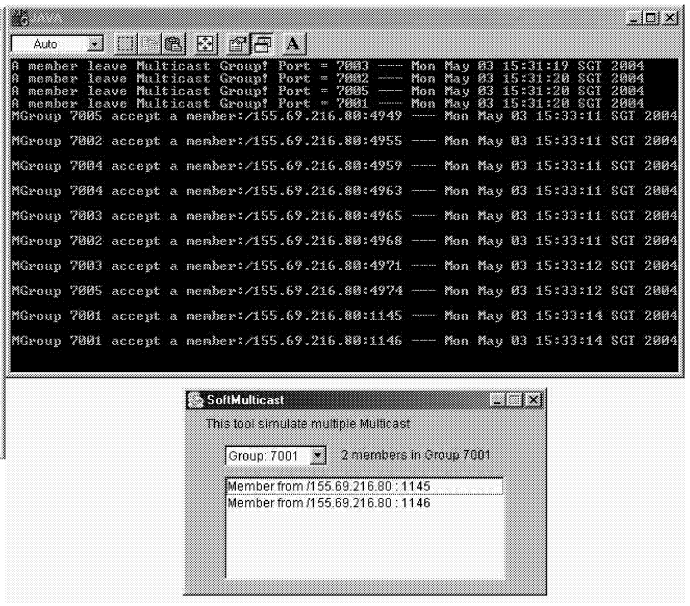


Figure 13. Sample Screen Capture of SoftMulticast Agent

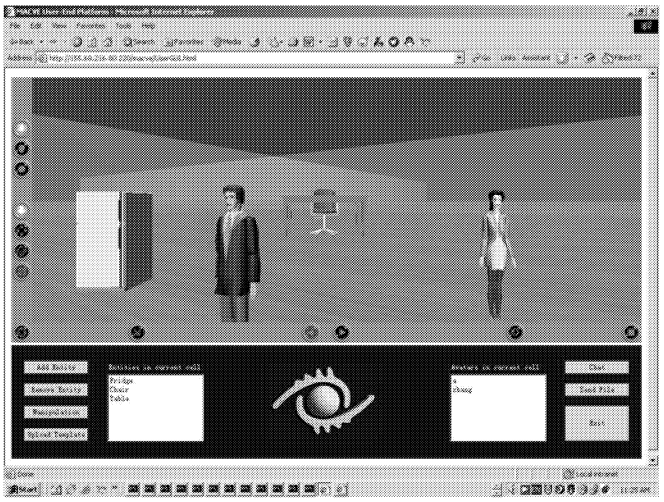


Figure 14. Sample Screen Capture of Mobile Agent based LCVE User Interface.

4 Conclusions and future work

In this paper, we have proposed a novel mobile agent-based architecture for LCVE. This architecture is flexible and autonomous. Our preliminary implementation of the proposed system demonstrates its feasibility. We will further evaluate the system performance in details in our future work. We argue that with our architecture, the LCVE system's scalability will be improved.

Our future work will also include improving the detailed collaboration mechanism for different mobile agents; and incorporating the proposed mobile agent-based architecture in our existing Smart CU3D CVE system. We intend to make our platform available for free to academic non-commercial use.

5 Reference

[1] Snowdon, D. and West, A. "AVIARY: Design Issues for Future Large-Scale Virtual Environments", *Presence* Vol. 3 N.4, 1994.

[2] Singh, G., Serra, L., Pang, W. and Ng, H., "BrickNet: A Software Toolkit for Network-Based Virtual Worlds", *Presence, MIT Press,* Vol. 3, No. 1, 1994, pp. 19-34.

[3] Hagsand, O. "Interactive Multiuser VEs in the DIVE System", *IEEE Multimedia,* Spring 1996, Vol. 3, N. 1, IEEE Computer Society, pp. 30-39.

[4] Greenhalgh, C. and Benford, S., "MASSIVE: A Collaborative Virtual Environment for Teleconferencing", *ACM Transactions on Computer Human Interfaces,* Vol. 2, N. 3, pp. 239-261, September 1995.

[5] Macedonia, M., Zyda, M., Pratt, D., Barham, P. and Zestwitz, S., "NPSNET: A Network Software Architecture for Large-Scale Virtual Environments", *in Presence,* Vol. 3, No. 4, Fall 1994, pp. 265--287, MIT Press.

[6] Singhal, S. and Cheriton, D., "Using Projection Aggregations to Support Scalability in Distributed Simulation", *Proc. International Conference on Distributed Computing Systems (ICDCS'96),* 1996.

[7] Funkhouser, T., "RING: A Client-Server System for Multi-User Virtual Environments", *Proc. Siggraph Symposium on Interactive 3D Graphics,* Monterey, April, 1996.

[8] Abrams, H., Watsen, K. and Zyda, M., "Three-Tiered Interest Management for Large Scale Virtual Environments", *Proc. VRST'98*, Taipei, September, 1998.

[9] Capin, T., Noser, H., Thalmann, D., Pandzic, I. and Thalmann, N., "Virtual Human Representation and Communication in VLNet", *IEEE Computer Graphics and Applications,* Vol. 17, N. 2, Mar-Apr 1997, pp. 42-53.

[10] Oliveira, J.C., Georganas, N.D., "VELVET: An adaptive hybrid architecture for very large virtual environments". *Proc. 2002 IEEE International Communication Conference (ICC'02).*

[11] Fnfrocken, S., and Mattern, F. "Mobile agents as an architectural concept for internet-based distributed applications: the WASP project approach." *Proc. KiVS'99 ("Kommunikation in Verteilten Systemen").* 1999 (29 October 2001).

[12] Cheng-Zhong Xu, Brian Wims, "A mobile agent based push methodology for global parallel computing". *In Concurrency - Practice and Experience* 12(8): 705-726 (2000).

[13] Michael R. Macedonia, Michael J. Zyda, and others, "Exploiting Reality with Multicast Groups: A Network Architecture for Large-scale Virtual Environments", *Proc. Virtual Reality Annual International Symposium (VRAIS'95),* March 11 - 15, 1995.

[14]Manuel Oliveira, Jon Crowcroft , Mel Slater, "Component framework infrastructure for virtual environments", *Proc. the third international conference on Collaborative virtual environments,* p.139-146, September 2000.

[15] Greenhalgh, C. and Benford, S., "MASSIVE: A Distributed Virtual Reality System Incorporating Spatial Trading", *Proc. the 15th International Conference on Distributed Computing Systems(DCS'95),* Vancouver, Canada,1995, pp.27-34.

[16] Barrus, J.W., Waters, R.C., and Anderson, D.B. (1996), "Locales: Supporting Lange Multiuser Virtual Environments", *in IEEE Computer Graphics and Applications*, 16 (6), Nov., 50-57.

[17] Barrus, J.W., Waters, R.C., Anderson, D.B. "Locales and beacons: efficient and precise support for large multi-user virtual environments". *Proc. 1996 IEEE Virtual Reality Annual International Symposium (VRAIS'96)*, pp. 204--213.

[18] Broll, W., "DWTP – An Internet Protocol for Shared Virtual Environments", *Proc. 3rd Symposium VRML'97*, Monterey, February, 1998, pp. 15-24.

[19] V. Kumar, "MBone: Interactive Multimedia on the Internet". *New Riders Publishing*, Indianapolis, Indiana, 1995.

[20] Broll, W., "SmallTool - A Toolkit for Realizing Shared Virtual Environments on the Internet", *Distributed Sys-tems Engineering, Special Issue on Distributed Virtual Environments*, No. 5, pp. 118-128 (1998).

COMTEC

COMMUNICATION NETWORKS

Load balancing in IP networks in normal operation and with failure resilience

Gerhard Haßlinger, Stefan Schnitter and Martin Franzke
Deutsche Telekom, T-Systems, Technologiezentrum
D-64307 Darmstadt, Germany
E-mail: {gerhard.hasslinger, stefan.schnitter, martin.franzke}@t-systems.com

KEYWORDS

Load balancing, failure resilience, traffic engineering, network planning

ABSTRACT

In the last decade the traffic volume transported over IP based telecommunication networks has increased at an even higher rate than the storage and computation capacities of computer systems. A growing user population exploits an extending spectrum of services with a trend to broadband Internet access. Technology on the optical and the IP layer is developing at corresponding pace to provide sufficient resources to meet the increasing demand. Nevertheless, high speed switching equipment is expensive, demanding for an efficient resource usage.

While network planning and management is controlling a process of steadily upgrading transmission capacities, traffic engineering methods are applied to optimize the throughput in the network platform according to current source to destination traffic demands. In a first step, traffic engineering can be applied to a predefined network topology, but a comprehensive approach has to take into account the process of link capacity adaptation to cope with increasing traffic loads and has to prepare for relevant failure scenarios.

We evaluate the efficiency of traffic engineering together with link upgrade strategies in order to get a maximum throughput in such a way that a considered traffic matrix can be multiplied by a maximum scalar scaling factor. We observed that load balancing can exploit a higher potential of improvement when failure scenarios are included. The influence of the network topology is shown in examples.

1 INTRODUCTION: TRAFFIC MANAGEMENT IN MPLS AND IP NETWORKS

Traffic flows on Internet service provider (ISP) platforms can be directed via classical IP routing protocols like open shortest path first (OSPF) or by making use of an underlying multiprotocol label switching (MPLS) architecture.

IP routing on the network layer assigns an arbitrarily configurable link weight for each link representing its length. The shortest path principle determines the route of each flow from source to destination with regard to the link weights. Since the traffic demand matrix of a backbone topology usually includes much more traffic flows than links, the paths of the flows in the network are correlated and cannot be chosen individually for each flow.

MPLS can provide individual paths to be chosen arbitrarily for each source to destination pair of the traffic matrix. In this way, the highest flexibility for constructing path designs with optimum load balancing is achieved [5][15]. In MPLS networks the complete set of provided paths is denoted as a label switched path (LSP) design.

If shortest routes are overloaded, some demands may be transported on detours in order to balance the traffic and to make full use of available capacities. Offline optimization algorithms can compute the most efficient LSP design.

While MPLS is accompanied with measurement for traffic flows per LSP, link measurement may be the only source of data in pure IP networks to observe the traffic development, in order to upgrade links when the load approaches a threshold and to estimate the traffic demand matrix [3][6][9][25]. Naturally, a lack of precise information about the traffic demand is an obstacle for the design of optimum network structures, which has been studied under different assumptions and objectives e.g. by [4][12][26].

In pure IP networks or when MPLS paths are not provided for all source to destination traffic demands, then the manipulation of IP link weights is a method of choice to balance the traffic while the shortest path principle remains valid. The IP weight optimization method has been investigated by [7][8]. It is apparent that link weight adaptation is restricted, since the number of links and thus routing weights available for manipulation is often much smaller than the number of traffic demands and possible LSP path designs. Nevertheless, this does not seem to be critical for usual network scenarios, where reasonable and often close to optimum results are reported, although on the other hand special cases can be constructed, which differ from the optimum by an arbitrary large factor [7].

Network wide load balancing improves the overall quality of service (QoS) properties, since it reduces the maximum link load in the network. The link load is a determining factor for the probability of temporary overload situations causing delay or packet loss. Vice versa, the network management can allow for a higher traffic demand to be delivered through the network due to load balancing, while a predefined QoS level can still be met.

We observed that the effect of load balancing in usual scenarios for service provider networks may vary in a wide range [6][8][9]. The long term difference in maximum throughput and utilization with and without load balancing has been investigated for network expansion processes by stepwise link upgrades in [16].

In this work we propose a strategy to upgrade link capacities with regard to predefined failure situations, such that all source to destination traffic demands can still be trans-

ported under each considered failure scenario. A corresponding link upgrade strategy is described in Section 4 with an example showing the long term results and a discussion of main properties and their implications. Alternative approaches for the treatment of failure scenarios and for resilient transport over MPLS networks have also been studied by [19][20]. We start with a summary of the algorithms and the effect of load balancing in normal operation without failures in Sections 2 and 3.

2 TRAFFIC ENGINEERING: ALGORITHMS AND TOOL SUPPORT

We have developed the TE-Scout tool to investigate a set of relevant methods for constructing an optimized path design which transports a maximum flow through a network regarding source to destination traffic demands. The basic application is for a single case of a predefined topology and traffic demand matrix.

There is some support to deal with advanced network planning scenarios: Relevant failure situations can be defined by sets of link and node breakdowns to be taken into account in the optimized path design. Bottlenecks in the network topology are indicated, which are characterized by a set of links with maximum load. In addition, network expansion strategies for increasing traffic demands can be analyzed, which suggest the most appropriate link upgrades to overcome current bottlenecks.

The main building blocks of the TE-Scout tool are described in [17], including the input, options for control of the path design optimization and the representation of results. Next we give a brief overview of the algorithms to obtain a balanced load distribution and a corresponding path design. The basic methods are:

- linear programming,
- the max-flow-min-cut principle and
- heuristic optimization approaches.

We define the network topology by a graph $G = (V, E)$ with a set $V = (v_1, \ldots, v_N)$ of nodes and $E \subseteq V \times V$ of unidirectional links with capacities cap_{jk} for each link $e_{jk} \in E$. The traffic demand matrix $T = (t_{jk})$ determines the flows to be delivered from a source v_j to destination v_k in the network.

The optimization goal is to minimize the maximum load on the links for a given network topology and traffic matrix T. Let $f^{PD}(t_{ij}, e_{mn})$ denote the portion of the demand t_{ij} from node i to j, which is transported on link e_{mn} for a considered path design PD, where $i, j, m, n \in V$.

Then $f^{PD}(e_{mn}) = \Sigma_{i,j} f^{PD}(t_{ij}, e_{mn})$ denotes the complete flow intensity on an edge e_{mn} and the utilization of e_{mn} is $l^{PD}(e_{mn}) = f^{PD}(e_{mn}) / cap_{mn}$. The maximum load for a path design is given by $l^{PD}_{\max} = \max_{m,n} l^{PD}(e_{mn})$. Finally, $l_{\min}$ denotes the smallest utilization achievable by load balancing regarding all links in the network using an optimum path design $l_{\min} = \min_{\forall PD} l^{PD}_{\max}$. The optimization goal can be equivalently expressed by considering the traffic matrix to be scaled by some factor $\lambda \in R^+$ and obtaining a maximum factor $\lambda_{\max}$, such that the traffic demands $\lambda_{\max} T$ can still be transported in the network. In the sequel, we refer to the maximum scaling factor $\lambda_{\max}$, which is in fact reciprocal to the minimum achievable utilization $\lambda_{\max} = 1 / l_{\min}$.

Linear Programming

Linear programming [10] is an exact method to determine the maximum admissible traffic load $\lambda_{\max} T$ and a corresponding LSP path design for all traffic demands $\lambda_{\max} t_{ij}$. As an inherent precondition, the linear programming solution may split a traffic demand t_{ij} over several paths, branching from the source or from other transient nodes.

A set of linear equations and boundary conditions describes the flows for transport of the demands t_{ij} based on the variables $f^{PD}(t_{ij}, e_{mn})$. They are related by equations for flow conservation at each node, such that the rate of all incoming traffic equals the rate of the traffic departing from a node, with regard to the traffic on the links as well as sources and destinations of traffic. We introduce variables $f^{PD}_s(e_{mn}) = \Sigma_j f^{PD}(t_{sj}, e_{mn})$ for the aggregated flow intensity of all demands originating at source s on a link e_{mn}. They are sufficient as a basis for linear programming and lead to an essentially reduced dimension of the linear equation system as compared to the complete set of variables $f^{PD}(t_{sj}, e_{mn})$. Then flow balance equations are set up for the traffic from each source s at each node n of the network:

$\forall\, s, n \in V$:

$$\sum_{m:\, e_{mn} \in E} f^{PD}_s(e_{mn}) = t_{sn} + \sum_{k:\, e_{nk} \in E} f^{PD}_s(e_{nk}) \quad \text{for } n \neq s;$$

$$\sum_j t_{sj} = \sum_{k:\, e_{sk} \in E} f^{PD}_s(e_{sk}) \quad \text{at the source } s;$$

where $\forall s, n, m : f^{PD}_s(e_{mn}) \geq 0.$

The flow balance at the source assumes that the path design has no loops, i.e. $\forall m, s: f^{PD}_s(e_{ms}) = 0$.

The capacity of each link is considered as constraint for the traffic rates [10]. The optimization can be used to decide, whether an admissible path design can be found with regard to all constraints. Moreover, we consider the optimization goal of minimizing the maximum utilization $l_{\min}$ in the network, which is equivalent to maximization of the scaling factor $\lambda_{\max} = 1 / l_{\min}$ such that a demand matrix $\lambda_{\max} T$ can still be delivered through the network. $\lambda_{\max} > 1$ means that the installed capacities are sufficient for the demands offered in T.

Traffic allocation via the max-flow-min-cut theorem

The max-flow-min-cut principle can be used to determine a minimum cut through the network, such that the sum of capacities of the links on the cut is a bottleneck for the maximum admissible traffic load $\lambda_{\max} T$.

A cut K is a minimum set of edges, which divides the set V of network nodes into two disjoint subsets S_K and D_K ($S_K \cup D_K = V$; $S_K \cap D_K = \{\}$), such that there is no path left from a node in S_K to a node in D_K when the cut is removed from the network. A mutual exchange of $S_K \leftrightarrow D_K$ defines a different cut.

For each cut K, we compute the ratio C_K / F_K of the sum $C_K = \sum_{i,j:e_{ij} \in K} c_{ij}$ of capacities provided on the edges of the cut to the sum of flows $F_K = \sum_{i \in S_K,\, j \in D_K} t_{ij}$ demanded from a source on one side of the cut to destinations on the other side. The minimum of those ratios $\min_K \{ C_K / F_K \}$ is an upper bound (max-flow-min-cut bound) of the scaling factor λ for possible increase of the traffic matrix, since $\forall K : \lambda F_K \le C_K$ must hold. The minimum cut in the network indicates the most restrictive bottleneck, Thus the links on the cut are candidates for capacity expansion in order to increase the admissible flow.

In addition, we have implemented a stepwise flow assignment method by allocating a multiple λT of the traffic matrix in each step on shortest paths, where λ is maximized such that the capacity of a link is exhausted in the shortest path design. Further steps increase the flow allocation using constraint shortest path first, where already fully utilized links are avoided. This can be continued whenever the network of all remaining non-exhausted links provides connectivity for all source to destination pairs, i.e. until a fully utilized cut in the network is encountered.
For the latter case, if there are no paths for traffic demands, which cross over the cut in both directions, then the max-flow-min-cut bound and thus a maximum possible flow $\lambda_{\max} T$ is achieved.
Otherwise we check several cases for possible further increase of the flow, but in general it is not difficult to construct examples, where the max-flow-min-cut bound cannot be reached. Then our allocation scheme finishes with several cuts being fully utilized with some paths crossing over each such cut in both directions. Applications to real network traffic scenarios show, that the latter case is relevant but not frequently encountered.
When all traffic demands arise from a single source or lead to a single destination, then the allocation scheme reduces to the well known method by Ford and Fulkerson [1][2], which is always able to construct an optimum LSP design exploiting the max-flow-min-cut bound.

Heuristic flow allocation

Heuristic algorithms represent the most flexible and scalable means for offline optimization when approximate solutions are sufficient. While linear programming presumes that traffic demands can be arbitrarily distributed on multiple paths, we use heuristics under the constraint that each demand must be directed on a unique path [24]. The latter assumption introduces an additional obstacle for load balancing and from a comparison with the optimum linear programming solution we can estimate the significance of possible branching of traffic flows over several paths for the traffic engineering gain. The delivery of each traffic demand t_{ij} on a unique path leads to an NP-hard optimization problem obviously including the NP-complete bin-packing problem already for a balanced assignment of a number of demands on two parallel links [11].

Therefore no scalable exact solution can be expected and we prefer simulated annealing, which is based on the search for paths minimizing the weighted costs for the links in a network-wide view. The cost function is increasing with the load on each link, leading to a tendency to shift paths away from highly utilized links. The paths for traffic demands are assigned sequentially where largest demand first, a random order or other alternatives may be chosen initially. Thereafter, simulated annealing [18][24] is used for stepwise improvement of the solution. Our experience shows that the traffic engineering gain obtained by simulated annealing is usually close to the optimum linear programming solution with less than 5% difference in most cases [15][24].

3 APPLICATION IN IP NETWORK OPERATION

Increasing Internet traffic volume & backbone link speeds

Under normal operation, network dimensioning is focused on periodical traffic peaks. Experience with Deutsche Telekom's Internet platform including broadband access for currently over 5 million ADSL subscribers shows that the series of peak rates observed on backbone links evaluated per day or per week is characterized by moderate short term variation and a steady trend of increasing traffic. The traffic increase is driven by the deployment of broadband access, where peer-to- peer networking presently contributes the prevalent portion of the traffic [14]. The increase rate is varying with the speed of broadband expansion in IP and access platforms and may exceed an often assumed factor 2 per year in some phases [21].

Link capacity in IP backbone networks is provided over optical transmission systems in the framework of synchronous transmission modules (STM) at line speeds of STM-1 ≈ 622 Mbit/s; STM-4 ≈ 2.5 Gbit/s; STM-16 ≈ 10 Gbit/s; STM-64 ≈ 40Gbit/s etc. Each upgrade from STM-n to STM-(n+1) establishes 4-fold capacity. In addition, upgrades to twice the capacity are applied using bundles of two parallel STM-n lines as intermediate step to STM-(n+1).

After an upgrade to twice or 4-fold capacity, the link utilization drops down to half or a quarter when the shortest path first (SPF) principle is applied without traffic redirection. As a consequence, considerable utilization gaps are encountered following each upgrade until traffic growth again approaches the load threshold and triggers the next upgrade [16].

Traffic engineering (TE) on the other hand can fill up the gaps by reoptimization of the path design after each upgrade, such that newly installed bandwidth is immediately used in order to reduce the load on other links [16]. The

required offline optimizations may be carried out using the TE-Scout tool.

Mean utilization for link upgrades with SPF

Traditionally, routing in IP networks is based on the shortest path first (SPF) principle without traffic engineering being applied. Then links can be upgraded independently of each other when the utilization of a link exceeds some threshold T_{Upgrade}, which is determined to guarantee quality of service (QoS) for the supported communication services [13], often including protection against link failures [9]. Two basic assumptions are made for modelling a standard link expansion process on Internet platforms without traffic engineering:

- The link capacity is extended by the same factor F_{Upgrade} ($F_{\text{Upgrade}} > 1$) at each time when the utilization on the link exceeds a predefined threshold T_{Upgrade}.
- Traffic demands are exponentially increasing over time at rate ω. Thus the demand from source v_j to destination v_k at time x is $t_{jk}(x) = t_{jk}(0) \exp(\omega x)$.

For shortest path first routing and without traffic engineering, the analysis of capacity expansion is simplified by the fact, that each link can be considered separately without changing the subset of traffic demands routed over the link under normal operation. Consequently, the utilization periodically increases from $T_{\text{Upgrade}} / F_{\text{Upgrade}}$ to T_{Upgrade} and again drops down immediately after the next upgrade:

$u_{jk}(x) = u_{jk}(0) \exp(\omega x) / (F_{\text{Upgrade}})^{N(x)} \quad (u_{jk}(x) < T_{\text{Upgrade}})$;

$c_{jk}(x) = c_{jk}(0) (F_{\text{Upgrade}})^{N(x)}$;

$N(x)$: number of upgrades during time x.

A period from an upgrade to the next one has duration $D_{\text{Upgrade}} = \ln(F_{\text{Upgrade}})/\omega$. Under both previous assumptions for the standard upgrade rule the same mean utilization u is reached on all links over a considered period:

$$u = \frac{T_{\text{Upgrade}}}{F_{\text{Upgrade}} D_{\text{Upgrade}}} \int_0^{D_{\text{Upgrade}}} \exp(\omega x)\, dx = T_{\text{Upgrade}} \frac{1 - 1/F_{\text{Upgrade}}}{\ln(F_{\text{Upgrade}})}. \quad (1)$$

Note, that the result is independent of the traffic increase rate ω and the time D_{Upgrade} between updates.

A reduction to 72% of the threshold is observed for $F_{\text{Upgrade}} = 2$ and only 54% can be achieved for $F_{\text{Upgrade}} = 4$. On the other hand, the reduction provides a potential for improvement to be exploited by traffic engineering.

4 CAPACITY EXPANSION FOR TRAFFIC ENGINEERING

Traffic Engineering and Quality of Service

When traffic engineering is applied, then link upgrades can no longer be executed for each link in separation and independently of the others. TE makes new installed capacity available to traffic currently flowing on distinct routes. A part of that traffic can be redirected over the additional bandwidth and at the same time reduces the load on other links.

A reoptimization step should be carried out soon after an upgrade or change in the topology to configure an updated optimum LSP path design. We use linear programming as implemented in the TE-Scout tool to balance the load throughout the network, where the optimization goal is to minimize the maximum load on links in the given network topology with regard to the traffic demands. This goal aims at minimizing QoS degradations, which are due to temporary overload.

The statistical multiplexing effect leads to smoothed variability of traffic composed of a large number of flows by subscribers in broadband access platforms. As a consequence, satisfactory QoS conditions are observed in the backbone when a fixed population of traffic sources generates even a considerable mean load up to a threshold in the range of 70%-90%, but QoS performance sharply declines, as soon as the threshold is exceeded. The implications of the utilization level on packet loss and delay as main QoS characteristics have been investigated for switching systems e.g. in [13].

Network expansion under normal operation

In order to perform capacity upgrades together with TE, a network-wide expansion strategy has to be established. A maximum admissible traffic flow in terms of a multiple $\lambda_{\max} \cdot T$ of the traffic matrix T can only be achieved, when all link capacities on a cut through the network are exhausted [13]. There may be further fully loaded links in an optimum LSP path design, but the admissible flow can be increased only by adding capacity on the link of the cut or, if there are several cuts then the capacity of all of them needs to be expanded.

The upgrading strategy can make use of the TE optimization methods by computing the increase in admissible flow $\lambda_{\max} \cdot T$ for all cases of single link upgrades. We proceed in this way and choose to upgrade a link to twice the capacity with a maximum ratio of the increase in the admissible flow to the capacity to be added. In this way, the mean utilization is maximized for the next upgrade step, although this is not guaranteed to be optimal over several steps. The effect of a combination of several upgrades may also be tested at high computational effort. These tests are included in our strategy, when there are several disjoint cuts through the network whose links have approached the threshold in an optimum LSP design. Then multiple link upgrades are required to improve the admissible flow.

The achievable long term throughput and corresponding mean utilization for traffic engineering with the previous upgrade strategy has been evaluated in [16] for several examples, where a relative improvement of more than 20% was observed in comparison to mean utilization of 72% for SPF as derived in Section 3 when upgrades are installing twice the previous capacity. As another positive effect, the

variation over time of the TE utilization curve is essentially smaller than for SPF.

Network expansion including failures scenarios

The path design allowing for a maximum throughput with balanced load can also be calculated for failure scenarios like a breakdown of a link, node or an arbitrary set of network resources. The optimization algorithm can be applied to any considered failure case. In particular, we define a set of relevant failure scenarios, which usually will include the most probable ones, e.g. all single link failures are considered in the following examples. Our aim is to secure the network against the set of failures, such that all traffic demands can be carried through the network in each relevant failure case [20].

Therefore we start from a predefined network topology, which may correspond to an available infrastructure of optical fiber links. Then we compute the maximum admissible traffic throughput $\lambda_{\max} T$ in terms of a linear scaling of the traffic matrix T for

- normal operation without failure and
- for all relevant failure scenarios,

where different path designs are computed to adapt balanced load solutions to each case separately. We take the smallest of the maximum admissible throughput results, i.e. the worst case, as the admissible throughput with regard to the set of failures.

Next, we start to upgrade the link capacities in order to increase the latter throughput. Therefore the failure case(s) with most impact on the throughput have to be considered. We take those failures into account, which yield less than 110% of the admissible throughput for the worst case.

Then we again test upgrades of each link regarding the improvement of admissible throughput for the considered failures. The improvement of an upgrade with regard to a failure is measured by the difference of the TE optimization results with and without the link upgrade, where an improvement beyond 110% of the worst situation is not accounted for.

In order to define a corresponding score function for comparison of link upgrades, let $E^* = (e_{jk} \mid j,k \in V;\ cap^*(e_{jk}) > 0)$ denote the set of links of the current network topology, each of which is equipped with a positive capacity cap^*: $E^* \rightarrow \Re^+$. $\lambda_{\max}(cap^*)$ is the maximum scaling factor, such that the throughput $\lambda_{\max} T$ can be transported as a multi commodity flow in the network assuming a linear growth of the traffic demand matrix T.

Let $F = \{F_1, ,\ldots, F_N\}$ be the set of failure scenarios to be taken into account, where each F_m is the set of remaining links and cap_m the capacity assignment for a failure scenario. In addition, let $cap_{(j,k),,m}$ denote a capacity assignment after a single link (j, k) has been upgraded for the topology corresponding to the failure scenario F_m, i.e.

$$cap_{(j,k),m}(e_{st}) = \begin{cases} cap^*(e_{st}) & \text{if } (s,t) \in F_m \text{ and } (s,t) \neq (j,k) \\ 2\,cap^*(e_{st}) & \text{if } (s,t) \in F_m \text{ and } (s,t) = (j,k) \\ 0 & \text{if } (s,t) \notin F_m \end{cases} \quad (2)$$

Then the maximum throughput $\lambda_{\max(}(cap_{(j,k),,m})$ is evaluated for all upgrades of links (j,k) and with regard to a subset of failure scenarios with most impact on the throughput, i.e. $\forall m$: $\lambda_{\max}(cap_m) < 1.1\ \lambda_{\max(}(F)$, where $\lambda_{\max(}(F) = \min_{\{m = 1, \ldots, N\}} (\lambda_{\max(}(cap_m))$ denotes the minimum for the worst case(s).

Then the score function s^*: $E^* \rightarrow \Re$ for link upgrades is obtained by the sum of improvements of the throughput regarding the subset of the most critical failure scenarios with regard to the ratio to the capacity, which has to be installed:

$$s^*(j,k) = \sum_{m=1}^{N} \max\left(0, \frac{\min(\lambda_{\max}(cap_{(j,k),m}), 1.1 \cdot \lambda_{\max}(F))}{\lambda_{\max}(cap_m)} - 1\right) / cap^*(e_{jk}). \quad (3)$$

A link (j, k) yielding a maximum score $s^*(j, k)$ is selected for the next upgrade step. In this way

- a choice for the next upgrade is preferred which has an effect on some of the most critical breakdowns
- and the number of relevant scenarios to compute the score function in the selection process is often small and thus favourable for the computation effort.

Optimization for TE load balancing is carried out by linear programming, where a single optimization requires less than half a second on a 1GHz personal computer for the following example networks with 18 nodes and 32 edges. Nevertheless, the computation of s^* and thus the decision about the next upgrade step may involve hundred tests of variants of the topology for different cases of failed links and extended capacities.

We have carried out a series of link upgrades with the next candidate being selected by the previously introduced score function establishing a network expansion strategy with regard to failures. Results are evaluated for the logical Sprint IP backbone topology consisting of 18 nodes and 32 bi-directional links [9], which is shown in Figure 1. In this example, each link upgrade is assumed to double the capacity bi-directionally, although in general, both transmission directions of a link may be treated separately. Initially, the same capacity is assumed for each link in both directions, i.e. all STM-1 links. The traffic matrix is given by

$$t_{ij} = \begin{cases} 2p & if\ (i \notin V_0\ and\ j \notin V_0)\ \ or\ \ i = j; \\ 1 & if\ ((i \in V_0\ and\ j \notin V_0)\ \ or \\ & \quad (i \notin V_0\ and\ j \in V_0))\ \ and\ \ i \neq j; \\ 2(1-p) & if\ (i \in V_0\ and\ j \in V_0)\ \ \ and\ \ i \neq j; \end{cases}$$

where V_0 is a set of 9 nodes in the eastern part of figure 1 and $p = 0.7$. The traffic matrix models different busy hours, which are relevant in the USA for the East ($p = 0$) and West coast ($p = 1$) as well as a sliding linear transition between both cases for $0 < p < 1$. It has been used for examples with varying p in normal operation in [17].

Figure 1 depicts an intermediate stage of the network expansion process, after 18 links have been upgraded to twice the initial capacity, which are printed as bold arcs.

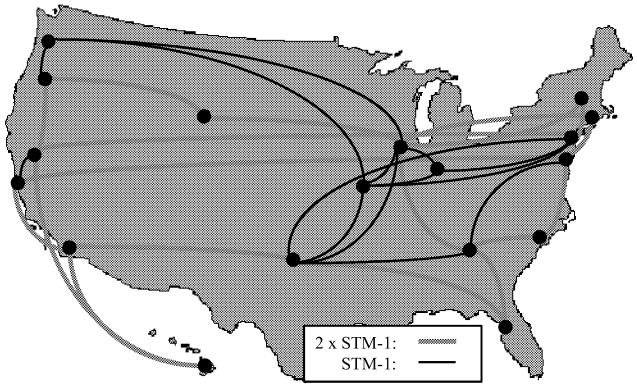


Figure 1: Example of the logical Sprint IP backbone topology [9] after 18 single link upgrade steps

As a measure of the efficiency of load balancing we introduce the effective mean load l. Let $d_{\min}(j, k)$ denote the smallest possible number of edges on a path from node v_j to v_k and $d(j, k)$ the mean number of edges on the paths to deliver the traffic flow demand t_{jk} in the optimized path design for load balancing. When portions of t_{jk} are transported over different paths by a linear programming solution, then the effective load is computed by the ratio of the sum of the maximum admissible traffic intensities $\lambda_{\max} T$ from source to destination with each of them being weighted by $d_{\min}(j, k)$ and the sum of link capacities $cap(e)$:

$$l \overset{def}{=} \sum_{j,k=1}^{N} d_{\min}(j,k)\, \lambda_{\max} t_{jk} \,/\, \sum_{e \in E} cap(e) \qquad (4)$$

The effective mean load is slightly smaller than the mean load as defined by the ratio of the sum of the flows on the links to the sum of the link capacities, or equivalently, by replacing $d_{\min}(j, k)$ with $d(j, k)$ in (4). When we compare a path design, which directs all traffic demands t_{jk} on paths with a minimum number $d_{\min}(j, k)$ of edges with another design involving detours over more than $d_{\min}(j, k)$ edges for some demands, then the latter design generates a higher mean load, even if the admissible throughput is the same. Thus considering the mean load does not yield a fair measure of the efficiency with a handicap for SPF, whereas the effective mean load defined in (4) only accounts for the admissible throughput $\lambda_{\max} T$ independently of the length of the paths [16].

Considering a single upgrade step, the effective mean load l decreases due to increasing capacity in the denominator, if the factor $\lambda_{\max}$ for the admissible traffic remains unchanged. An increase of l is observed each time when an improvement of $\lambda_{\max}$ more than compensates for the additional capacity.

Figure 2 shows the effective mean load after each of the first 100 single link upgrades for several examples, where the lowest curve corresponds to the considered case. The curve shows some series of up to 10 successive upgrade steps without improving $\lambda_{\max}$ followed by single upgrades yielding essential improvement. Often several links have to be upgraded together until an increase of the admissible flow is encountered.

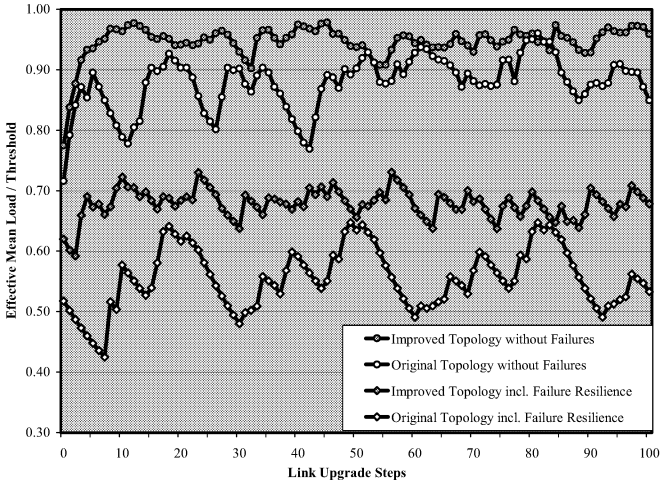


Figure 2: Resource utilization during a process of stepwise capacity upgrades

When we observe upgrade processes with the given upgrade strategy over long periods, then the capacity assignment adapts to the demands in the traffic matrix during an initial phase. Thereafter upgrades are carried out almost periodically, such that a subset of the network links is upgraded to double capacity once per period. Approximate periodicity can already be observed in some case studies of Figure 2 prior to 100 upgrade steps. In the considered example, the effective mean load approaches about 56% over a periodic upgrade cycle. A maximum at 66% is reached for the capacity assignment of Figure 1 after 18 upgrades and later on for analogous cases with twice, 4-fold etc. capacity assignment.

For comparison, we also evaluated the effective mean load in normal operation for the same topology, traffic matrix and initial capacity assignment but without regarding link failures. Therefore an entirely different upgrade strategy always prefers an upgrade to a maximum possible effective mean load in the next step, which has been studied in more detail by [16]. The corresponding curve for the "Original topology without failure" in figure 1 achieves an effective mean load of about 89% of the threshold. This means, that the additional capacity, which has to be spent for circumventing single link errors amounts to a factor of about 89% / 56% ≈ 1.6.

Next we compare these results to the mean load for shortest path routing, which allows each link to be upgraded separately and independently of the rest of the network. Under normal operation, a link is upgraded when the load approached the threshold. For upgrades to twice the capacity, this leads to slightly more than 72% mean load relative to the threshold over an upgrade period as approved by equation (1) at the end of Section 3 [16]. In order to evaluate SPF with regard to single link failures, we compute for

each link the ratio of the load on the link in normal operation to the maximum load being observed on the link when another links has been removed from the network. In order to protect against the worst case yielding the maximum load on the link, we have to reduce the upgrade threshold by the previously computed ratio. Therefore the mean load on the link is further reduced below 72% by the considered ratio to the achievable mean load in normal operation. In this way, we evaluated the mean load on each link and for the network in total, yielding 42% in the example.

The example of the logical topology of the Sprint network in Figure 1 has some nodes connected by only two links to the rest of the network, which obviously is a handicap with regard to failure resilience. Therefore we next consider a modified topology, where three links have been added in order to achieve a connectivity degree of 3, such that there are three link disjoint paths from each source to destination in the network. On the other hand, we removed two links, which have not been upgraded in the previous example using our strategy for normal operation [16].

We have analyzed the upgrade process for the modified topology in the same way as for the original one, with the same traffic matrix and initial assignment of one unit (STM-1) of capacity on each link. Figure 3 shows the topology with capacity assignment after 23 upgrade steps. The corresponding curve "Improved topology incl. failure resilience" in Figure 2 shows a maximum of 73% effective mean load at this stage. Figure 4 shows the capacity assignment after 174 upgrade steps, when the maximum of 73% is reached a second time but with an entirely different capacity assignment. The long term effective mean load approaches 68% for the improved topology. This confirms that the efficiency of traffic engineering depends on the network topology and improves for a higher degree of connectivity.

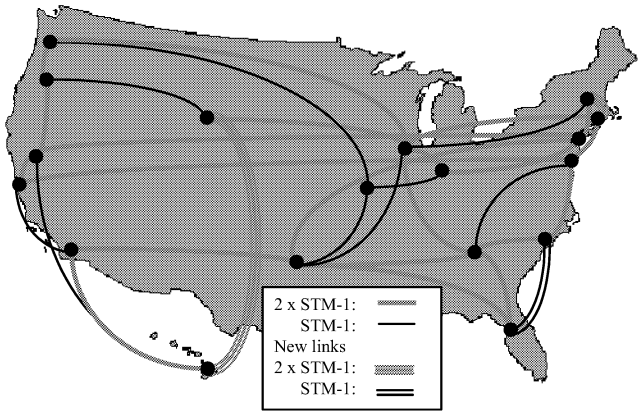


Figure 3: Modified network topology for connectivity degree of 3 after 23 link upgrades

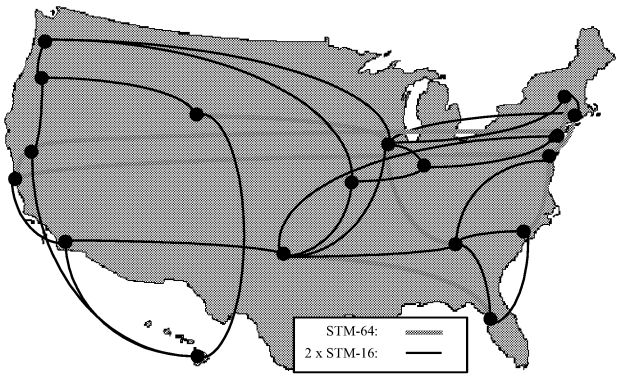


Figure 4: Modified network topology for connectivity degree of 3 after 174 upgrades

Table 1: Long term mean effective load during a link upgrade process	Original topology (Figure 1)	Modified topology (Figure 3-4)
Traffic Engineering (TE) for normal operation	89 %	94 %
Shortest path routing (SPF) for normal operation	72 %	72 %
Traffic Engineering (TE) including failure resilience	56 %	68 %
Shortest path routing (SPF) including failure resilience	42 %	47 %

5 CONCLUSIONS

We investigated strategies for upgrading the installed capacity in telecommunication networks in order to obtain an optimum throughput and resource utilization with regard to single link failures or in general, a predefined set of failure situations. Therefore a tool for offline traffic engineering in MPLS based IP networks is used to compute optimized load balancing solutions for a predefined network and traffic matrix. While exact analysis of the alternatives for the next upgrade step is carried out, the upgrade strategy introduces a heuristic behaviour over several steps and for the long term development.

After an initial phase, the process of extending the link capacities usually adapts to the demands for source to destination flows in traffic matrix and often comes close to a periodical upgrade cycle after some time. The best results in intermediate stages are useful for an appropriate resource provisioning. The results also give hints on the relevance of each link for the achievable throughput with implications for possible improvement in the network topology, where a higher degree of connectivity leads to improved efficiency of traffic engineering in case studies. Alternative strategies are for further study as well as the consideration of more

detailed cost functions extending the main criterion of optimum resource utilization.

ACKNOWLEDGEMENT

We would like to thank Hanna Aliakseyeva for her support in evaluating the results.

REFERENCES

[1] R.K. Ahuja, T.L. Magnanti and J.B. Orlin, Network flows: Theory, algorithms and applications, Prentice Hall (1993)

[2] V.K. Balakrishnan, Network optimization, Chapman & Hall, London (1995)

[3] N. Benameur and J.W. Roberts, Traffic matrix interference in IP networks, Proc. NETWORKS, Munich (2002) 151-156

[4] I. Berry et al., An integrated GA-LP Approach to Communication Network Design, Telecommunication Systems 12 (1999) 265-280

[5] A. Elwalid, C. Lin, S. Low and I. Widjaja, MATE: MPLS Adaptive Traffic Engineering, Proc. IEEE INFOCOM (2001)

[6] A. Feldmann, A. Greenberg, C. Lund, N. Reingold and J. Rexford, NetScope: Traffic engineering for IP networks, IEEE Network, special issue on Internet traffic engineering, Ed. Z. Wang (March 2000) 11-19

[7] B. Fortz and M. Thorup, Optimizing OSPF/IS-IS weights in a changing world, IEEE JSAC (2002) 756-766

[8] B. Fortz, J. Rexroth and M. Thorup, Traffic engineering with traditional IP routing protocols, IEEE Comm. Magaz. (Oct. 2002) 118-124

[9] Fraleigh et al., Packet-level traffic measurements from the Sprint IP backbone, IEEE Network (Nov./Dez. 2003) 6-16

[10] M. Franzke and A. Pönitz, Global shortest path solutions for the traffic engineering problem, Proc. Networks'02 Conf., Munich (2002) 275-278

[11] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, NewYork (1979)

[12] B. Gavish, Topological Design of Communication Networks: The overall Design Problem, Europ. J. Oper. Res. 58 (1992) 149-172

[13] G. Haßlinger, Quality-of-service analysis for statistical multiplexing with Gaussian distributed and autoregressive input modeling, Telecommunication Systems 16 (2001) 315-336

[14] G. Haßlinger: Peer-to-peer Networking: Main aspects and conclusions from the view of Internet service providers, WSEAS Transactions on Computers, Vol. 3/3 (2004) 587-591

[15] G. Haßlinger and S. Schnitter, Optimized load distribution in MPLS networks, Telecommunications Network Design and Management, Kluwer Akad. Publ. (2002) 125-142

[16] G. Haßlinger and S. Schnitter: IP Network expansion for growing traffic demands with shortest path first routing and traffic engineering, Proc. 11th Internat. Network Strategy and Planning Symposium, Networks'04, Vienna, Austria (2004) 81-86

[17] G. Haßlinger, S. Schnitter and M. Franzke: TE-Scout: A Traffic Engineering Tool for Optimised Path Design in IP Service Provider Networks, Proc. 12th Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB/PGTS), Dresden (2004) 35-44

[18] M. Kolonko, Some new results on simulated annealing applied to the job shop scheduling problem, Europ. J. Operational Research, 113 (1999) 123-136

[19] M. Menth, J. Milbrandt and S. Kopf, Capacity assignment for NAC budgets in resilient networks, Proc. 11th Internat. Network Strategy and Planning Symposium, Vienna, Austria (2004) 193-198

[20] M. Menth, A. Reifert, and J. Milbrandt, Self-Protecting Multipaths - A Simple and Resource-Efficient Protection Switching Mechanism for MPLS Networks, Proceedings of the 3. IFIP-TC6 Networking Conference Networking'04, Athens, Greece (2004)

[21] A. Odlyzko, Internet traffic growth: Sources and implications, Proc. SPIE Vol. 5247 (2003) 1-15

[22] M. Pioro and D. Mehdi, Routing, Flow and Capacity Design in Communication and Computer Networks, Morgan Kaufmann Publ. (2004)

[23] E. Rosen, A. Viswanathan and R. Callon, Multiprotocol label switching architecture, RFC 3031, Proposed IETF standard (2001)

[24] S. Schnitter and G. Haßlinger, Heuristic Solutions to the LSP-design for MPLS Traffic Engineering, Proc. Networks'02 Conf., Munich (2002) 269-273

[25] S. Schnitter and M. Horneffer, Traffic Matrices for MPLS Networks with LDP Statistics, Proc. 11th Internat. Network Strategy and Planning Symposium, Vienna, Austria (2004) 231-236

[26] R.D. Wollmer, Optimal investment in capacity expansion of arcs in stochastic multi commodity network flows, Proc. 6. INFORMS Telecom. Conf. WA1.4, Boca Raton (2002)

CR-LDP Usage for setting up a LSP with bandwidth guaranteed in a SIP over MPLS network

C.Zhang, C.G.Guy
Department of Electronic Engineering
The University of Reading, Reading, RG6 6AY, UK
Tel.: +44-118-378-8757, Fax. +44-118-378-8583
E-mail: Chen.Zhang@reading.ac.uk , C.G.Guy@reading.ac.uk

Abstract: The next generation communication systems will provide multimedia services and need to support high quality of service (QoS) in a more flexible and intelligent manner than previous generations. In order to provide different QoS to meet different user demands, we have proposed a SIP (Session Initiation Protocol) over MPLS (Multi-Protocol Label Switching) network. There are two entities to be presented: TE-SIP (Traffic Engineering-SIP) Server in the Access Network and a Local Decision Server in the Core Network. In this paper, we present a method using CR-LDP (Constraint Routed Label Distribution Protocol) to set up a LSP (Label Switching Path) with bandwidth guaranteed for providing QoS.

1. Introduction

The future network will have to handle a huge amount of IP data traffic, including a significant amount of real-time traffic with demands for assured QoS. The future telecommunication network is likely to be based on an all-IP network, so providing QoS in the Internet will play an important role.

There are many problems with using IP networks to carry voice or multimedia traffic [Veltri and Salsano 2002]. To overcome these problems, the IETF (Internet Engineering Task Force) has proposed some protocols. [Zhang and Guy 2002] Proposed a SIP over MPLS network to provide QoS for multimedia traffic flow, which chooses SIP as a call handling protocol and an MPLS-based network as the core network to provide QoS.

The SIP [Rosenberg et al 2002] protocol supports establishing and terminating multimedia communication sessions on the Internet, so SIP can be used for a broad range of types of communications, such as Internet telephony, multimedia conferences, and distance learning, on both wired and wireless networks. SIP is independent of the lower-layer transport protocol, which allows it to take advantage of new transport protocols. Compared with H.323 which was published by ITU, SIP is simple and developable. There are many discussions which have been published about the advantages and disadvantages of H.323 and SIP. MPLS [Rosen et al 2001] is an advanced forwarding scheme and an MPLS- based network is a high-speed network. It intends to provide QoS using traffic engineering, which is a procedure of selecting the most appropriate path. MPLS-TE is a procedure of selecting LSP for different QoS requests.
How to set up the LSP is a big issue. [Zhang and Guy 2003] have presented a method RSVP-TE (Resource Reservation Protocol with Traffic Engineering extensions). In this paper, we present another method using CR-LDP to set up the LSP (Label Switching Path) with bandwidth guaranteed for providing QoS. The proposed network architecture is discussed in Section 2; Section 3 describes the TE-SIP server in the Access Network and GWLSR in the MPLS network. Then, the procedure of using CR-LDP to set up LSP (Label Switching Path) with bandwidth guaranteed in a SIP over MPLS network is presented. Conclusions are given in Section 5.

2. Network Architecture of the SIP over MPLS based network

Figure 1 shows the Network architecture for a SIP over MPLS based network [Zhang and Guy 2002]. This Architecture is divided into three sub-networks: SIP client, Access Network and Core Network.

- The SIP client: SIP client is a SIP phone or a PC, which can require SIP service.
- Access network domain: is an Internet Service Provider (ISP), which can provide SIP service. When the SIP client starts a SIP call-setup session, the Access Network will provide call control functions and signaling gateway functions. The access network is a bridge between the SIP client and the core network.
- Core network domain: is an MPLS based network. There are two node types: The Gateway LSR (GWLSR) and Label Switch Routers (LSR). The MPLS network is connected with other networks, for example PSTN, other MPLS network, DiffServ network, etc.

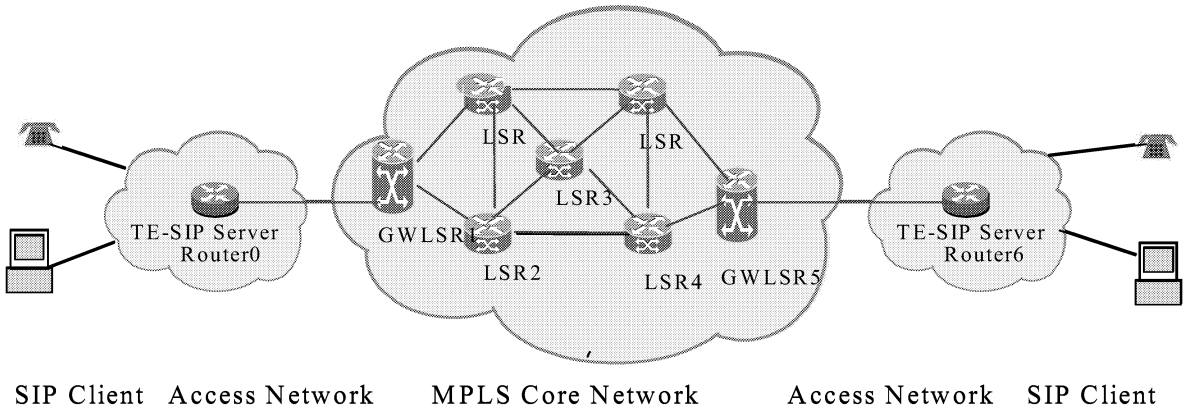


Figure1 Network architecture of SIP over MPLS based network

3. TE-SIP server and GWLSR

Figure 2 shows the TE-Server and GWLSR internal Architectures [Zhang and Guy 2003]. The TE-SIP server is in the access network and the GWLSR is on the border of the MPLS network. When the client sends a SIP message to its proxy server, it receives a reply from its server. If the proxy server can deal with TE information with GWLSR, this proxy server is a Traffic Engineering enabled SIP Server (TE-SIP server). The message is sent to the destination proxy server of the access network. If this proxy server can deal with TE information, the call path will be set up. Otherwise, it will setup a normal call. Meanwhile, the TE-SIP server will communicate with the GWLSR for requested TE.

The main function of the GWLSR is making the decision to choose which LSP to forward the traffic flow over. [Zhang and Guy 2003] have proposed a local decision server and a routing algorithm for GWLSR. The local decision server decides the routing forwarding, which includes explicit route and resource reservation, according to the different QoS user demands and the network attributes that includes link state database, network topology and resource available.

The following section describes CR-LDP to setup LSP with bandwidth reservation for forwarding traffic flow.

4. CR-LDP to setup LSP with bandwidth reservation

CR-LDP (Constraint Routed Label Distribution Protocol) specifies a mechanism to distribute label, set up, and maintain LSPs along explicitly routed or constraint-routed paths. CR-LDP extends LDP by adding some functionality [Jamoussi et al 2001] . Our proposed network intends to provide QoS for real time multimedia traffic flow. How to provide the end-to-end bandwidth guaranteed LSPs for traffic engineering is the most important question. We have studied the RSVP-TE to set LSP with bandwidth guaranteed to provide QoS management, and in this section we will discuss using CR-LDP to set LSP to provide QoS.

When the caller TE-SIP server receives the 200 OK message with the complete TE session indicators, it completes the TE session setup by performing the TE request to the GWLSR on the caller access network for the caller-to-callee traffic flow. The call set-up protocol is described in [Jamoussi et al 2001]. The GWLSR is the edge point between Access and Core Network. According to the routing table principle, GWLSR will forward the traffic flow into MPLS Network.

In order to establish the single-LSP traffic flow over MPLS bearer structure, the procedure is as follows:

1. A bi-directional LSP is created either by using an MPLS control protocol (CR-LDP).
2. As voice or audio connections arrive at the GWLSR, the traffic flow will be assigned to the label, and then according to MPLS forwarding mechanism, it will be forward to destination LSR.

When the GWLSR (ingress) receives the TE QoS request from ER of Access Network, it will generate a CR-LDP label request message, which contains an ER-TLV (Edge Router Tape-Length- Value) containing the explicitly routed path. The ER-TLV contains a few ER-HOP TLVs, which is a list of nodes that could potentially be used to satisfy the constraint routing.

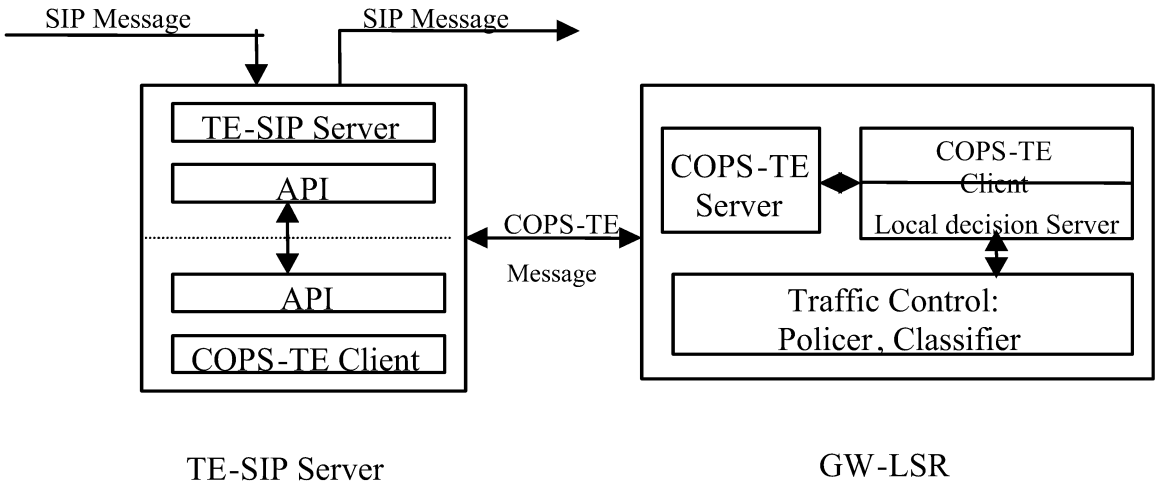


Figure2 TE-SIP Server and GWLSR internal architecture

GWLSR will forward the CR-LDP label request message with ER-HOP TVLs intact to next LSR mentioned in the explicit route. The LSR removes the topmost ER-TLV and forwards the rest to the next hop. When the label request message reaches the GWLSR (egress), the egress sends a CR-LDP label mapping message back to the source along the same path traversed by the label request message.

To satisfy the bandwidth reservations for the real-time traffic, the PDR (Peak Data Rate) and PBS (Peak Burst Size) parameters of the traffic parameter TLV are specified and sent along with the label request message.

The CDR (Committed Data Rate) and CBS (Committed Burst Size) parameters are set equal to those values of PDS and PBS. The GWLSR also configures the QoS control modules in ingress and egress for admission control. This configuration drops packets that exceed the PDR and PBS specification. Negotiation flags for these parameters are generally set so that the intermediate LSRs and GWLSR can increment or decrement the PDR and PBS values depending on the resources in the core network. The negotiation parameter values are sent back in the CR-LDP label mapping message to make the GWLSR and LSRs adjust the domain resources accordingly.

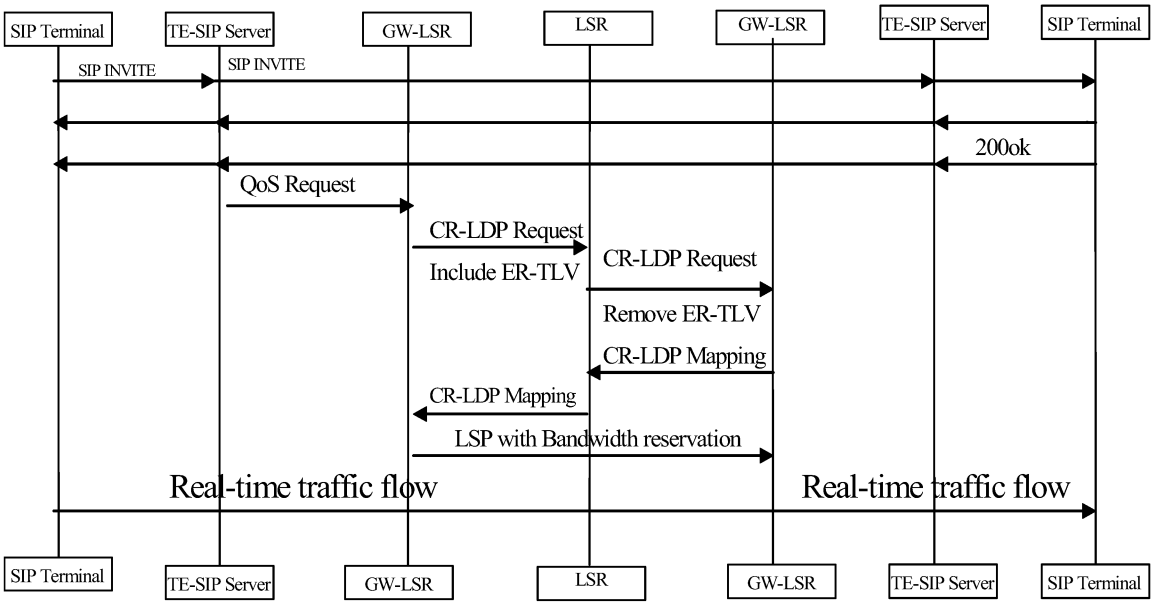


Figure 3 the procedure of setting up LSP with bandwidth reservation

After using CR-LDP to set up a LSP with bandwidth guaranteed, the real-time traffic flow can be sent from caller to the callee.

The procedure of setting up LSP with bandwidth reservation is shown in the figure 3.

5. Conclusion and further work

In this paper, we propose a network architecture to provide Traffic Engineering in a SIP over MPLS based Network. This model adopts the SIP protocol as signaling protocol to set up a call and MPLS as a core network forwarding mechanism. A TE-SIP server is developed to handle the combination of SIP call setup signaling and TE information at Access Network. It communicates with GWLSR, which will provide the TE action for different users. We present a method using CR-LDP to setup a LSP with bandwidth reservation, it can be used to provide QoS to different traffic flow according to demands.

Further work will concentrate on network simulation and improving the network architecture. Network simulation will demonstrate our proposed network attributes and then can improve the network architecture according to the simulation results.

6. References

Jamoussi B, L. Andersson R. Callon R. Dantu L. Wu P. Doolan N. Feldman A. Fredette M. Girish E. Gray J. Heinanen T. Kilty A. Malis, .Constraint-based LSP Setup using LDP, work in progress, draft-ietf-mplscr-ldp-05, March 2001.

Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, Sparks, M. Handley, E. Schooler "SIP: Session Initiation Protocol" RFC3261,IETF Network working group, June 2002

Rosen E, A. Viswanathan, and R. Callon "MPLS Architecture ", RFC3031, IETF Network Group Jan.2001,

Veltri L and S. Salsano, "SIP Extensions for QoS support in DiffServ Networks" IETF Internet Drafts<draft-veltri-sip-qsip-00.txt>, April 2002, work in progress

Zhang C, C.G.Guy, "A Framework for Traffic Engineering In a SIP-over-MPLS based network", LCS2002, London, Sept.2002, ISBN: 0-9538863-2-8, pp129-132

Zhang C, C.G.Guy, "TE-SIP Server Design for a SIP-over-MPLS based network" ICCT2003, Beijing, April 2003, ISBN: 7-5635-0686-1, pp1758-1761

Zhang C, C.G.Guy "Routing Algorithm at Local Decision Server at GWLSR of the SIP over MPLS based Network", HETNET2003, Ilkley, UK, July 2003, ISBN: 0-9540151-3-4, pp14/1-14/10

Zhang C, C.G.Guy "COPS Usage for QoS management between TE-SIP server and GWLSR over a SIP over MPLS network", ITC-CSCC2004, Tohoku University, Japan, 5-8 of July, 2004

MOBILE COMMUNICATIONS

SPECIFICATION AND DESIGN OF AN ADVANCED HOT SPOT CONTROLLER BASED ON RADIUS AUTHENTICATION SERVER

Panayiotis S. Alefragis, Nikolaos S. Voros
Department of Applied Informatics in Management and Finance
Technological Educational Institution of Messolonghi
Nea Ktiria, Messolonghi 30200, Greece
E-mail: {alefrag, voros} @ teimes.gr

Vassilis D. Triantafillou
Department of Telecommunication Systems and Networks
Technological Educational Institution of Messolonghi
Nea Ktiria, Messolonghi 30200, Greece
E-mail: triantaf @ teimes.gr

Konstantinos E. Tsatsaronis
ATMEL Hellas S.A.
ATMEL MMC Athens Branch
48 Papazachariou str.
Agios. Dimitrios, 173 42, Greece
E-mail: ktsatsaronis@athens.atmel.com

KEYWORDS

Hotspot controller, authentication, wireless LAN infrastructure

ABSTRACT

The widespread use of wireless communications and the international trend to integrate them with traditional computer networks, has leveraged the creation of wireless access points for mobile users, commonly know as hot spots. This paper provides an insight of the design details encountered during the design of a wireless hotspot controller that allows user authentication for the mobile stations that move within the area of the hotspot. The architecture presented relies on RADIUS authentication server and has been tested in various real world environments.

INTRODUCTION

The sudden growth of communications and mobile devices in the early decade of this century has made a huge blast to networks and data exchange mobility. The need of users to have access to their corporate network or to the Internet, from any place, has driven vendors to produce new products especially for that purpose.

This paper presents the intrinsic features that must be taken into account for the design of publicly available Wireless Local Area Networks (WLANs), commonly known as Hotspots, in a standalone product solution. The hotspot architecture presented is based on RADIUS server and offers a low-cost and modular solution for the implementation of a hotspot controller.

The rest of the paper is organized as follows: Section 2 describes rationale of the idea presented in the rest of the paper, while in parallel provides a brief overview of existing design attempts. In section 3 the proposed architecture, along with the main design issues encountered, is described. Finally, the paper concludes by referring to the potential extensions of the proposed hot spot controller architecture.

RATIONALE

The concept of implementing a hotspot controller in a single standalone product that can have multiple uses is a desired market solution

for wireless hotspots. An important advantage of having such a product is that it provides a simple, cost effective and reliable interconnection platform.

Most access point gateways available in the market will support the basic functionality of a small standalone hotspot (Ziakas D. 2003). A basic web interface is usually used to interconnect the standard for user Authentication, Authorisation and Accounting (AAA) server(s) (Hassell J. 2002).

The key concept introduced in this paper is the support of different payment methods allowing users the flexibility to choose the way of payment they are keen on. In addition, a connection to a central payment server is provided in order to process payment requests through a secure connection for each client machine.

THE PROPOSED ARCHITECTURE

A wireless LAN hotspot is usually designed to be *public*. It is usually positioned in a public place with itinerant human traffic, for example a railway station, an airport or a cafeteria. This is in contrast with a private WLAN, which is usually designed to be exclusively available to a user group belonging to a particular company or institution and not to the general public. However, many key design points for hotspots are in common to the deployment of both public hotspots and private WLANs. Figure 1 outlines the network topology of a small hotspot that allows users to connect to it once the network has been detected.

Hotspot Controller Specification

A RADIUS Hotspot Controller is designed for use in a TCP/IP network environment and makes use of IP networks for the communication of the various devices with a RADIUS Authentication Server. In Figure 1 it can be seen that the authentication and billing servers are not located in the same area of the hotspot controller. Each request must be routed-through an Internet connection in order to be processed by the AAA server(s).

Considering the deployed security of the authentication server, it must at least ensure that security features such as the Extensible Authentication Protocol (EAP is a protocol used between a user station and an authenticator or authentication server) over RADIUS (RFC 2869 2004) are implemented. Moreover, Robust Secure Network (RSN) is also needed in order to enhance security (Edney J., Arbaugh W. 2004). All the aforementioned requirements are supported by RADIUS Authentication Server (Hassell J. 2002).

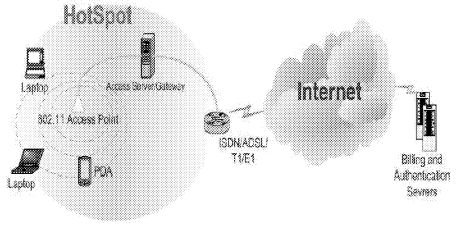


Figure 1: A small Hotspot network topology

It is also very important to have support of vendor-specific attributes, both from the hotspot controller and the RADIUS Authentication Server, in order to serve more specialized features. Such a proprietary feature is for example Microsoft's Challenge Handshake Authentication Protocol (MS-CHAP) authentication method (RFC 2548 2004; RFC 2759 2004). Support of this authentication method enables users to establish remote PPP sessions without being necessary to configure an authentication method on the client.

Hotspot Controller Design

The hotspot controller developed has two main APIs to communicate with the Radius Authentication Server and the end user application. The second provides a command line like interface functionality, called Login

Process API (LPA) (see Figure 2), and it is indented for use by the end users of the system; the first one is designed for allowing the communication with the RADIUS Authentication Server called the RADIUS Server API (RSA part of Figure 2).

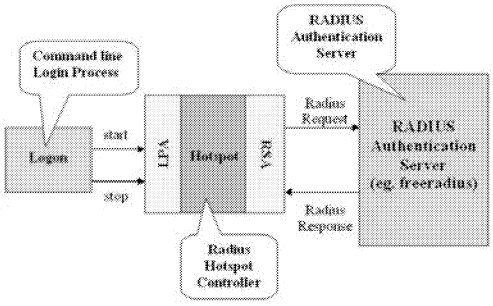


Figure 2: A brief description the Hotspot Controller Design

Figure 2 illustrates the interaction of a user when he/she accesses the hotspot area for the first time. A user starts (or stops) sessions from a command line interface providing his credentials needed in order his requests become *trusted*. The LPA provides the appropriate information about the user to the main part of the hotspot controller using datagram sockets. The RSA establishes the connection(s) with the RADIUS Authentication Server(s) by opening an internet socket with the RADIUS Server to allow the transactions, through port 1812 for authentication or 1813 for accounting (Hassell J. 2002; RFC 2865 2004; RFC 2866 2004).

The architecture of the hotspot controller is outlined in Figure 3 and its constituent parts include:

- *LPA* - The API responsible for the interaction between the user and the hotspot controller.
- *Hotspot Controller* - The main part responsible for scheduling the events, listing the attributes and packaging the packets to be sent.
- *RSA* - The API responsible for the communication with the RADIUS Authentication and Accounting Server(s).

Upon the user initiation of a session from the login command line (described as "Logon" in Figure 2) the session handler (as shown in Figure 3) triggers the timer to record the timestamp and schedules an event. The session handler (see Figure 4) sends the type of the attribute (authentication or accounting) along with the number of the attributes sent from the login process (username, password etc), to the attribute handler in order to commence the construction of the attribute value pair lists. As soon as the lists are created, the packet handler forms the information in RADIUS packets according to the types of information that will to be sent (authentication or accounting) and establishes a connection with the RADIUS server. The RADIUS Server API manages to send the requests to the RADIUS authentication or accounting server(s) and schedules an event.

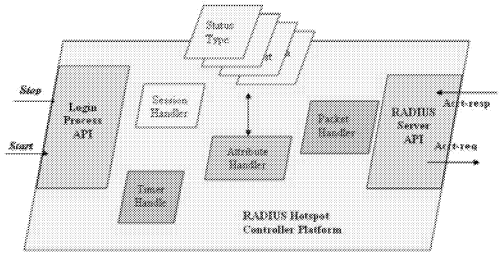


Figure 3: Radius Hotspot Controller Platform

The timer schedules events on a specific interim that is determined at the beginning of the execution as an argument - the minimum value is 60 seconds (RFC 2867 2004) - and waits for a response packet matching the requested session id. If there is no response or a response timeout from the RADIUS Server occurs, a backup server from the configuration file is selected and the same procedure is replicated again with the same session id that the primary request used.

If a response package arrives from the RADIUS server, the RADIUS Server API transfers the packet to the packet handler. The packet handler examines the header of the packet and extracts its content. If the packet contains information that denies the access to the user, the session handler is trigged, stops the session and deletes it from the session list. If the packet contains an Auth-Accept or Acct-Accept header (Valid Authentication/ Accounting response from the RADIUS Server), the packet handler sends its data to the attribute handler, where it creates the attribute value pair list with the attributes needed for the specific session (accounting or/and authentication attributes).

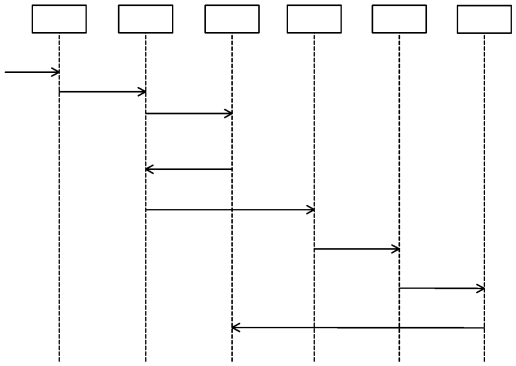


Figure 4: The interaction between the user and the AAA server(s)

Consequently, the session handler sends an acknowledgment to the device driver in order to permit the exchange of data between the user's device and the requested destination device. The session handler takes a new timestamp from the timer, schedules a new event and creates a new interim session event. A new attribute value pair list is created by the attribute handler, and is sent the packet handler. The packet handler sends the data to the RADIUS Server API (as shown in Figure 3) again, where they are forwarded to the destination IP. This process is repeated until the user sends a *session stop* message. When a *session stop* arrives to the session handler, the latter informs the device driver to deny data exchange from sender, and triggers a timestamp message in order to record the exact time user's session stopped. In parallel, the session handler deletes any related pending events and the user session from the session list.

Hotspot Controller Implementation

The RADIUS hotspot controller was developed under Linux. The implementation has been partially based on parts of the specific open source code, especially for functions related to packet handling and new attribute value pair lists creation. This choice was mainly made because of the popularity, the reliability and the stability of the specific open source RADIUS Authentication Server libraries.

CONCLUSIONS

In the previous paragraphs we presented the design of a hotspot controller based on a RADIUS based server. The main goal of the proposed hotspot controller architecture was to develop a system small enough to be imported even in handheld devices or access points. The use of off-the-shelf components for the development, makes the produced devices an excellent cost effective solution which can be used by small corporate networks or in small areas like cafeterias, airport lobbies etc. Moreover, the specific system can be easily extended in the case that new attributes or extensions are required from future versions of the RADIUS protocol.

REFERENCES

Cheswick W., Bellovin S. and Rubin A. *Firewalls and Internet Security*, Addison-Wesley, 2003.

Edney J., Arbaugh W., Real 802.11 Security: Wi-Fi Protected Access and 802.11i, Addison Wesley, 2004.

Flickenger R., *Building Wireless Community Networks*, O'Reilly & Associates, 2001.

FreeRADIUS Server Project, v.0.9.3, 2003, Available at: http://www.lh.freeradius.org/

Hassell J., *RADIUS Securing Access to Private Resources*, O'Reilly & Associates, 2002.

IPDR, Public WLAN Market Growth Depends on New Billing & Settlement Infrastructure Accounting & Settlement: A Billion Dollar Question, 2003, Available at: http://www.ipdr.org/documents/wlanwp.pdf.

RFC 2548, *Microsoft Vendor-specific RADIUS Attributes*, 2004, Available at: http://www.faqs.org/rfcs/rfc2548.html.

RFC 2759, *Microsoft PPP CHAP Extensions (MS-CHAP)*, version 2, 2004, Available at: http://www.faqs.org/rfcs/rfc2759.html.

RFC 2865, Remote Authentication Dial In User Service (RADI http://www.faqs.org/rfcs/rfc2866.htmlUS), 2004, Available at: http://www.faqs.org/rfcs/rfc2865.html .

RFC 2866, *RADIUS Accounting*, 2004, Available at: http://www.faqs.org/rfcs/rfc2866.html.

RFC 2867, *RADIUS Accounting Modifications for Tunnel Protocol Support*, Available at: http://www.faqs.org/rfcs/rfc2867.html

RFC 2869, *RADIUS Extensions*, 2004, Available at: http://www.faqs.org/rfcs/rfc2869.html.

Ziakas D., *Creating publicly available WLANs, Hotspot in a Box*, 2003, Available at: http://whitepapers.zdnet.co.uk/0,39025945,60075853p-39000522q,00.htm.

BIOGRAPHIES

Dr. ALEFRAGIS PANAYIOTIS obtained his Diploma and his PhD from the Electrical & Computer Engineering Department of Patras University, Greece. He is currently scientific associate at the Department of Applied Informatics in Management and Finance of the Technological Educational Institution of Messolonghi, Greece. His main research interests include Parallel and Distributed Processing, Object Oriented Programming, Scheduling, Optimization and Networks. He has extended professional experience in the above fields. He has published 12 papers in refereed conferences and journals. He has participated in ESPRIT, IST and National R&D projects.

Dr. TRIANTAFILLOU VASSILIS obtained his Diploma and his PhD from the Computer Science and Engineering Department of Patras University (Greece). He is currently an Associate Professor Department of Telecommunication Systems and Networks of the Technological Educational Institution of Messolonghi, Greece. His main research interests include research in the area of Networks, Telematics and New Services. He has extended professional experience in Design and Analysis of Networks and design and implementation of Open Distance Learning Tools. He has published over 25 papers in various well-known refereed conferences and journals. He has participated and technically supervised various R & D projects such as ESPRIT, ISPO.

TSATSARONIS KONSTANTINOS obtained his Diploma in Applied Informatics in Management and Finance, at the Technological Educational Institution of Messolonghi, Greece in 2004. He works as a software engineer at the WLAN Group of ATMEL Hellas S.A. and his main research interests are WLAN Security, Networks and RADIUS protocol.

Dr. VOROS NIKOLAOS received his Diploma in Computer & Informatics Engineering (1996) and his PhD in Electrical Engineering & Computer Technology (2001) form University of Patras, Greece. During the last years he has been working as hardware/software codesign specialist in various Greek telecommunication companies. He is also scientific associate at the Department of Applied Informatics in Management and Finance of the Technological Educational Institution of Messolonghi, Greece. He has published more than 30 scientific articles in international journals and conferences, while he is also co-author of the books *System Level Design Model with reuse of System IP* and *System Level Design of Reconfigurable Systems-on-Chip*, published by Springer.

A News Flash System for Mobile Devices Using J2ME

Khaled Belhoul
Mohamed Jamal Zemerly
Arif Al Hammadi
Etisalat University College
Sharjah P.O Box: 980
United Arab Emirates
E-mail: khaled84@hotmail.com, {jamal,arif}@ece.ac.ae

KEYWORDS
J2ME, GPRS, Mobile Devices, Ubiquitous Computing, News Flash.

ABSTRACT

An application (termed NewsFlash) that delivers important news articles to mobile device users is currently under development using Java 2 Micro Edition (J2ME). The application is considered ubiquitous as it runs on many different Java enabled mobile devices while on the move. The news items are delivered according to the user preferences. The system uses the GPRS/MMS mobile technology rather than the old SMS which was found to be a nuisance to some users of already available similar systems. NewsFlash consists of a web service, wireless connectivity and a range of information devices. In this paper, the critical parts of such system and the way the NewsFlash application was developed will be elaborated.

INTRODUCTION

Information can be considered as one of the most important assets people own. Therefore, access to information should always be available and not restricted to certain places such as libraries or computers that are connected to the Internet. The aim of the NewsFlash project is to provide an alternative and efficient way in which people can receive only the information they are interested in. This is achieved by designing and implementing a client-server system that allows users to receive NewsFlash messages on the move.

The main objective of the system is to enable its subscriber to specify type, frequency and the language of the news items to receive. This user control is particularly very important for financial market users where information critical to the price movement of a stock, for example, is required as soon as a news article is available during stock market opening times. Whereas in other times the news articles can be packed together in one message and sent in a time suitable to the preferences of the user. This can be achieved as the system uses the General Packet Radio Service (GPRS) (Sheikh 2004) and The Multimedia Messaging Service (MMS) (MMS 2005). The combination GPRS/MMS gives the ability to the mobile user be always connected and to send/receive messages comprising a combination of text, sounds, images and video. This will give an edge for this system over other systems that use the Short Messaging Service (SMS) which can only send text messages, one at a time, and of maximum size of about 160 ASCII characters.

According to the user preferences, MMS messages that contain either one or more news items are sent in certain preferred times. The message will also include all size of each article and fees (charged to the user according to the size of the requested message). If the client is interested in any of the headlines and sees that the fee is suitable, on a click the client will receive the entire article with associated images, graphs, audio and video clips. This is especially useful for sport articles where, for example, an instant replay of video clip that may contain that last score change and the way it was done (e.g. last goal scored in a football game).

It is important to mention that in order to maintain the on the move style of the project, the client application, from which the client sends and receives data, has been implemented using Java 2 Micro Edition or J2ME. Therefore, it will be installed on the client's mobile. Java 2 Micro Edition or J2ME is a development and runtime environment designed to put Java software on consumer electronics and embedded device. Usually, these devices are smaller and less powerful than traditional computing devices. As such, J2ME technology is an effort to condense and reduce standard Java into as small a footprint as possible (White and Hemphill 2002).

SIMILAR SYSTEMS

There are some systems that address the NewsFlash problem domain but they use different technology and approach. Here, the news sending system that depends on the SMS and the E-wap will be compared to the NewsFlash system that uses the MMS when sending the news articles.

Short Messaging Service (SMS)

There are some products in the market that use the SMS technology to send news headlines to the subscribers. However, these systems are limited in their performance. For example, Al Jazeera.net (Jazeera 2005) provides an application called (AlJazeeraMobile) that sends recent news in the form of SMS to registered mobile devices. However, these messages are limited in size which leads to two problems. The first problem is sending large number of messages especially when the client has registered to

different news categories. The second problem is that these messages have limited capabilities, for example; images, graphs, audio and video clips can not be displayed. Some subscribers might receive around twenty different SMS messages within two hours. Users of such a system may see this as a nuisance and consequently discontinue their subscription. However, the NewsFlash system overcomes this problem by grouping all headlines that belong to different categories into one MMS message. This is a relief to the clients who have been receiving large number of SMS in the other systems.

E-wap

The E-wap (Etisalat 2005**)** is another technology that has been utilized by other news sending systems in different products launched to the market. The problem with these products is that the user has to spend time browsing and looking for the appropriate news article. The user will then be charged on the time he has spent whether the right material has been found or not. While in the NewsFlash system the user receives MMS messages without paying any fees. Whenever a particular article of interest has been clicked on to be downloaded then the charging will start i.e. the user is only paying for the article he/she is interested in. The amount to be charged is also made visible along with the news headline.

SYSTEM DESCRIPTION

In order to implement the NewsFlash application, the system has been decomposed into five main modules. Each module has been implemented under different Java language specification or implementation. Figure 1 shows the function decomposition of the system.

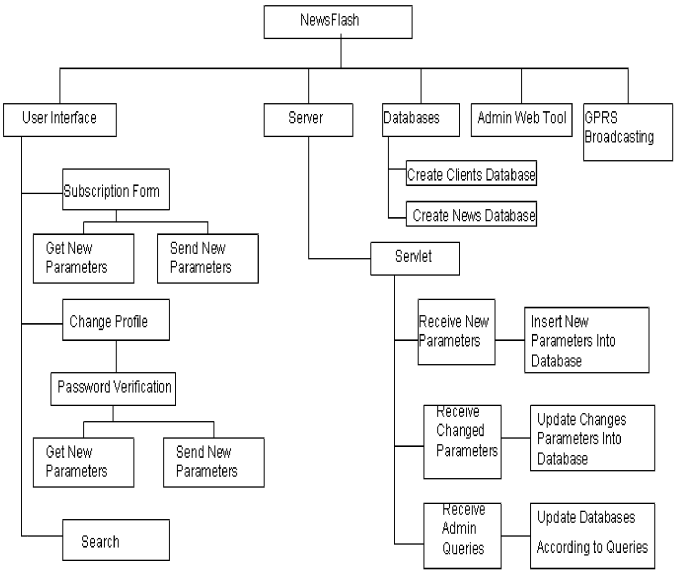


Figure 1: System's Function Decomposition

From the function decomposition of the system it can be seen that the five main modules are the user interface, server, databases, admin web tool and GPRS broadcasting. All five modules will be addressed in details in the next sections.

User Interface

The user interface has been implemented using J2ME where it is embedded in the client application that can run on any Java enabled mobile that implements the MIDP 2.0. The Mobile Information Device Profile (MIDP) is used with the Connected Limited Device Configuration (CLDC) and contains classes that provide local storage, a user interface, and networking capabilities to an application that runs on a mobile computing device such as Palm OS devices. MIDP is used with wireless Java applications (Keogh 2003).

The main target of this module is to provide a friendly user interface from which data can be collected and then sent to the second module which is the server. In order to make the application more portable all components that make up the user interface have been implemented within the High Level API. Since it is guaranteed that any mobile that implements the MIDP 2.0 can display those components. Moreover, any connection initiated by the client application is done using the HTTP protocol for the same reason, i.e. it is supported by all mobile devices that implement the MIDP 2.0.

The user interface consists of the main screen from which the user can navigate to the send subscription form, change profile form and search form. Figure 2 is a screen snapshot that shows the user interface main screen.



Figure 2: User Interface Main Screen

In order for a user to register to the system all fields in the subscription form have to be filled. Afterwards the data is sent to the server through an HTTP connection where the data is verified and entered into the system's database.

After being accepted to the system, the user can change the profile at anytime through the change profile form, where a password has to be verified. Another feature that is provided to the user is the search option that allows the user to search

for any news item in the news database. Figure 3 shows the search screen that can be accessed through the search option.

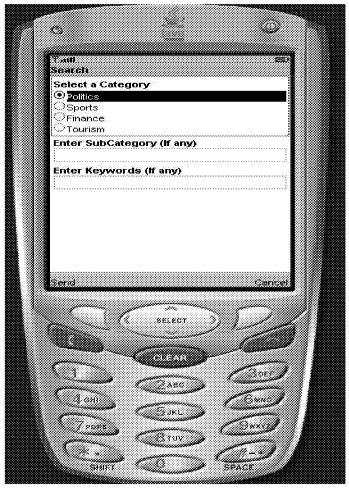


Figure 3: User Interface Search Screen

The selection of the main category in the search screen is a must. Although the selection of the sub-category and the keywords refine the searching process, they are made optional so that the searching process is simplified.

The output of this module is in the form of HTTP post requests with the corresponding data. These requests are directed to the server that makes up the second module of the system.

Server

The server used in the system is the Apache Tomcat 5.0 (Apche 2005). The servlet mainly resides on top of the server. The main functionality of the servlet is to act as the system's interface. It responds to different requests initiated by the client application or the system admin and updates the databases accordingly. Figure 4 shows how the servlet handles the requests directed to it.

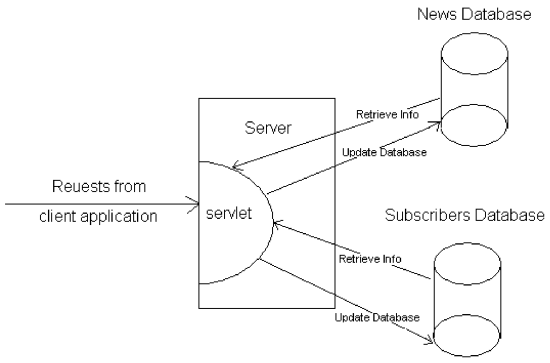


Figure 4: Servlet in Action

The servlet can access the databases using Java DataBase Connectivity (JDBC) (Sun Microsystems 2005). JDBC technology is an API (included in both Java 2 Standard Edition, J2SE, and Java 2 Enterprise Edition, J2EE releases) that provides cross-DBMS connectivity to a wide range of SQL databases and access to other tabular data sources, such as spreadsheets or flat files. With a JDBC technology-enabled driver, you can connect all corporate data even in a heterogeneous environment (Sun Microsystems 2005).
The servlet has other functionalities to perform which include verifying the requests and validating the data accompanied with it.

Databases

There are two main databases in the system which are the subscribers' database and the news database. Both are implemented using MYSQL (Sebesta 2003). The subscribers' database is built from information extracted from the user interface on the mobile application which are; the mobile class (mobile's features such as GPRS and third generation), language selection, news category, user name and mobile number. Hence, NewsFlash can be heavily customized for example, the user's customization can be Football, French League, and Time range: 10:am-6pm.Note that time range specifies the time that the subscriber would like to have messages broadcasted to him/her. On the other hand, the news database holds information about: the news headlines, the news article's source or URL, the news language and category it belongs to and some optional entries such as the news sub-category and some keywords (for searching purposes). This type of information about the news is entered by the system's administrator through the system's administrator web tool.

Admin Web Tool

The admin web tool provides a means of control over all of the vital information stored in the databases to the system's administrator. It consists of a website where buttons that corresponds to certain queries are made available. These queries are: show all clients, show all news, delete from clients, delete from news, add to clients and add to news. Once a button is hit certain info, if needed, is gathered from the administrator and sent to the servlet through http get requests with the information being added to the URL. Once it is received by the servlet the appropriate JDBC procedure is called and the response is sent back in an html form. In Figure 5 the response of the show all clients query is shown. It can be seen that it is an html table with content that dynamically changes depending on the content of the subscribers' database.



Name	Password	PhoneNumber	Politics	Sports	Finanace	Tourism	Language	TimeFrom	TimeTo
Hassan	123456	5666654	true	true	false	false	arabic	1230	1530
Abdualmun3	123	622509	false	true	true	false	Both	1230	2130

Figure 5: System's Response to Admin Query

The admin web tool also allows the administrator to change the frequency of the GPRS message broadcasting. For example, it can be set to twice a day or every couple of hours.

GPRS Broadcasting

The GPRS broadcasting is the final module which is a stand-alone Java application whose role is to send the correct GPRS message to the corresponding subscriber. This is obviously done through a GPRS gateway. One of the major events in the system is the reception of a news article. Once a news article is received, the language and category is checked. Then the news article is saved in the news database. According to the language and category the subscribers' database is checked and the list of subscribers who are supposed to receive this particular news is generated. This list which consists of the subscriber's phone number and all the news headlines that are supposed to be sent to him/her are then stored in a temporary database. After certain periodic intervals, depending on the client's preferences, the message broadcasting function checks the temporary database and sends the GPRS messages to the subscribers. Figure 6 shows the transition changes that occur when a news article is entered to the system.

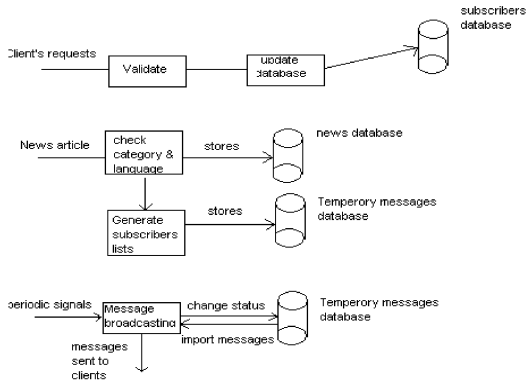


Figure 6: System's State Transition Diagram

In this module subscribers who should receive the same messages are grouped within the temporary messages database so that the messages are broadcasted to all of them at the same time. This feature improves the performance of the sending mechanism.

RESULTS AND DISCUSSION

As mentioned earlier the system is currently under development. Therefore, the final results which include the MMS message that a subscriber should receive has not been reached yet. However, Figure 7 shows how would the MMS message for a client who has subscribed for sports and finance news look like. The figure shows four news items each with the size and price that the user has to pay for (GPRS cost) if the full document is requested. The first MMS message is free since the user pays a registration fee to receive the news flashes. If the user chooses to download the highlighted article the server will extract it from the source and send it to him/her.

There are some extensions that can be added to the system. One of them is to increase the security aspect of the system by using more sophisticated wireless encryption techniques. More over, a news grabbing agent can be developed and integrated to the system so that the news articles are automatically labeled without the aid of the administrator.

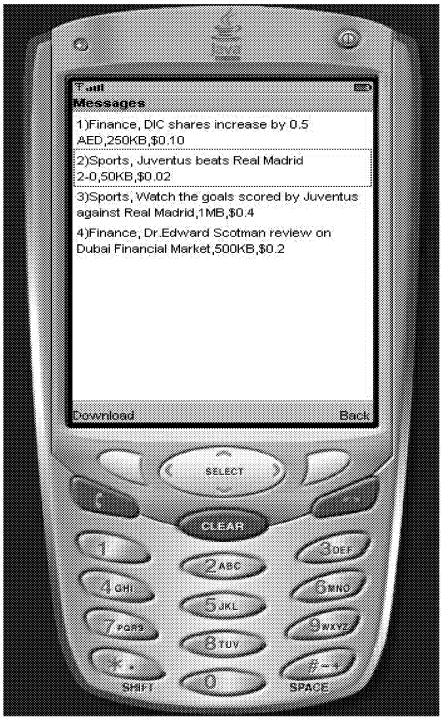


Figure 7: Expected Result, MMS Message

To enhance the system's performance, the third generation mobiles' features can be utilized. By doing so, live conferences, sports events and other long video or audio clips can be delivered through the system. The current problem that faces such enhancement is that the mobile's third generation emulator for J2ME developers has not been released yet in the public domain.

Another way to increase the participation of the subscribers in the NewsFlash system is to integrate an existing mobile based chat system. After combining both systems, a subscriber receiving the news article requested will be notified if there is an existing chat room that discusses the topic of the news article. If the subscriber clicks on the link he/she will be directed to the chat room where the topic can then be further discussed and any MMS files can be exchanged between the room's members.

CONCLUSION

This paper describes a news flash system that is being developed using J2ME for Java enabled mobile devices. The news flash system allows users on the move with little time on their hand to browse the internet or watch television to receive news items related to their interest at time intervals they specify. This system also shows how different technologies such as GPRS/MMS can be integrated to the system. The news flash also has some unique characteristics over other similar systems such as handling different languages, extensive subscriber's customization and sending full news article with related pictures, graphs, video and audio clips.

REFRENCES

Apache Jakarta. 2005. http://jakarta.apache.org, accessed 12 March 2005.

Etisalat. 2005. http://www.ewap.ae, accessed 12 March 2005.

Jazeera TV. 2005. http://www.jazeera.net, accessed 12 March 2005.

Keogh J. 2003. *J2ME: The Complete Reference.* McGraw-Hill/Osbrane 2600 Tenth Street Berkeley, California 94710, U.S.A.

MMS. 2005. http://www.mobilemms.com/mmsfaq.asp, accessed 12 March 2005.

Sebesta R. W. 2003. *Programming the World Wide Web.* Pearson Education, Inc, MA. 522-556

Sheikh A. U.H. 2004. Wireless Communications Theory and Techniques. Kluway Academic Publishers 101 Philip Drive Assinippi Park Norwell, Massachuesets 02061 USA. 647-650

Sun Microsystems. 2005. http://java.sun.com, accessed 12 March 2005.

White J. and David Hemphill. 2002. *Java in Small Things*. Manning Publications Co. 209 Bruce Park Avenue Greenwich, CT 06830.3-17

LOCATION INFORMATION SYSTEM – A DESCRIPTION OF A POSITIONING MIDDLEWARE SYSTEM

Kimmo Koskinen, Radek Spáčil, Jouni Ikonen
Lappeenranta University of Technology
Department of Information Technology
P.O.Box 20, 53851 Lappeenranta, Finland
E-mail: kimmo.m.koskinen@lut.fi

KEYWORDS

Location information, positioning, WLAN, service architecture

ABSTRACT

The spreading of wireless communication networks has brought a possibility to use the location information of the users in various applications. To make this possible, applications need APIs (Application Programming Interface) for accessing the location information. The location information of a user is private information, which needs to be protected from misuse. For this reason, the user should have complete control over the usage of his own location information. There is a need for middleware systems which gather location information, provide an API for the services and allow users to authorize the use of their location information. This paper presents Location Information System (LIS), a middleware system for location based services in Wireless Local Area Networks (WLAN).

INTRODUCTION

The development of Location Based Services (LBS) presents a wide application area. This area includes activities such as gathering, categorizing storing, accessing and authorizing the usage of location information. This paper discusses one example implementation of a platform for providing LBS services. The solution provided is not complete system for covering all possible usage cases, but a proof of concept for gathering position information from a wireless local area network and allowing services a limited access to the gathered information. Positioning system forms the active part of the location service, usually by gathering sensor information and combining this information with a location model. A positioning system itself may consist of sensors and a position calculation system which evaluates sensor data to calculate the position of an object. Survey and categorization of positioning systems can be found from [Hightower 2001]. Generally current positioning systems are application and/or sensor system dependent. Hightower et. al [Hightower 2002] introduced an idea of location stack, to make position data generating systems and applications independent of each other. Idea is very good, but real implementation is quite complex (of cource debending on level of detail the stack provides, e.g. how security and access to data is controlled).

Some of well known positiong systems are:

- Active badge system [Want 2001], where the locatable objects carry an "Active badge", a small tag-device which transmit an infrared signal with unique code in short intervals. This signal is observed by a network of sensors which are placed in an area where the tag's are desired to be locatable. A master station polls the sensors for tag sightings and this data is then available for applications.
- Microsoft RADAR [Bahl 2000] and commerical Ekahau Positioning Engine (EPE) [Ekahau 2005], which both use WLAN (802.11b/g) networks for positioning. The locatable devices are a WLAN network interface cards, which are identified by their hardware address (MAC). Location based services can retrieve the location of a devices from the manager station through a specified APIs.
- Global Positioning System (GPS) differs from previous type of systems as object to be located calculates their own location by themselves. In GPS, the locatable device has a receiver which collects the signal which is transmitted by satellites orbiting the earth. The receiver device can calculate it's own position from at least three individual satellite signals. Still, the GPS receiver itself does not know the location of other GPS receivers. If the location information of one device needs to be accessible by many other parties, there must be a system distributing this information.

Different positioning systems offer different APIs for the location based applications. Application use these interfaces

to retrieve the location information of the terminals. Depending on the communication network, terminals can be identified with, for example, IP addresses or MAC addresses (e.g. in 802.11b WLAN networks). Identifying different terminals is necessary to be able request the location of a specific terminal. For the users to be able to authorize the usage of their location information, the terminal has to be associated with a user. Then it depends on the policy of the user to allow or disallow the usage of the location information. There may also be some system specific policies, such as "a user may always get his own location information", which can be implicitly applied. Also, there may exist policies for emergency situations in which the location information of the user can be given to authorities such as the police [FCC 2005]. The user may also have some straightforward methods to disallow the usage of the location information. For example, if there is a software agent gathering and forwarding sensor data from the terminal of the user, the user can just "switch off" the terminal or the agent. However, when multiple applications can use the location information of a user in various ways, there is a need for centralized management of the usage of location information. Still, forcing a system level access policy is not straightforward.

CASE WLPR.NET AND RELATED WORK

The wireless network built by Laboratory of Communication Engineering at Lappeenranta University of Technology—WLPR.NET—covers the campus of the Lappeenranta University of Technology and part of the Lappeenranta city. The WLPR.NET is built according to *Lappeenranta model* [Juutilainen 2002], which allows the access network (built with devices conforming to 802.11b/g standards) to be centrally managed (as opposed to *community networks*) and still have an option for the user to choose independently an Internet Service Provider (ISP). This network model emphasizes the possibility to create local services. One type of local service are Location Based Services (LBS). The creation of a unified positioning platform in the local access network will benefit the possible service providers and increase the attractiveness of the local network. The current cell based positioning system in the WLPR.NET network has minimal requirements from the client devices. No additional software is needed from the client device to be able to use the positioning system. Location data of the terminal devices of the users is gathered from the access points in the network mainly with RADIUS (Remote Authentication Dial In User Service) service. At the first stage, the location data was stored in a database to which every application had access rights. This however created problems for the management of the application, as they had to be separately registered by the database administrator. The functionality of the applications was tied to the database structure which could be subject to change. Also the users in the network did not have sufficient control over the usage of their location information. Moreover, this model restricted the possibilities of the applications by allowing only request-response type of operation. If the applications needed to react to changes in location information, they had to poll the database for updates which causes unnecessary traffic. It was seen practical to provide an alternate way to make use of location information in the communication protocol level and to provide a model for creating LBSs.

Applications for Location Information Systems

Use of location information in different applications has been studied in, for example [Hightower 2001, Hightower 2002], the area of ubiquitous computing. Location information in ubiquitous computing has been used for example in the Aura project at the Carnegie Mellon University [Aura 2005]. An example of context-aware application where location information has been used is the *Portable Help Desk* [Galan 2002] application developed in the Aura project. Another application of positioning enabled network was demonstrated by Lancaster GUIDE project [Davies 2002], a hand-held computer based tourist guide for visitors to Lancaster. The location information of hand-held computers in a WLAN network was used in creating this system. However, the technical solution depends on non-standard changes of 802.11b network protocols. It is necessary to note that GUIDE project started in time where 802.11 standard was not widespread and therefore special solution was needed.

APIs for Location Information Systems

General APIs for location systems have been defined by many organizations. The Location Working Group (LOC) [OMG 2005] (formerly known as Location Interoperability Forum, LIF) of the Open Mobile Alliance (OMA) has developed a specification for Mobile Location Protocol (MLP). This protocol is an application level XML (eXtensible Markup Language) protocol for getting the position of mobile stations. The XML messages can be transported by different transport level protocols such as HTTP (Hyper Text Transfer Protocol) or SOAP (Simple Object Access Protocol) [W3C 2005]. The protocol itself is independent of the network technology: the bearer and location derivation technology. The Parlay Group [Parlay 2005] has defined Open Service Access (OSA) API which consists of 14 parts. This definition enables service providers to use the services offered by mobile networks. Part 6, Mobility Service Capability Feature (SCF), provides a definition for accessing the location information of the user in mobile networks. There exist mappings for at least Java, CORBA (Common Object Request Broker Architecture) IDL (Interface Description Language) and WSDL (Web Service Description Language) of

the Mobility SCF. All the above mentioned API's strive to provide some common way to retrieve location information of users in the network. The problems is that they are defined primarily for mobile telecommunication networks and not for wireless data transfer networks such as WLAN. Implementing these interfaces for use in this work would be a too cumbersome task. Secondly a common model of authorizing the usage of location information by the services must be taken into account.

WLPR LOCATION INFORMATION SYSTEM ARCHITECTURE

The architecture of the LIS is a collection of services joined by a database and interfaces for:

- collecting and updating location information
- querying location information
- registration of services and users
- management of user rights and service attributes

Requirements for the LIS

The API for LIS should be available in many programming languages to not limit possible applications. Also the problem of managing service authorization and user rights must be solved. Open standards have to be used to guarantee free usage of the LIS API. Extensibility allows fast development and the ability to make modifications easily. Implementation based on XML ensures good extensibility.

CORBA [Corba 2005] was one option for the API but it was seen to be too heavy for LIS. The possibility of using SOAP [W3C 2005] for the LIS API was investigated. This lead into studying the Web Services framework [Webservices 2002]. Web Service technologies provide a way to describe the interface of a service with WSDL [WSDL 1.1]. With WSDL, the LBS API can be specified in a platform and programming language independent way. SOAP is a XML -based messaging framework for exchanging XML -messages. The WSDL definition of a service can be mapped to SOAP. SOAP can then be mapped to many transport protocols. HTTP mapping for SOAP was chosen in this work. The use of WSDL and SOAP provided a way to both describe the LIS service interface and to define a protocol for communication between the LIS and LBSs. Users have to be able to control the usage of their location information. This can be done by specifying access control to the location information. This is problematic to implement in the service API, because this interface is meant only for communication between the LIS and the LBS. This was solved by storing the access policies to LIS. These policies define which services are able to request the location of the user. The user can edit these policies with a HTML based interface. This is done through a secure HTTPS session.

Overview of the LIS

An overview of the LIS and the WLAN positioning system is presented in Figure 1. The user uses his terminal to interact with the location aware service. The location aware service uses a SOAP interface to query the location information of the users in the WLAN network. As the user is using the service in the WLAN network, the location information of his terminal gets recorded with the use of RADIUS service. WLAN access points are configured to used 802.11 Open Systems authentication. This causes the access points to make authentication requests to the RADIUS service whenever the terminal of the user is activated (i.e. turned on) or when the terminal changes the access point where it is connected to. The cell based location data from RADIUS access reports is stored into a *Location Database* with the help of a *taildaemon* program. The *taildaemon* parses the RADIUS log files and extracts the cell (identified by access point IP address) of the terminals (identified by MAC address) in the network. The LIS then uses the *Location Database* when location information of a user is requested.

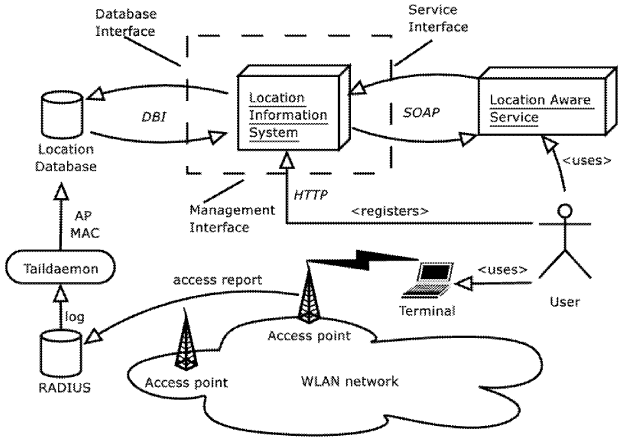


Figure 1. LIS overview

User and service management interface

In the LIS, both users and LBSs have accounts. Both account types have username/password authentication. Users have a *nickname* and LBSs a *LBS name*. Both account types have attributes related to them. These attributes are managed through a HTML web page interface, which is the *Management Interface* in Figure 1. To get an account, users register to the LIS. For LBSs, the registration is done by the service provider. Both service providers and users can log into the system with their *user name/password* pair to make changes to the account attributes. User accounts contain the MAC ad-

dress of the user's WLAN device which is detected when the user registers for the first time and also on subsequent logins. The user can use multiple devices, therefore there is always one *active* device, which has the same location as the user. Users can see a list of LBSs registered to the LIS and can select which LBSs are able to use the location information of the user. The accounts can also be removed by the corresponding user or service provider who owns the account.

Service API

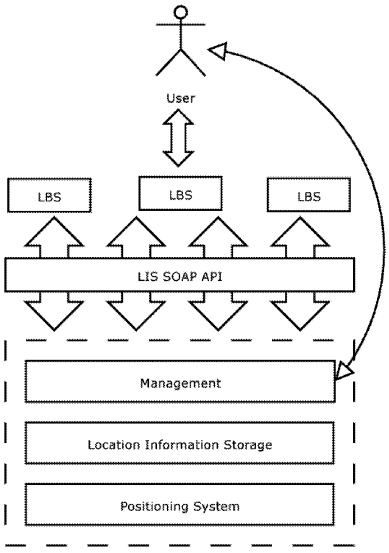


Figure 2. LIS layer architecture

The LIS can be divided into a layer architecture with interfaces between each layer. The layer architecture of LIS is presented on Figure 2. The lower most level is the *Positioning System*. Positioning system produces location data which is stored in the *Location Information Storage* layer. *Management* layer contains functionality for user and LBS account management. The SOAP API, specified with WSDL, provides two types of operations for the LBSs, *request-response* and *trigger* operations. The following *request-response* operations provide immediate answers. LBSs can use *GetUserLocation* operation for fetching the location of a specific user. The LBS will get in return a location identifier from which the LBS can fetch more attributes with *GetLocationInfo* operation. *GetLocationList* operation returns a list of all cell id's present in the cell positioning system.

Every operation specifies the *LBS name* and the *password* in the request. When the operation is executed, the authenticity of the LBS is checked first. After this the operation itself is authorized. Figure 3 presents the authorization procedure. Currently, the *GetUserLocation* specifies also *caller* parameter, person who is making the positioning request. However, user-to-user authorization is not implemented yet and this parameter is not used. The current functions and their parameters are defined in Table 1.

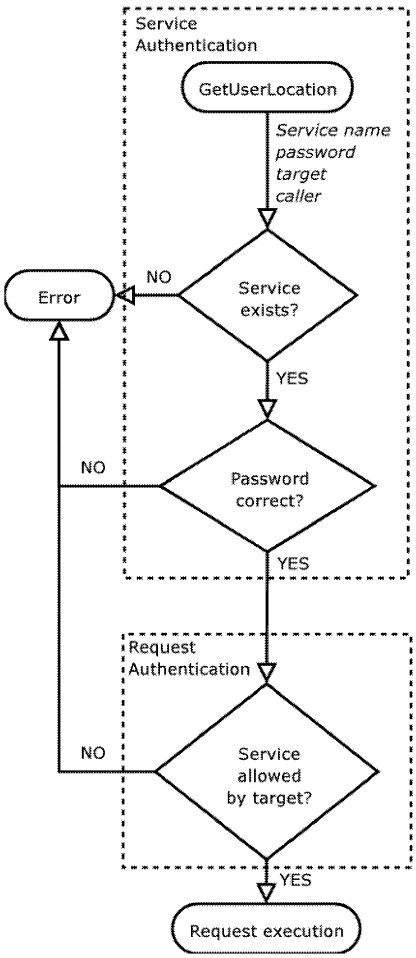


Figure 3. Request authorization

Trigger operations provide the LBS a conditional way of reacting to the change in location information. *SetTrigger* operation is used to place a trigger on a specified user. The LBS will get in return an identification of the trigger. Whenever the user moves, a SOAP callback function *FireTrigger* is executed by the LIS. This callback function is implemented by the LBS. *FireTrigger* function has the following parameters: new location of the user and the identification of the trigger which had been fired. *DeleteTrigger* can be used by the LBS to remove the triggers which have been set. The trigger notification should be authorized before information is actually sent to the LBS. This is because the user targeted by the trigger might have disallowed the LBS to use his location information after succesfull *SetTrigger* operation. The triggers are implemented in a PostgreSQL database with PL/pgSQL. PL/pgSQL is a loadable procedural language for the PostgreSQL database system. It allows to extend the database functionality with SQL like procedural language. The functionality of the trigger mechanism is presented on Figure 4.

Function	Parameters
request-response methods	
GetUserLocation	target,caller,service, password
GetLocationInfo	location id,location type, service,password
GetLocationList	service,password
trigger methods	
SetTrigger	target,URI,proxy,service, password
DeleteTrigger	trigger_id,service,password
FireTrigger	trigger_id,location

Table 1. LIS API methods

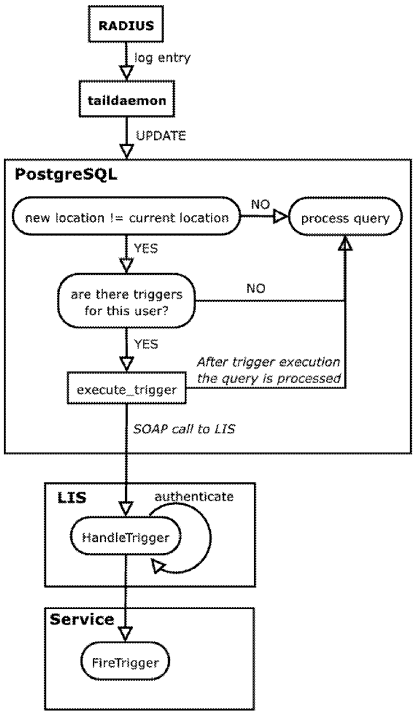


Figure 4. Trigger functionality

LOCATION INFORMATION SYSTEM APPLICATIONS

Currently there exists one application which uses the LIS system. This is the Floating Note application described in [Multaharju 2004]. The idea in this application is to present a new way to use the location information provided in a WLAN network in the communication between the users of the network. Users can post *notes* in their current location on a forum in which message threads are arranged by the cell locations of the network. Different locations contain different threads of messages. The floating note uses the *request-response* interface provided by LIS for fetching location information of the user. The user has an account in the Floating Note application and this account contains the user name of the user's LIS account. When the Floating Note requires to know the location of the user, the SOAP request is authorized based on the allowance rules of the targeted user. Other possible applications would include tracking applications or context aware applications which would react to the change in the location information of the user. For example, a notification application where a user could set a condition such as "remind me to borrow a book from library when leaving from work". The notification would contain the work area as one parameter, and when the user leaves the work area, a notification message is displayed to the user.

DISCUSSION

The LIS layer architecture gives a possibility to combine location data from different positioning systems. This would include at least GPS data. Applications could decide on what single positioning system to use but combining data from different sources could also be beneficial. This leads into the discussion of combining the different reference systems used by different positioning systems. Positioning systems use a reference model in which to express location information. These *coordinate reference systems* may vary from positioning system to another. For example, cell positioning systems provide a discrete one-dimensional reference space (a list of cell id's) for the location data. Other positioning system may use multidimensional, continuous coordinate reference systems. An example of this kind of system is GPS (Global Positioning System) which can offer location information expressed in *latitude*, *longitude* and *altitude*. There exists also "hybrid" positioning systems which offer a mixed discrete and continuous presentation of location information. For example Ekahau EPE [Ekahau 2005] presents location information with *floor-maps* (discrete set of maps) and continuous x and y coordinates inside the floor-maps. Despite of the different reference models, mappings from one reference model to another can be established. Discrete cell locations can be mapped to a continuous reference model by defining cell coverage areas in the continuous reference model coordinate space. Mappings from multidimensional reference models to another can also be established. It should be noted that coordinate transformation operations are usual in Geographic Information Systems (GIS). Coordinate transformation operations are needed because reference systems may differ, for example, from country to country. As long as applications are using the coordinate system of only one positioning system, they can not benefit from the possibilities offered by other positioning systems. This differentiates the applications by the use of a particular positioning system. There exists a need to have a uniform way to refer to location data. A location model with a computable location identifier has been presented in [Changhao 2002]. The authors define a location identifier (Aura Location Identifier, ALI) which is used for referring to location data. In ALI

different location representations (coordinate based, hierarchical) are combined and expressed with a URI (Universal Resource Identifier) -like presentation. This method allows to refer to the location data produced by different positioning systems in a uniformed way. It also allows to execute operations for different ALIs, such is "what is the distance between two ALI references". A method of combining location data, which is described in [Angermannr 2001], uses probability density functions of each location from different sensing systems to calculate resulting position. By super-positioning data from several sensors it is possible to reach higher accuracy of the resulting location. OpenGIS has taken a different approach for the combination of different coordination reference systems by defining a coordinate transformation service [Opengis 2005]. This service can be used to conduct coordinate transformation calculation and thus to ease the task of the location application. Another item of development would be more complex triggers. Some services would benefit from the possibility to define area targets for the triggers which would fire when a user enters or leaves the area. Eventually the trigger mechanism could contain a logic execution environment which would allow to construct more complex rules for firing the triggers. The procedure currently allows only coarse selection of which service is allowed to use the location information of specific users in the operations provided by LIS. Future development would include specification and implemention of user-to-user authorization. This would rely on the service to provide the correct caller and target pair in the operations (for example, "User A want's to know the location of user B"). This is needed for handling the user-to-user authorization policies in the LIS (as opposed to letting the LBS itself create user-to-user policies and decide whether "user A is allowed to know the location of user B"). However, this issue can be problematic because there is a need to form trust relationships between users, LBSs and the LIS. A system for providing location policies has been presented in [Hengartner 2003]. The authors discuss the problems on defining location policies and an implementation based on public key infrastructure.

CONCLUSION

This paper presented a platform for creating Location Based Services (LBS). The LIS platform provides an interface for accessing the location information of users in an open WLAN network. The LIS provides also a way for the user to control the usage of their location information. Division to several independent modules with pre-defined interfaces ensures good modularity and scalability. LIS uses a cell based WLAN positioning system but the architecture allows to acquire location information from other positioning systems. The API provides two different methods for the services to access the location data. *Request-response* and *trigger* methods. Request-response method provides immediate answers for the queries. Trigger method provides the service a way to react to the change in location information (event based). Privacy is ensured by service and user rights management system, both services and users have accounts in the LIS. Management interface allows the users to define which service can access their location information. On protocol level, every method call is authenticated by service name and password. It is necessary to say the current implementation is not suitable for applications which need very small response times. At this stage of development, the emphasis lays on modular architecture, rather then speed.

References

[Hightower 2001] Jeffrey Hightower, Gaetano Borriello, "Location Systems for Ubiquitous Computing", *Computer*, August 2001, Vol. 34, no.8, pp.56-66.

[Hightower 2002] Jeffrey Hightower, Barry Brummit, Gaetano Borriello, "The Location Stack: A Layered Model for Location in Ubiquitious Computing", *Proceedings of the 4^{th} IEEE Workshop on Mobile Computing Systems & Applications*, Callicoon, New York, June 2002, pp. 22-28.

[Want 2001] Want, R., Hopper, A., Falcao, V., Gibbons, J., "The Active Badge Location System", *ACM Transactions on Information Systems*, January 1992, Vol. 10, No. 1, pp 91-102.

[Bahl 2000] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System", *Proceedings of IEEE Infocom 2000*, Tel-Aviv, Israel,March 2000, pp. 775-784.

[Ekahau 2005] Ekahau, *Ekahau website*, `http://www.ekahau.com`, visited: February 13, 2005

[FCC 2005] Federal Communications Commission, *FCC: Enhanced 911 - Wireless Services*, `http://www.fcc.gov/911/enhanced/`, visited: February 14, 2005

[Juutilainen 2002] Juutilainen, M., Ikonen, J. and Porras, J., "Evaluation of a Next Generation Public Wireless Multi-ISP Network", *Proceedings of 27^{th} conference on Local Computer Networks*,IEEE, Tampa, Florida, 6-8 November 2002, pp 405-414.

[Aura 2005] Carnegie Mellon University, *Project Aura, Distraction-free Ubiquitous Computing*, `http://www-2.cs.cmu.edu/$\sim$aura`, visited: February 13, 2005

[Galan 2002] Garlan, D., Siewiorek, D., Smailagic, A., Steenkiste, P., *Project Aura: Toward Distraction-Free Pervasive Computing, IEEE Pervasive Computing, April-June 2002*, `http://www-2.cs.cmu.edu/$\sim$aura/docdir/aura-pervasive02.pdf`, visited: February 13, 2005

[Davies 2002] Davies, N., Cheverst, K., Friday, A., Mitchel, K., "Future Wireless Applications For A Networked City: Services for Visitors and Residents", *IEEE Wireless Communications*, Feburary 2002, Vol.9, Issue 1, pp 8-16.

[OMG 2005] Open Mobile Alliance, *Location Working Group Home Page*, `http://www.openmobilealliance.org/tech/wg_committees/loc.html`, visited: February 13, 2005

[Parlay 2005] The Parlay group, *Home page*, `http://www.parlay.org`, visited: February 13, 2005

[Corba 2005] Object Management Group, *OMG's CORBA Website*, `http://www.corba.org`, visited: February 13, 2005

[W3C 2005] W3C, *SOAP Version 1.2 Part 1: Messaging Framework*, `http://www.w3.org/TR/2003/REC-soap12-part1-20030624/`, visited: February 13, 2005

[Webservices 2002] W3C, *Web Services Activity*, `http://www.w3.org/2002/ws`, visited: February 13, 2005

[WSDL 1.1] W3C, *Web Services Description Language (WSDL) 1.1*, `http://www.w3.org/TR/wsdl`, visited: February 13, 2005

[Multaharju 2004] Multaharju, M., Koskinen, K., Ikonen, J., "Floating Note—A Location Based Messaging Application", *Proceedings of 2nd Workshop on Applications of Wireless Communications*, Lappeenranta, Finland, August 2004, pp. 9-16.

[Changhao 2002] Changhao, J., Steenkiste, P., "A Hybrid Location Model with Computable Location Identifier for Ubiquitous Computing", *UbiComp'02*, Goteborg, Sweden, 2002.

[Angermannr 2001] Angermann, M., Kammann, J., Robertson, P., Steinga, A., Strang, T., "Software representation for heterogeneous location data sources within a probabilistic framework", *International Symposium on Location Based Services for Cellular Users*, Locellus, February 2001, pp. 107–118.

[Opengis 2005] OpenGIS consortium, *Web Coordinate Transformation Service*, `http://www.opengis.org/docs/02-061r1.pdf`, visited: February 13, 2005

[Hengartner 2003] Hengartner, U., and Steenkiste, P., "Protecting Access to People Location Information", *Proceedings of First International Conference on Security in Pervasive Computing*, SPC 2003, Boppard, Germany March 2003.

COMMUNICATIONS SECURITY

ON THE SECURITY OF NON-RSA EMV PAYMENT CARDS

Mårten Trolin
Royal Institute of Technology (KTH)
Nada, SE-100 44 Stockholm, Sweden
Tel: +46 8 790 60 00, Fax: +46 8 790 09 30
marten@nada.kth.se

KEYWORDS
EMV, smart-cards, software protocols

ABSTRACT

We show a weakness in the specification for offline capable EMV payment cards. The weakness, which applies to cards without RSA capability, enables an attacker to duplicate a card and make transactions that cannot be tied to the original card.

INTRODUCTION

A large part of today's electronic purchases are made with different kinds of payment cards. The majority of the cards used today have a magnetic stripe where the card data is stored. Over the last years, card skimming, where the content of the magnetic stripe is copied, has become a major problem. The countermeasure is to move from the magnetic stripe to smart-cards where the data is stored on a chip instead. To make sure also smart-card based payment cards will have the same global acceptance as the magnetic stripe, Europay, MasterCard and Visa have together developed the EMV specification.
The preparations for moving from payment cards based on magnetic stripe to smart-card based cards have been going on for more than ten years. Some card issuers have already converted their card base to EMV smart-cards, and more are about to make the switch.
The base for EMV smart-cards is the EMV specifications (EMV 2000), which define the protocol between the card and the terminal. Payment organizations, in particular Visa and MasterCard, have developed their own extensions to the EMV specifications (MasterCard 2001; 2002, Visa 2000).
In this paper we will examine a potential problem in the configuration of an EMV card. In particular we will show how to avoid this problem with a card based on Visa's VSDC specification, and that it cannot be avoided when using MasterCard's M/Chip specification.
EMV specifies two possible security levels for cards, Static Data Authentication (SDA) and Dynamic Data Authentication (DDA). A DDA card must support RSA, whereas an SDA card does not. A DDA card therefore has a private RSA key that can be used for offline authentication. An SDA card has no such key. The issue we discuss in this paper relates only to SDA cards.

SMART-CARDS

A smart-card is a tiny computer with its own CPU and storage. The data stored on the card cannot be read or written directly but only through certain functions. This means that a smart-card may have keys that can be used for encryption but cannot be read in clear. Another possibility is to have a PIN that must be entered before certain functions can be used. More information about smart-cards can be found in (Jurgensen and Guthery 2002).
The fact that the data on the card can only be accessed through predefined functions means that we can define data to be *public* when it can be accessed in clear and *private* when it is used only for internal processing by the card. When analyzing protocols involving smart-cards, a reasonable security model is to assume the data can only be accessed and modified using the predefined functions. The weakness analyzed in this paper follows that security model.

CRYPTOGRAPHIC FUNCTIONS

In this section we will briefly discuss the cryptographic functions that are used in an EMV transaction.

Digital Signatures

Assume Alice sends a message to Bob, and Bob wants to make sure the message really is from Alice, and that it has not been modified. If digital signatures are used, Alice has a *public* and a *private* key. Alice can give the public key to anyone that she wants to communicate with, and the private key is kept secret. She uses the private key to sign the message, and Bob can verify the signature using Alice's public key.
The signature scheme used by EMV cards is RSA.

Public Key Based Authentication

Assume Alice wants to convince herself that the person she is talking to actually is Bob and not someone impersonating Bob. Assume also that Alice and Bob have never met or exchanged information before, so Bob cannot have given Alice a public key. In a public key infrastructure (PKI) this

is solved as follows. Alice trusts a third party, the Certificate Authority (CA), and has access to the CA's public key. Bob has had his public key signed by the CA's private key together with information identifying the key as Bob's. The signature is called a *certificate*. Now Bob can give his public key together with the certificate to Alice. Alice convinces herself that the key indeed belongs to Bob by verifying the certificate.

Message Authentication Codes (MACs)

The above solution of course works also when Alice and Bob know each other. However, public key cryptography has the drawback of being slow, and therefore is used only when no other possibility is available.
If Alice and Bob know each other, they may exchange a symmetric key k which only they know. Using the key, Alice can compute a *message authentication code*, MAC, for a message. Since also Bob knows k, he can verify the MAC. Also Bob can use k to compute a MAC which Alice can verify. Since the key k is knows only to Alice and Bob, the MAC is unintelligible to anyone else.

THE EMV SPECIFICATION

The Participants

The main participants in an EMV transaction are

- the card. The card is held by a customer (the cardholder) and contains information (both secret and public) set by the issuer. The card is identified by its card number.
- the terminal. The terminal is located at a merchant and is able to interact with cards. The terminal has internal settings defining when to allow offline transactions and when to require transactions to be processed online.
- the acquirer. The acquirer receives data from the terminal. In this principal description, the acquirer does nothing but forwarding the information to the issuer.
- the issuer. The issuer issues cards to his customers and processes transactions where his cards are involved.
- the payment organization. The payment organization is responsible for forwarding information between issuers and acquirers, and also for issuing public key certificates to the issuers.

In Figure 1 a schematic overview is given.

Transaction Outcome

The EMV specification describes in detail the data flow between the card and the terminal during a transaction. The result of an execution the protocol is one of the following

1. Transaction is approved offline.
2. Transaction is denied offline.
3. Transaction is sent to the issuer for online authorization. (In this case the protocol also defines the behavior if no response is received from the issuer, but this will not be discussed in this paper.)

Since most transaction are either approved offline or sent online, we will consider only these two cases here. The principle is that a transaction can be approved offline *only* if *both* the card *and* the terminal agrees on it, but is sent online if at least one of them requests it.
Both Visa and MasterCard have written their own extensions to EMV. Here they define which of the public EMV parameters that can be used, and also what the internal behavior of the card should be. Visa calls their application VSDC (Visa 2000). MasterCard has published two separate documents, one giving the external interface in the form of minimum requirements (MasterCard 2001) and one defining the internal behavior by describing the application M/Chip (MasterCard 2002). However, also MasterCard is moving towards a unified document giving both internal and external details.

Cryptographic Authentication

There are two different forms of authentication in EMV transactions, authentication between the card and the terminal and authentication between the card and the issuer.

Card-terminal Authentication
In the very beginning of a transaction, the terminal verifies the authenticity of the card. This can be done either with *Static Data Authentication* or with *Dynamic Data Authentication*. In the first case, the card has a (static) signature on some subset of its parameters which is verified by the terminal. In the second case, the terminal verifies that the card also knows a private key by letting it sign a random nonce.

Card-issuer Authentication
During a transaction, the card may generate one or two Message Authentication Codes, MACs. These MACs are generated with a symmetric key known only to the issuer and the card. Therefore the MAC can only be verified by the issuer and not by the terminal or the acquirer.

Card Configuration

SDA and DDA
EMV gives the option of using low-cost cards without RSA capabilities as well as more expensive RSA enabled cards. Cards without RSA capabilities support only *Static Data Authentication* (SDA) whereas cards with RSA support can handle also *Dynamic Data Authentication* (DDA).
For both SDA and DDA, the issuer receives a certificate from the payment organization. The issuer certificate and the issuer public key, IPK, are stored publicly on the card. For SDA, the issuer signs a set of card parameters of his choice

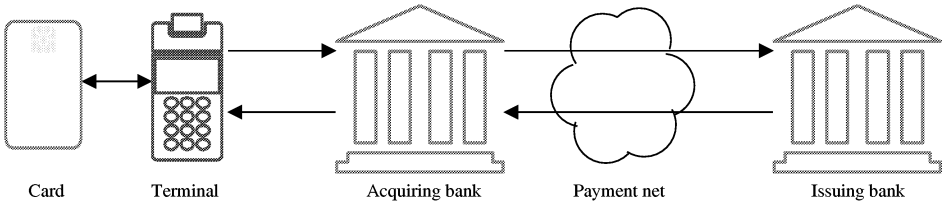


Figure 1: Data Flow of an EMV Transaction

with his private key and places the signature on the card. The signature is called *Signed Static Application Data*, SSAD. In the case of DDA the card is given its own RSA key pair. The card private key is stored internally on the card, but the card public key is signed by the issuer. (Even if the card supports DDA, an SDA signature is still put on the card.)

Card parameters
Apart from the keys and certificates mentioned above, several parameters describing under which circumstances to allow offline transactions are stored on the card. In this paper we are only interested in one parameter, namely *Lower Consecutive Offline Limit* (LCOL). The LCOL gives the number of transactions that can be performed offline, i.e., without contacting the issuer.
Also the parameters *Application Transaction Counter* (ATC) and *Last Online Application Transaction Counter* (LATC) are stored on the card. The ATC contains the number of transaction the card has performed and the LATC holds the index of latest transaction that executed online. They are both initialized to zero.

Symmetric keys
When a card is issued, the issuer generates a symmetric key that is stored on the card and used to generate MACs on transaction messages. Up to three other symmetric keys are also stored on the card, but since they play no role to this paper, we ignore them. The keys are also stored by the issuer, but not disclosed to the merchants or the acquirers.

The Protocol

In Figure 2 an overview of the protocol used for deciding the result of the transaction is shown. Here we only describe the case of an SDA card.

1. The terminal reads all public data from the card.

2. The terminal receives transaction data, including the amount and currency of the transaction.

3. The terminal uses the public key of the payment organization to verify that the issuer certificate is valid and then uses the issuer public key (IPK) to verify that the SDA signature is valid (SSAD). If any of these steps fail, the processing is terminated.

4. The terminal reads the ATC and the LATC off the card and computes the number of offline transactions as (LATC − ATC). It compares this value the *Lower Consecutive Offline Limit* (LCOL) read in Step 2 and proceeds as follows:

   - If the number of offline transactions is lower than LCOL, the terminal makes a request for an offline transaction to the card.
   - If the number of offline transactions is equal to or higher than LCOL, the terminal makes a request for an online transaction to the card.
   - If the LCOL is not present, the terminal makes a request for an offline transaction to the card.

5. After receiving the terminal's decision, the card may perform additional internal risk analysis before it makes the final decision. However, if the terminal requested an online transaction the card may not override the decision to make the transaction offline.

   - If the card receives a request for an online transaction, it creates a MAC over transaction data and returns to the terminal. This data is sent to the issuer who verifies the authenticity of the MAC.
   - If the card receives a request for an offline transaction, then depending on the result of the internal analysis, it either approves the transaction offline or requests an online authorization. In both cases a MAC is created.

THE PROBLEM

In this section we will describe the potential problem, and also how to avoid it where possible.

Making a Pure Online Card

In many cases, it is desirable to have a card that can only function online. There are two ways to achieve this:

- Set the LCOL to zero. This way the terminal will always make a transaction go online.
- Make the internal risk analysis of the card such that it always makes the decision to go online, regardless of what the terminal's decision is.

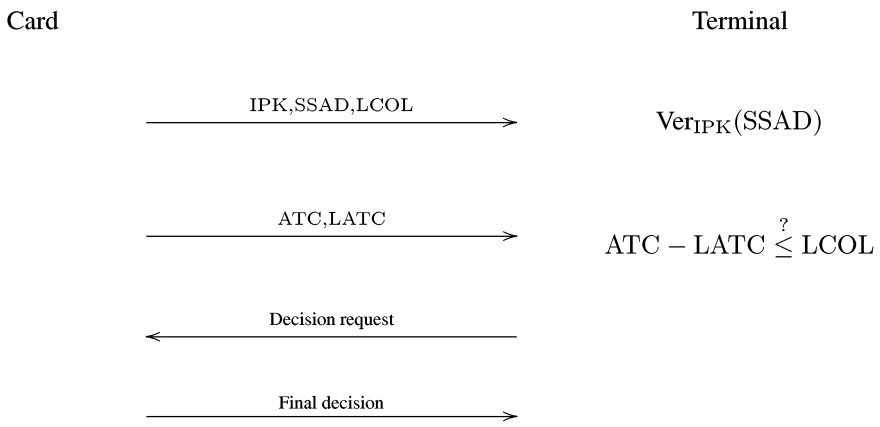


Figure 2: The Protocol Used for Deciding the Outcome of a Transaction

The most obvious reason to make a card online-only is to make sure the cardholder does not spend money he does not have. However, for an SDA card there is also another reason. Since the MAC cannot be verified offline, someone might copy the card, keeping the original SDA signature, but replace the symmetric key. Then a terminal would accept the card (since the SDA signature is valid), and when (and if) the issuer detects that the MAC is invalid, it is already too late.

The essence of the attack described here is to copy the card and modify the copy in such a way that it will allow offline transactions.

Copying an SDA Card

If we assume that the hardware is secure, the adversary can copy all the public data on a card, but not the internal data. Also, when the card is copied, he can modify data that has not been signed by the issuer, but if he changes the data included in the SDA signature the card will not be accepted. When a card is copied the adversary can change the internal behavior of the card by replacing the original program code by his own.

Making an Online Card Work Offline

As mentioned above, there are two ways to make a card online-only. Here we will discuss different attack scenarios depending on what method is used.

If the LCOL parameter is set to zero, then it may either be signed or not be signed with the SDA signature. If it is not signed, the adversary can simply copy the card and modify the LCOL to contain a non-zero value. The card will be accepted offline, since the SDA signature is still valid. He will not be able to copy the symmetric key, so the copy cannot be used online, but as long as the card only is used offline, it will work. In case the LCOL is signed, it cannot be changed and the attack does not work.

If the LCOL is not present on the card, but the internal risk analysis of the card is used to make all transaction go online, then the attack is a little bit different. When the adversary copies the card, he replaces the card application with an application that always accepts to make the transaction offline. Also here he cannot copy the symmetric key, but he will be able to use the card offline.

Note that for any of these attacks, the adversary only needs access to the card for a few seconds so that he can read all the public data. Since the commands for doing this are standardized, any card-reader could be used for this.

Copying an Offline-Enabled Card

If the card has LCOL non-zero, but the LCOL is not in the SDA signature, the adversary can of course use a similar method to get an arbitrary number of offline transactions (with invalid MACs, making it impossible to tie the transaction to the card). However, even if then LCOL is signed, the adversary can issue an attack similar to those described above. He can copy all the parameters on the card, but modify the card application to that it always responds that no offline transactions have been performed prior to the current. That way the terminal will always accept to make the transaction offline (since the number of offline transactions is lower than LCOL) and the issuer will not be able to detect that the MAC is invalid.

Protecting Against the Attack

As we can see, the only way to make the card secure against the proposed attack is to set the LCOL to zero and include it in the SDA signature. In other words, there is no way of making a secure offline SDA card. Note that this attack does not work for DDA cards.

However, the specifications for M/Chip (MasterCard 2002) (used for MasterCard) don't allow the use of theLCOL, leaving only card-based risk analysis for making a card online-only. As we have seen, such an approach is always susceptible to the attack by modifying the application. (The M/Chip specifications do define the LCOL, but only as private pa-

rameter used internally by the card.)
It can be noted that inclusion of the LCOL in the data signed with SDA is not in the published recommendations. One step for reducing the potential threat is to update the recommendations to include the LCOL and also note that it should be set to zero.

CONCLUSIONS AND RECOMMENDATIONS

We have demonstrated how EMV cards with a certain configuration can be attacked, and we have also pointed out how to configure a card correctly to avoid this attack. We have seen that cards based on M/Chip cannot be configured in the proposed way, and are therefore always susceptible to the attack. Here the only solution is to move to (more expensive) DDA cards.
One possibility to solve the problem is the change the EMV specification so that a terminal always goes online when a non-DDA EMV card is used. Although the consequence is that issuers using low-cost card cannot benefit from the advantages of offline transaction, from a security perspective this approach would be the most efficient.

REFERENCES

EMV, 2000. *EMV2000 Integrated Circuit Card Specifications for Payment Systems.* EMVCo. Available from `http://www.emvco.com` as of Sep 2003.

Jurgensen T. and Guthery S., 2002. *Smart Cards: The Developer's Toolkit.* Prentice Hall PTR, 1st ed. ISBN 0-13-093730-4.

MasterCard, 2001. *Terminal Requirements for Acceptance of Chip Pay Now (Debit) and Pay Later (Credit) Cards, version 4.0.* Available from `http://www.mastercardintl.com` as of Sep 2003.

MasterCard, 2002. *M/Chip 4 Card Application Specifications for Credit and Debit version 1.0.* Available from `http://www.mastercardintl.com` as of Sep 2003.

Visa, 2000. *VIS Visa Integrated Circuit Card 1.4.0.* Available from `http://international.visa.com` as of Sep 2003.

SECURING DATA BY STEGANOGRAPHY ON SMALL DEVICES

Alessandro Amoroso
Marco Masotti
Michele Pio Di Pasquale
Università di Bologna
Dipartimento di Scienze dell'Informazione
Mura Anteo Zamboni 7, 40127 Bologna, Italy
E-mail: {amoroso, masotti, dipasqua }@cs.unibo.it

KEYWORDS

Steganography, Smartphone, Security, MMS.

ABSTRACT

Steganography is a well known, and long standing, technique to hide data into data. The actual small devices, such as the smartphones, provide a complete set of capabilities to cipher and decipher text into pictures.

In this paper we propose an application that enables the user to exploit the capabilities of smartphones to hide and retrieve data into pictures taken by means of the onboard camera. The limited computational power and storage capacity of a small device are enough to allows their user to take advantage of a sophisticated technique such as the steganography.

By means of our application a smartphone becomes a self-contained device that enables the user to easily hide, store, and retrieve secure data. Moreover, those data can be securely sent and received by means of the Multimedia Messaging Service, MMS, a common technique to exchange images in the actual scenario of cellular telephony.

SCENARIO

The main idea of the steganography is to embed hidden data into other data, such that even the existence of the secret is concealed. The name is derived from Greek, and literally means "covered writing". Historically, the methods used to implement steganography included invisible inks, tiny pin punctures on chosen characters, grids that cover almost all the text but few characters, small differences between characters, and so on. The basic idea is that none can gather that a document is hiding a secret message; conversely, the receiver can extract the secret from the carrier, or cover, data.

The advent of computers improved the range of applicability or the steganography. Nowadays the kind of medium that can be used to embed a secret vary from pictures to music. Even the secret can be represented in any kind of medium. By means of complex computation it is possible to embed, or to detect, large secrets into huge carriers.

A branch of data hiding close to the steganography, is the *watermarking*, that has common usage in the protection of copyrights of digital media, such as art images, music, and video [Johnson, Duric and Jajodia 2000]. The main idea of watermarking is that a unique piece of data is embedded within a medium without noticeably altering it. At any time the watermarker can extract the secret data from the medium to prove their originality. The conceptual difference between the two branches of hiding data is the object of communication. In watermarking the carrier data are the communication, while the hidden data provide copyright protection; on the contrary, in steganography the secret is the communication, while the carrier data are a cover. Moreover, the presence of watermarking is often declared, and any attempt of remove it leads to a deterioration of the carrier quality [Wang and Wang 2004].

Another well known technique to store a secret is its *encryption*, *i.e* the technique to transformation data into different form. The resulting data can be understood only by knowing how to return them to their original form. The major difference with steganography is that the existence of the secret is not concealed; encrypted data manifest their existence. Encryption and steganography are aimed to different purposes. The encryption encodes data in a form that unintended observer cannot determine their meaning. Conversely, the steganography attempts to hide data such that an unintended observer does not suspect that any secret is concealed. To gain a higher decree of secrecy it is possible combine

the two techniques: the secret is encrypted before being embedded into the carrier data. Steganography relies on the same assumptions of the encryption, then they share the same constraints. They both require that the sender and the receiver of a document had previously agreed on the adopted method, and that they had securely exchanged some key [Artz 2001].

The aim of this work is to study the feasibility of using a small device as a complete platform to perform steganography. The actual smartphones are equipped with a camera, they store locally the photos and they have some power of computation. The screens of small devices are usually few hundred pixels per side, and can display, at the best, thousands colors. The CPUs of small devices are rarely equipped with floating point co–processors, and run at a pace that is orders of magnitude smaller with respect to the processor the PCs. Finally, the RAM onboard of the small devices is orders of magnitude smaller than the one into the PCs. The above limitations depict a scenario that resembles the early 90's of the last century. We would test the capability of such devices as complete steganography machine.

The smartphones are widely diffused, and the possibility of store small amount of secret data, such as passwords, SSN, ID, credit card number, into them seems to be very captive to a huge number of people. Moreover, the secure data stored in a smartphone can be securely sent, and received, by means of the MMSes. This method establish an hidden communication channel that actually seems to be unused.

The remainder of the paper is structured as follows. In the next section we present the main ideas of the steganography. The subsequent section describes a Java application that implements a steganographic method on a smartphone, and we discuss its performances. A Concluding Remarks Section ends the paper.

STEGANOGRAPHY AND STEGANALYSIS

The most important issues in steganography are: *security*, *payload*, and *robustness*. The security of the steganography means that the hidden contents is "invisible" both perceptually and statistically, *i.e.* an eavesdropper cannot detect the presence of a secret by using any accessible means. The payload, or *capacity*, of a carrier data is the amount of secret message that can be embedded into the carrier data. Usually those two issues are in contrast, and they require a trade-off. Finally, the robustness of a steganography means that it the secret can successfully resist to further elaboration of the carrier, such as compressions, *i.e.* after a decompression the secret is not corrupted and still accessible.

The capacity is limited by the size of the carrier data, and usually it is a small fraction of it. In other words, the amount of secret data is inversely proportional to the constraints on the integrity of the carrier data. Basically, the steganographic process is made by two step; it starts identifying the redundant bits in the cover medium, *i.e.* those that can be modified without "significantly" altering the medium integrity. Then, it generates the "stego" medium by replacing some of those bits with data from the secret. The significance of the alterations is crucial to the *steganalysis*, that is the countermeasure to steganography [Provos and Honeyman 2003].

If the embedding of a secret in a carried data leaves detectable traces, then it is possible a successful attack [Johnson and Jajodia 1998]. While a simple comparison with the unmodified carrier and the stego medium immediately reveals the changes, the original cover is kept secret. The steganalysis works in passive mode, it usually involves two major techniques: perceptual and statistical analysis [Fridrich et al. 2001]. A winning steganalysis could be the mere *detection* of the presence of a secret. As an example of such steganalysis applications see [Agaian and Rodriguez 2004].

Another important issue in steganalysis is its computational complexity. The detection of a stego data could mean the screening of a huge quantity of documents (even millions), then this operation has to be faster [Wang and Wang 2004].

In this paper we consider the application of steganography to pictures, and more precisely to photos. The bitmap format of pictures is not commonly used due to the size of the resulting files. A widely adopted compression format to store and send picture is the JPEG format [Wallace 1991]. Due to the compression algorithm, any JPEG image has a known structure, therefore even a slight modification of it would be easily detectable. Moreover, the JPEG is a *lossy* compressed format, *i.e.* it provides a close approximation of the original image, but not exactly the same. The approximation introduced by the JPEG compression could destroy the embedded secret.

PROPOSED APPLICATION

We propose an application that implements something in the middle between watermarking and steganography. The key idea is to encipher small payloads in pictures, such as the former, keeping it secret, such as in the latter. We put a control into the application, that refuses payload bigger than a certain percentage of the cover data. This method protects distracted users against "unsecure" steganography.

The idea of our application is that instead of openly crypt some sensible data we hide them inside the picture stored in the smartphone. If someone take possession of our phone they cannot detect and decrypt our data.

The number of steganography methods that can take advantage of the JPEG format is low. Nevertheless, we need to apply some of those because JPEG is the most common format for store picture on the smartphones. After a screening of the available steganography methods, we selected the F5 [Westfeld 2001].

Our application is based on the implementation of F5 released by its author, that can be downloaded at `wwwrn.inf.tu-dresden.de/~westfeld/f5.html` The F5 steganography method takes advantage of the JPEG coding, adding further steps during the process.

Westfeld suggests that the maximum size of the secret embedded by means of F5 should be no more than 14% of the cover size to be undetected by visual analysis. Moreover, he suggests that a secret smaller than 1% of the cover should be undetected under any kind of steganalysis [Westfeld 2001].

EXPERIMENTAL ANALYSIS

The aim of our tests was to measure how long the program runs both to cipher, both to decipher secrets into several images. We would like to test the feasibility of such an application in the area of small devices. Running on the experimental device, the application was unable to manage a cover image bigger than 20KB, due to "memory full" problems. On the contrary, the same program does not exhibit the same behavior running on a personal computer. The image size seems to be a stiff constraint, but it still offers enough room both to store significant photos (see Fig. 1), both to embed small data, such as credit card numbers, PIN, and passwords. Nevertheless, we will try to implement a new version of the program with a more sparing memory demand.

The Java application has been tested on a smartphone Sony–Ericsson P800, operating under the Symbian 7.0 – UIQ 2.0. The standard software of the P800 includes the J2ME Virtual Machine. Table 1 (source `http://my-symbian.com`) shows some technical specification of the device in comparisons with respect to other smartpones. The P800 offers the same computational power of the other smartphones, and its display is one of the less performing in the category. The built-in camera is standard for its category.

In our tests we measured the time to encipher and decipher a secret into a carrier message. To keep the secret about 1% of the cover, we tested four pure ASCII files, whose size was 50, 10, 200, and 400 bytes. The latter

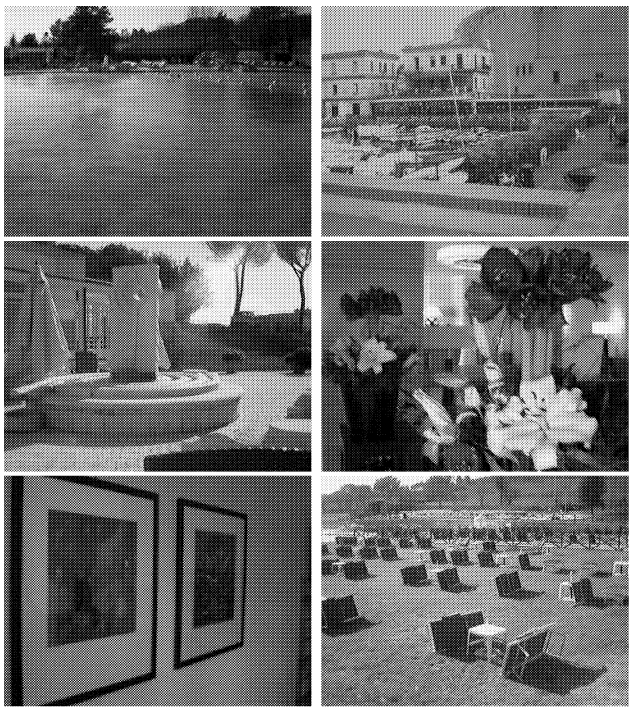


Figure 1: Spa (11,440 Byte), Naples (14,740 Byte), Fountain (14,861 Byte), Flowers (15,243 Byte), Paints (17,862 Byte), Shore (19,477 Byte)

represented about 2% of the biggest cover image. We tested the program on several cover photos, stored as JPEG files; some of them are shown in Fig. 1. Tose pictures have been shot with the internal camera of the P800, their size is 320×240 pixels at 16 Millions of colors, the captions report the size, in byte, of each file. Those pictures show the efficacy even of such a small photo size.

The preliminary results show that the target smartphone could be used as a standalone steganography machine. Fig. 2 shows, in the upper graph, how long the program takes to cipher a secret into the picture shown in Fig. 1, while varying the secret size. Similarly, the lower graph of the same figure shows how long the program takes to decipher the embedded secret text form the cover photos.

In both those graphs the x axis represents the size of the secrets, and the colored columns represent the computation time for each photo, as shown in the legend. Roughly, the cipher time is close to 15 seconds in any case, while the decipher time is slightly more than 9 seconds. The variance of computational times due to the secret size is limited to few percentual points. While the maximum size of the secret remains bounded to 1% of the cover image, the computational times may vary by few tenths of second.

Table 1: Smartphones Comparison

	SonyEricsson P800	**SonyEricsson P800**	**Panasonic X700**	**Nokia 6630**	**Siemens SX1**	**Motorola A920/925**
Memory card slot	16 MB, up to 128 MB	32 MB, up to 1 GB supported	miniSD	64 MB card bundled	MMC	32 MB, up to 256MB
Internal storage	12 MB	64 MB	4 MB	10 MB	24 MB	8 MB
CPU - 32 bit RISC	ARM9 (156 MHz)	ARM9 based (156 MHz)	ARM9	ARM9	ARM9	ARM9 (120 MHz)
RAM	16 MB (7-9 available)	32 MB (∼21 available)	8 MB (∼4 available)	16 MB (∼10 available)	?	32 MB (21 available)
Display	4096 colors, 208x320 pixel	262000 colors, 208x320 pixel	65536 colors, 176x208 pixel	65536 colors, 176x208 pixel	65536 colors, 176x208 pixel	65536 colors, 208x320 pixel
Built-in camera	640x480 24-bit	640x480 24-bit	640x480 24-bit	1.3 Megapixels	640x480 24-bit	640x480 24-bit

Fig. 3 shows the enciphering time with respect to the shared secret key used to compute the steganography. The x axis shows the size of the key, while the colored columns represent the time to encipher a secret into the "shore"photo. The color of the columns represent the size of the secrets, as shown by the legend. The maximum variance of the computational times introduced by the key length is close to 3%.

We think that the running times presented above could be considered by the user as a reasonable price to pay in order to achieve high level of secrecy. The secret embedded in a cover can be securely stored, sent, and retrieved, within few seconds.

Considering that the maximum size for a standard SMS is 160 characters, the constraint mentioned above is not too limiting. It takes a small picture, *i.e.* a single MMS, to securely send the same amount of data.

Moreover, the key used to encipher the secret can be considerably shorter than the secret. This means that the steganographic approach can be used as a kind of "leverage" to to the human memory: remembering a shorter key allows the user to retrieve longer secrets.

CONCLUDING REMARKS

We shown that steganography can be an effective technique to hide small amount of data by means of smartphones. Those devices contains all the necessary capacity to accomplish every task of the steganographic process, from shooting the cover pictures to encipher, and decipher, secrets into them.

A stego-document could be used both to store secrets, both to send them by means of hidden channel. Keeping the payload smaller than 1% of the cover photo practically assure that the secret will be ignored by the steganalysis programs.

The telecom companies can block those possible secret channels by adopting a simple countermeasure. Since the steganography is not very robust, the telecom companies could embed small amount of background noise, or a recurring pattern, into any sent MMS, doing so leads to break any possible carried secret.

ACKNOWLEDGMENT

This work has been partially funded by the project *FIRB/"WebMinds"* of the Italian Ministry of Education, University and Research.

BIBLIOGRAPHY

[Agaian and Rodriguez 2004] Agaian S.S. and B.M. Rodriguez. 2004. "Steganalysis Based on Cluster Classification". In *Proceedings of the IASTED Internet and Multimedia System and Applications* (Kauai, Hwaii, USA, Aug 16-18), IASTED Zurich, Switzerland, 414-419.

[Artz 2001] Artz D. 2001. "Digital Steganography: Hiding Data Within Data". *IEEE Internet Computing*, (May-June), 75-80.

[Fridrich et al. 2001] Fridrich J.; Goljan M.; and R. Du. 2001. "Detecting LSB Steganography in Color and Gray–Scale Images". *IEEE MultiMedia*, (Oct-Dec), 22-28.

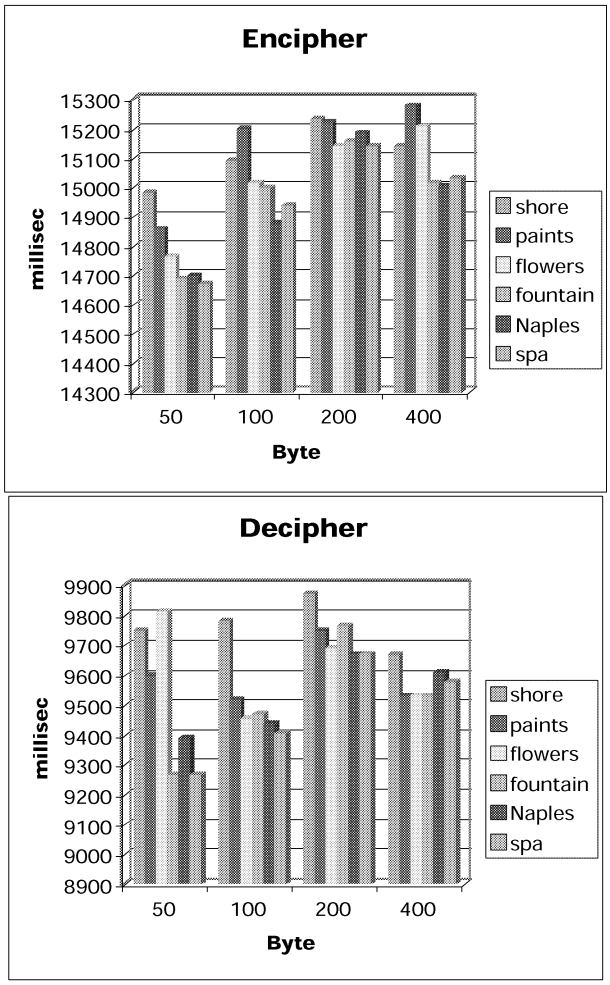


Figure 2: Enciphering and Deciphering Time for the Above Photos

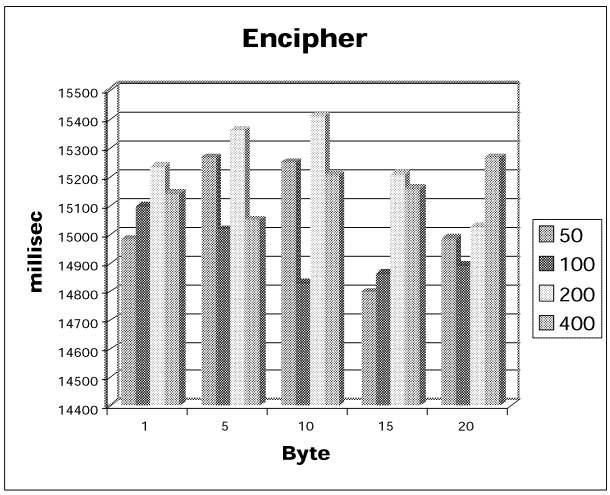


Figure 3: Enciphering Times for the "Shore" Photo by Varying the Length of the Secret Key and the Size of the Secret

[Johnson and Jajodia 1998] Johnson N.F. and S. Jajodia. 1998. "Exploring Steganography: Seeing the Unseen". *IEEE Computer* 31, No.2 (Feb), 26-34.

[Johnson, Duric and Jajodia 2000] Johnson N.F.; Duric Z.; and S. Jajodia. 2000. *Information Hiding: Steganography and Watermarking – Attacks and Countermeasures*. Kluwer Academic Publisher.

[Provos and Honeyman 2003] Provos N. and P. Honeyman. 2003. "Hide and Seek: An Introduction to Steganography". *IEEE Security & Privacy*, May-June, 32-44.

[Wallace 1991] Wallace J.K. 1991. "The JPEG Still Picture Compression Standard". *ACM Communications* 34, No.4, 31-44.

[Wang and Wang 2004] Wang H. and S. Wang . 2004. "Cyber Warfare: Steganography vs. Steganalysis". *ACM Communications* 47, No.10 (Oct), 76-82.

[Westfeld 2001] Westfeld, A. 2001. "F5-Steganographic algorithm: High capacity despite better steganalysis". In *Lecture Notes in Computer Science 2137*. SpringerVerlag, Berlin, 289302.

AUTHORS BIOGRAPHY

ALESSANDRO AMOROSO is Assistant Professor at the Department of Computer Science of the University of Bologna since 1994, and he got his laurea degree at the same university in 1987. The main research areas of Amoroso are the distributed systems and the multimedia. He participated to several scientific projects of National Research Council (CNR), National Energy Board (ENEA) and University of California at San Diego (UCSD).

MICHELE PIO DI PASQUALE got his degree in Computer Science at the University of Bologna in 2004, currently he his working as consultant in computers.

MARCO MASOTTI got his degree in Computer Science at the University of Bologna in 2003, currently he his involved on some research projects in the same university.

INTRUSION DETECTION VIA BEHAVIOURAL PROFILING ON MOBILE AND WIRELESS NETWORKED DEVICES

Ibrahim Zincir, Steven Furnell and Andrew Phippen
University Of Plymouth
Portland Square A304
Plymouth, PL4 8AA, UK
E-mail: izincir@plymouth.ac.uk, sfurnell@plymouth.ac.uk,
andy@jack.see.plymouth.ac.uk

KEYWORDS
Wireless, Behaviour, Profile, Security, IDS.

ABSTRACT

Wireless networks have gained an increasingly important role in our daily lives. They give us the much needed flexibility and mobility with one major concern, security. Since they use radio frequencies to transmit data from one node to another, the hacker does not have to gain physical access to the network wires and to pass through firewalls and gateways. As the threats are coming from multiple angles with a variety of different platforms, traditional security methods do not provide sufficient protection. Hence, more powerful solutions are needed. It is considered that using Behavioural-Based Intrusion Detection System (IDS) will assist the identification of malicious user activity, and help to improve the resulting security of wireless networks.

INTRODUCTION

The world has changed a lot since Marconi made a long trip from Italy to England to present his magic box to the British Telegraph authorities he developed to be used as a wireless telegraph system. One hundred years later, digital satellites broadcasting HD TV channels, mobile phones are connecting us to each other for any type of purpose a man could not even imagine a century ago, laptops surfing the Internet via hot spots, and PDAs checking emails at the airports are common scenes that we are all used to seeing. Today wireless is the shining star. The idea of connecting everything, anywhere at anytime pushes enterprises to pour more money than ever into wireless technologies. There are 1.5 billion cell phones in the world today, 3 times more than the number of computers and these mobile phones provides the equal amount of processor power to the desktops of 1990's (Newsweek 2004). Wireless technology is everywhere.

The major problem with them is as the information between the information is transmitted between the nodes through radio frequencies instead of cables, and the threats are envisaged to come from multiple angles with a variety of different platforms, traditional security methods will not provide sufficient protection against security challenges that is governed in wireless networks. We believe that using behavioural profiling as an Intrusion Detection System (IDS) to detect the malicious user/behaviour will help us to improve the security of wireless networks. In this system, historical user profiles are created and then compared with real-time activity in order to detect malicious behaviour.

BEHAVIOUR PROFILING

A behaviour profile is a collection of information that can be used to describe basic characteristics of an individual defined under the rules created by the system administrator. These specifications are designed depending on for when, where and how the profile will be used. In a way a profile can be seen as a business card as it contains some basic definitions such as name, location, and phone number of the user. In order to create stable profiles one must not include every bit of information but must use sensible specifications to enable a common way of defining the behaviour.

One of the biggest problems of modern IDSs is that they try to catch the malicious activity by analysing the user signatures and/or detecting the misuse. In misuse detection, the monitoring system looks for an activity that corresponds to known intrusion signatures or system vulnerabilities; it monitors for explicit patterns and diagnoses the specific attacks from the signatures. The major problem for this method is how to differ the intrusive signatures from the right ones, and how to include all variations of attacks into a signature. Moreover, this type of a technique explicitly works on known attack patterns. This results in the IDS remaining vulnerable to unknown attack patterns (McHugh et al. 2000, Brox 2002, Babaoglu 2003). In order to detect novel and unknown attacks a Behaviour Based IDS implements not only misuse tools but also an anomaly detection approach. Anomaly means something that is not nominal or normal. In anomaly detection, the intrusions are detected by looking for an activity that is different from a user's or system's normal behaviour. The system creates a normal activity profile and then compares this profile with the ones that are connected to the network in real time. As a result, it has the ability to detect symptoms of new kinds of threats. The major problem for this system is it may misjudge the normal activities as intrusive, and the intrusive activities as normal. Thus, the profiling must be done with great accuracy and detection criteria have to be made carefully (McHugh et al. 2000, Brox 2002,Babaoglu 2003). In addition, during the process, the system has to work as fast as possible since the malicious activity has to be detected as quickly as possible once it occurs.

Behaviour-Based intrusion detection techniques detect an intrusion by observing a deviation from the expected

behaviour of the user. The model of a valid behaviour is extracted from the collected information. User profiling means developing a profile for a user of a network. The profile is built by using history of events and actions. The detection system first monitors and records the user's behaviour. Then, by examining these files, it creates a specific profile for that specific user. Whenever the user connects to the network the system then compares in order to check whether the user is who he claims to be. The intrusion detection system later compares this profile with the current one. When a deviation is observed, an alarm is generated. In other words, anything that does not match to the expected behaviour is considered potentially intrusive. Therefore, the intrusion detection system may be much powerful, but its accuracy will be a question mark since there will be a lot of false alarms.

The advantages of Behaviour-Based IDS are numerous (Lawrence Livermore National Laboratory 1996); they can detect attacks to new vulnerabilities and they are not dependent upon operating systems and they can also detect misuse. However, there are also several challenges to overcome before implementing such a system. Wireless intrusion detection is a very new technology and hence before applying it to any network one must be very careful since there may be bugs and vulnerabilities that can create the biggest threats. Finally, an IDS is only as effective as the individuals who analyze and respond to the data collected by the system. A Behaviour-Based IDS, like a standard IDS, can require vast human resources to analyze and respond to threat detection. There are four main rules to be able to deal with these risks: know & plan, protect communications, protect wireless devices, protect airwaves and monitor 24/7.

The biggest problem with this system is anything that has not been met previously is accepted as intrusive. As a result, it causes a lot of false alarms which in turn results a drawback (Newman et al. 2002, Timm 2001). In addition, creating a profile for a user is not that easy. The whole specifications of a person cannot be learned in a period of time. Also, behaviour can easily change by time, and the user can gain new habits. This creates another challenge since the system has to update itself from time to time to learn the new user profile, and between these periods more false alarms and vulnerabilities occur since an attack during the learning procedure may create more headaches.

There are two types of error as a result of the comparison of the profiles: false positives and false negatives. False positive means a registered user is identified as malicious user after the process of matching the profiles. False negative means a malicious user is identified as a registered one after the process of matching the profiles. To have a successful IDS both of these rates have to be lower. Specifically the false negative has to be zero since a malicious user cannot be given access to the network. To solve these problems, the IDS must be designed to have the ability to update itself periodically regarding to environment since the behaviour profile of a user may change over time. Otherwise, the system will lose its credibility and instead of protecting the network, it will damage the system. Thus, the IDS has to continue to obtain real time data to update the user profiles. A major concern in this approach is how the mechanism will work when a change in a profile happens. When will it reject and when will it accept it? It cannot continuously update itself with every new activity of the user. How long should the learning period be? Another problem is how the system will protect itself from a malicious user who will try to train the system by time to impersonate a registered user. How can the system react to this?

Response to the malicious activities is another important issue in Intrusion Detection Systems. Usually, system administrators prefer passive response systems as they want to see the alert and take the appropriate action by themselves, in order to prevent adverse effects upon legitimate users in the event of false alarms (Papadaki 2004). Once the administrator is convinced that there may be a malicious user inside the network, he then may ask the system to send a text or email to that user asking his specific password and/or PIN number especially created for situations like this. One last challenge for behaviour profiling is the privacy issue. As a lot personal information may be collected about the registered user, the system will be totally automated and all the log entries will be encrypted and no one will have the access to read these log entries.

In short, there are many important challenges in designing a Behaviour-Based IDS; but we believe by creating a successful profiling mechanism we will be able deal with all of these problems.

CREATING BEHAVIOUR PROFILES

There is no opposition against the idea that wireless networks are not totally secured against targeted attacks. As the sensitive information theft becomes the most growing concern for the enterprises it is inevitable that there may not be any attacks. As a result, it is important to develop effective security solutions and policies. Since the medium cannot be controlled properly, wireless networks bring big risks as much as big advantages. However, with the solutions and the policies they can still be secured. Using a successful IDS that monitors and analyze the traffic inside the network by 24/7 will help the solution to the problem. To this end, our project will create historic profiles of the clients and then compare them with the real-time ones in order to detect malicious activity. As explained above, choosing which characteristics of a registered user to audit in order to create historical profiles is a major challenge. In our project, our IDS will collect data both on the mobile device and on the network. The service provider will install a program which will log the information regarding how the device is used while it is not using the wireless network, as well as it will continuously monitor the user whenever he or she is connected to the network.

Our Behaviour Based IDS will be implemented on a mobile device (e.g. a smart phone) that will have the ability to operate under other wireless networks (WiFi, WiMax, and/or Bluetooth) as well as the network service provided by the operator (GSM, GPRS and/or 3G/4G). Mobile devices have gained an increasingly important role in our daily lives as

675 million phone sales is expected by 2006 while 30 million Personal Digital Assistants (PDAs)(Reed 2002). We use them for many different purposes as they provide flexibility and mobility. Since few individuals like to share their mobile phone with others they are very personal, and as they are regularly connected to the network they are easily traceable. As a result, the way they are used is very unique depending on the owner and there are many different characteristics that can be used to define a profile for the user.

Today, mobile telecommunication companies keep records of their customers' behaviours for many different purposes. In 2001 Ericsson conducted research similar to our project, trying to detect fraud use of the network by creating user behavioural profiles. The system uses pattern recognition software built into intelligent agents-called sentinels-that assign behavioural profiles of subscribers on a network. If the software detects unusual activity on an account, it will send a text message to the mobile phone, the user will then have to punch in a PIN to identify themselves, if they fail to do so, the phone will be cut off (Rowe 2001).

According to The Register, a website for latest technological inventions, American researches Nathan Eagle and Sandy Pentland from Massachusetts Institute of Technology are developing applications for mobile phones which can learn user's daily habits so that they can become mobile digital secretaries. The smart phone learns about owner's behaviour by logging calls and recording when the digital camera implemented on the phone is used. In addition, it takes note of Bluetooth pairing bonds in order to understand who the user socialise with. The software has been installed on 100 Nokia 6600 and the collected information is downloaded onto a server in MIT. The results will be used to investigate how social networks build as well as technologies (Leyden 2004). Another research is made again by Nathan Eagle and Sandy Pentland in order to explore a mobile device user's situation by recording his conversations with the individuals around him and the commands he gives to his mobile device while he is waiting in the queue in a restaurant. The device then analyse these conversations and commands in order to offer different solutions to the user (Eagle and Pentland 2003). According to Schilke et al, the biggest problem with the personalisation is there is too much information for users to be deal with, and they cannot access it at the right time and at the right place as well as at the right format. The main dimensions are time, location and interest. The time dimension is the user's repeating behaviour, the location dimension is where the user is, and the interest dimension is understanding what the user is doing and why (Schilke et al. 2004).

Other research conducted in order to understand whether the different cultures really have different ways of using mobile devices shows that approximately 35% of the users are using wireless networks only to download files and to send emails (Lee et al. 2002).

According to Donner (2004), it would not be wrong to describe the call register (the log of the incoming and outgoing calls) is a distinct characteristic of a mobile phone user. Donner's research showed that nearly 66% of the outgoing calls were made to family members and friends and 65% of incoming calls were made by the family members and friends were answered.

Another application that tries to predict user's behaviour is created by Zeynep Inanoglu and Ron Caneel of the Media Lab at the Massachusetts Institute Of Technology. It is a voicemail system called Emotive Alert that labels messages according to the caller's tone of voice in order to identify which messages are the most urgent (Biever 2005).

These findings suggest that there are many different characteristics of wireless device owner's that can be used to define as their profiles. In addition to which applications are used at when and where also how the device is used to interact with other devices and networks are going to be taken into consideration in our project. Table 1 describes the some of the basic characteristics that can be used in the behavioural profiling.

Table 1: Basic Characteristics Of Mobile Devices' Users

Outgoing Calls
Incoming Calls
Video Calls
Location
Time
SMS
MMS
Favourite Websites
Email Addresses
IP Address Of Access Points
Bluetooth ID
Digital Camera Usage
Key Stroke
Voice Commands
Opened Applications
Media Player
Downloaded Files

IMPLEMENTING A BEHAVIOUR-BASED IDS

In order to design a novel structure of a Behaviour-Based Intrusion Detection System, the first step is to choose what to log on both at the mobile device and the network. Since the IDS will be designed for a mobile device which can be active in all the known wireless networks and bandwidths, there are too many different kinds of information regarding to the user to be logged. After analysing the other behavioural profiling research summarized above, the most accepted distinctive characteristics of the users are chosen to be collected.

To prevent too much memory consumption and processor speed, the system will record only very basic processes inside the mobile device. The program installed by the network service provider will log the Bluetooth ID of the connection point when the device is connected to other networks through Bluetooth and it will log the IP address of the access point if it is connected to other networks through

any IEEE 802.1X technologies. It will also record the addresses that emails are sent and the websites that are visited during the connection to these networks. When the device is connected to its registered network, all the outgoing and incoming calls, the numbers that the SMS and MMS messages are sent, all the addresses of the emails that are sent and finally all the websites that are visited will be logged by the network. Of course, all these log entries will include the time and the location dimensions. In the end, this collected data will then be combined and used to create behaviour profiles for the registered users. Table 2 shows a basic explanation of this.

Table 2: Data To Be Collected

Mobile Device		Network
Bluetooth	IEEE 802.1X	GSM/GPRS/3G/4G
Bluetooth ID	*IP Address*	*SMS, MMS Messages*
Email Addresses	*Email Addresses*	*Email Addresses*
Websites visited	*Websites visited*	*Websites visited*
		Call Register

In our project, we will first implement the program into a mobile device which will be used to collect information about different people who will use the device. Then the program will record the data regarding to the specifications described in the table above. We will then download this data into our server in order to analyse this information. During this analysing process we will try to understand and create characteristics for each user such as: who does he calls first, which web sites he browses most, which numbers he calls most and at what time, what is the general time he spends on calls, how often does he use SMS and MMS and to whom does he send, whether he uses his phone like a PDA, when and how, to whom he sends SMS and/MMS messages, does he make any video calls and if so to whom, does he use internet from the service operators or does he use his own wireless connections such as Bluetooth, or WiFi to browse the net, and if he does surf the Internet which web sites does he prefer to visit most. When the device uses its WiFi abilities to connect to a computer network, what is the IP address of the servers, how long does the connection to the network take place? If Bluetooth technology is used, what is the Bluetooth ID? In the end, we will create a user profile for each of the specific user of the system. This user profile will be updated regularly in order to adapt itself to the changing behaviour of the user.

The IDS that is going to be designed, as explained, will store the collected information inside the service provider's network and will monitor the traffic continuously. There will not be any analyse on the user side and hence the decisions will be made much more quickly and accurately. The IDS will learn to predict the next action of the user by comparing the historical profile (that consists of applications being run, usage times, types and combination of tasks, process durations, types of files used and etc. for a given time period) with the real-time one of the same specific user. The system will constantly observe the new trend and will update the user behaviour over time. This update will be made regularly by a time basis of 3 months. The main problems will be how to react to a rapidly changing behaviour and how to decide the 'normal' behaviour. In order to have a standard the system will carry out the following steps in order to identify the malicious activities:

1) When the mobile device is activated inside the network the system will begin to collect real-time data in order to create a real-time profile
2) This real-time profile will then be compared to the historical profile in order to catch malicious activity
3) If the two profiles (historical & real-time) will match then the user will be accepted as legitimate and the system will continue to monitor the user in specific time intervals
4) If the two profiles (historical & real-time) do not match then the user will be regarded as suspicious and the system will ask the device to send the data that it is collected
5) Then this data will be compared with the baseline profile created as a safeguard
6) If the two profiles (baseline & real-time) will match then the user will be accepted as a legitimate and the system will continue to monitor the at specific time intervals
7) If the two profiles (baseline & real-time) do not match then the user will be regarded as malicious and it will alert the system administrator to make the appropriate response

Figure 1 gives a brief explanation of how this will work:

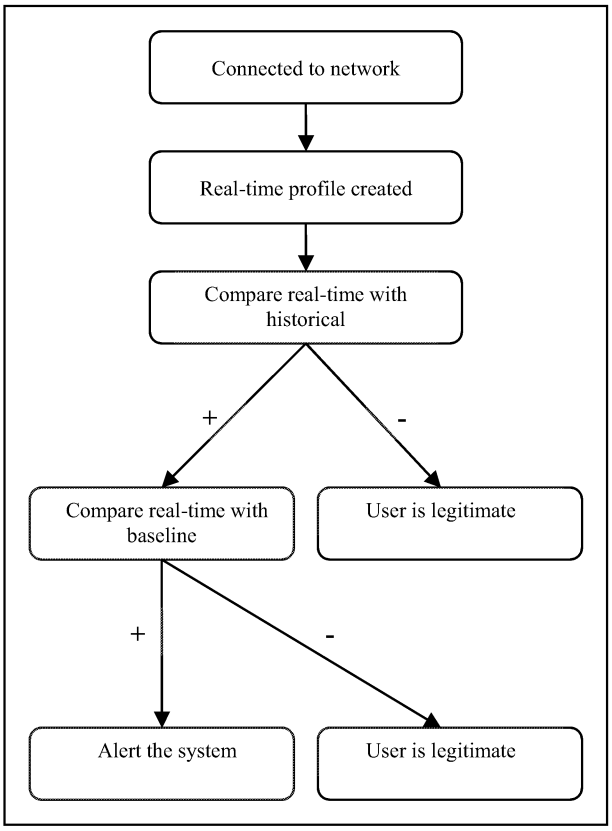


Figure 1: Architecture Of Behaviour-Based IDS

In our project, the profiles will be compared by using the artificial intelligence methods to catch the malicious activity inside the network. Then, once the user is identified as

suspicious, the system can begin to watch the activities closer and then by again making comparisons, it can alert the system in case of need. It can then respond according to the decision of the service provider. The IDS will use dynamic profiling and the host and the network systems will act together in order to catch intrusive behaviours.

Figure 2 shows the basic structure of the Intrusion Detection System based on behavioural profiling.

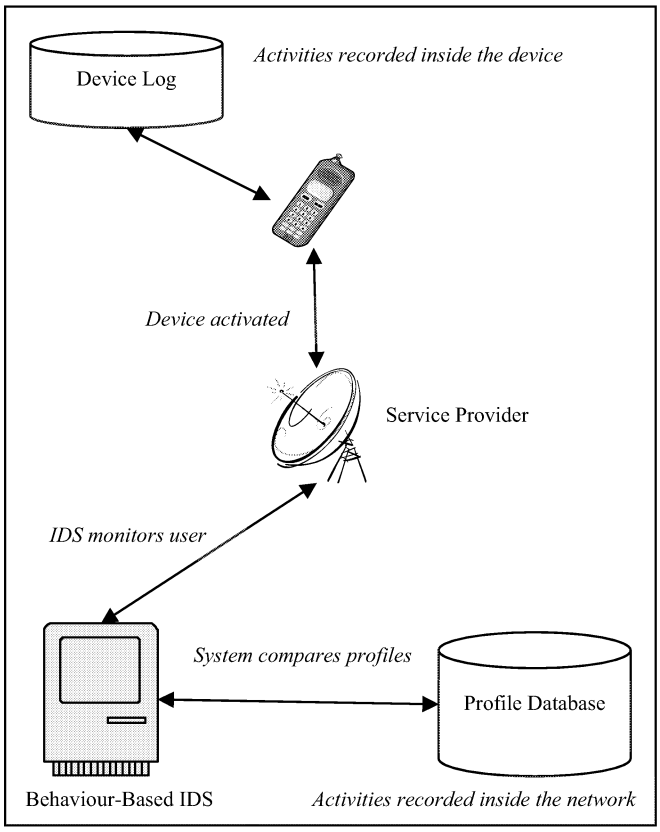


Figure 2: Behaviour-Based IDS

CONCLUSION

In our project, the profiles will be compared using a neural network technique to catch the malicious activity inside the network. Then, once the user is identified as suspicious, the system can begin to watch the activities closer and then by again making comparisons, it can alert the system in case of need. It can then respond according to the decision of the service provider. The advantage of behaviour based IDS is it provides better protection against new attacks that are not known or met. Since many activities can be recorded in order to create a much healthier behavioural profile it is not very easy for any hacker to guess which ones are important. Also, by updating the profiles regularly in large time intervals such as three months, it would not be easy for a malicious user to train the system by time.

The aim of the project, as explained, is the design of a stable, durable and trustable Intrusion Detection System that has the ability to adapt itself into new challenges and attacks supporting minimum false positive alerts. The authors are currently trying to implement a program into the mobile devices in order to collect data to be able to create profiles. After this step, we will create the novel architecture that will then be followed by the design of the prototype.

REFERENCES

Babaoglu O., 2003, "IDS: Intrusion Detection Systems", January 2003.http://www.cs.unibo.it/babaoglu/courses/security/lucidi/IDS.pdf

Biever C., 2005, "Voicemail Software Recognises Callers' Emotions", *New Scientist*, 8 January 2005. http://www.newscientist.com/article.ns?id=mg18524813.100

Brox A., 2002, "Signature Based or Anomaly Based Intrusion Detection: The Practice and Pitfalls", February 2002, http://www.itsecurity.com/papers/proseq1.htm

Donner J., 2004, "The Mobile Behaviours", *The Global and The Local In Mobile Communication Conference*, Hungary, June 2004, http://www.columbia.edu/~jd2210/whowhy.pdf

Eagle N. and Pentland S., 2003, "Handhelds That Listen And Learn", *IEEE Computer Magazine, Special Issue on Handheld Computing.* September 2003.

Lawrence Livermore National Laboratory, 1996, "Intrusion Detection And Response", *National Info-Sec Technical Baseline*, http://all.net/journal/ntb/ids.html

Lee Y., Lee I., Kim J. and Kim H., 2002, "A Cross-Cultural Study On The Value Structure Of Mobile Internet Usage", *Journal Of Electronic Commerce Research*, vol. 3, no. 4.

Leyden J., 2004, "Smart Phone Predicts Owner's Behaviour", *The Register*, 25 November 2004, http://www.theregister.co.uk/2004/11/25/super_smart_phone

McHugh J., Christie A. and Allen J., 2000, "Defending Yourself: The Role Of Intrusion Detection Systems", *IEEE Software,* September 2000, pp. 42-51.

Newman D., Snyder J. and Thayer R., 2002, "Crying Wolf: False Alarms Hide Attacks", *Network World*, 24 June 2002, http://www.nwfusion.com/cgi-bin/mailto/x.cgi

Newsweek, 2004, "Wireless Is Everywhere", *Newsweek Magazine*, 7 June 2004, pp. 68.

Papadaki, M. 2004. *Classifying and Responding to Network Intrusions*, PhD thesis, University of Plymouth, Plymouth, UK.

Reed Electronic Research, 2002, "The Mobile Industry- A Strategic Overview", October 2002, http://www.rer.co.uk

Rowe G., 2001, "Something In The Way She Phones", *New Scientist,* February 2001.

Schilke S. W., Bleimann U., Furnell S.M. and Phippen A., 2004, "Multi-Dimensional Personalisation For The Online And Offline World", *Internet Research*, vol. 14, no. 5, pp. 379-385.

Timm K., 2001, "Strategies To Reduce False Positives And False Negatives In NIDS", 11 September 2001, http://www.securityfocus.com/printable/infocus/1463

MEDIATEC

MULTIMEDIA COMMUNICATION TOOLS

Creation and Application of Metadata for Content Production in Broadcasting Stations
-- A Case Study of Metadata using high school baseball games in Japan--

Tomohisa Akafuji
Asahi Broadcasting Corporation
2-2 Oyodo-minami, Kitaku,
Osaka, 531-8501, Japan
E-mail: aka@abc.asahi.co.jp

Kaname Harumoto
Osaka University
2-1 Yamadaoka, Suita
Osaka, 565-0871, Japan
E-mail: harumoto@eng.osaka-u.ac.jp

Mitsuru Yasukata Yozo Yamamoto Masaaki Kobayashi
Panasonic AVC Networks Company; Matsushita Electric Industrial Co., Ltd.
2-15 Matsuba-cho, Kadoma, Osaka, 571-8503, Japan
yasukata.mitsuru@jp.panasonic.com yamamoto.yozo@jp.panasonic.com kobayashi.mark@jp.panasonic.com

Keishi Kandori
Asahi Broadcasting Corporation
2-2 Oyodo-minami, Kitaku,
Osaka, 531-8501, Japan
E-mail: kk@abc.asahi.co.jp

Shinji Shimojo
Osaka University
5-1 Mihogaoka, Ibaraki
Osaka, 567-0047, Japan
E-mail: shimojo@cmc.osaka-u.ac.jp

KEYWORDS
Metadata, Content Production, Internet, Annotation

ABSTRACT

The progress of digitization in broadcasting stations enables the production of advanced content from video materials. Metadata plays an important role in describing the semantics of content. We focus on the significance of metadata for multimedia content production in broadcasting stations. A method of sharing valuable metadata in content production is proposed and a method of annotating metadata for multimedia content is introduced as examples of the utilization of metadata. In this paper, the knowledge and experience obtained through the implementation of system and content production using real baseball contents is described.

METADATA IN BROADCASTING STATIONS

The digitization of video material in broadcasting stations makes storing and processing video contents by computer easy. Here, metadata attached video data are classified into three groups according to each property of metadata.

The first group includes information about the date, hour, and location of the video materials that were shot. A UMID (Society of Motion Picture and Television Engineering; Unique Material Identifier) is standardized metadata. Such video information is automatically generated by the UMID-enabled systems in the video and audio frames. Because the accuracy of such basic data is precise, it is reasonable to assume that the UMID can be built into devices to create metadata.

The second group includes the semantics of content for editing support and for creating high-quality content. The handling of this metadata differs from purpose to purpose. Furthermore, what we call a "*killer application*" using this metadata is desired. We focus our attention on this second group of metadata. The creation and application of the metadata are especially considered. Metadata of each group is discussed regarding the standardization for content exchange and content sharing. Because the representation and standardization of metadata are not essential problems to this paper, they will not be discussed in detail.

The last group concerns copyright and the conditions for content reuse. One reason why the distribution of high-quality content is not promoted, is technical problems concerning DRM (Digital Rights Management) and security. Another reason is that there are various frameworks of content production for broadcast and their characteristics complicate copyrights management.

From the viewpoint of the utilization of metadata, we consider the creation of applications and content using metadata of the second group in broadcasting stations to be insufficient still. The reason for this insufficiency is a problem of cost-benefit performance in the handling of the metadata. The creation of metadata and the standardization of metadata are sometimes mixed up which also results in cost problems. We should deal with these two problems separately in the current situation where the availability of metadata such as RDF (W3C) is not proved.

In this paper, an approach is considered to the utilization of metadata with emphasis not only on a structured approach, but also on a bottom-up approach such as content production. In broadcasting stations, valuable metadata have been accumulated for a long time. We try to create content for the Internet by sharing and creating metadata using real baseball game content. This paper contributes to the accumulation of knowledge and experience in multimedia content production using metadata.

CONTENT OF BASEBALL GAMES

Asahi Broadcasting Corporation broadcasts the National High School Baseball Tournament in Japan via terrestrial television, satellite, radio, and Internet. Our web site is the only site from which the high school baseball games are broadcast live on the Internet. Various contents of baseball

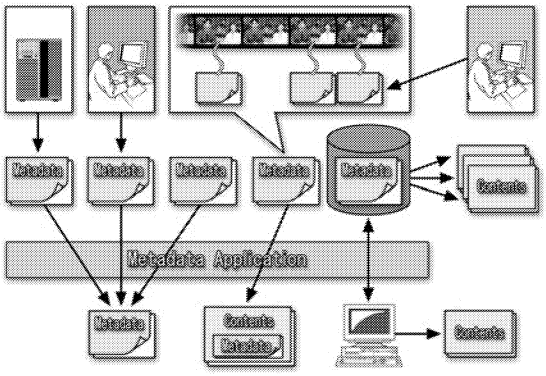


Figure 1: The relationship between content and metadata

games are provided on our web site. The types of content include score, players' statistics and information, and are created using video, audio, and metadata for broadcast. The advantages of broadcasting stations, which have valuable contents or metadata, are maximized.

Our content is created from diverse metadata for each purpose. These metadata are created by reusing metadata for broadcast, by combining various metadata, and by human annotation. This relationship between content and metadata is shown in Figure 1.

Score content

Acquisition of score data

Metadata for broadcast tends to be created and stored separately according to content. We developed new mechanisms for sharing metadata. Metadata about the semantics of content can be shared among diverse media. The purpose of sharing metadata among broadcasting stations is not only for cost reduction, but also because these data are always accurate and kept up-to-date. Score content of baseball is a good example regarding sharing metadata.

In general, the most important metadata of baseball content contains the score. Score data are created for broadcast and include running score, inning, SBO, and batting lineup. These data are very reliable and valuable because these data are used for CG displayed on the TV screen and data-casting. These data are also available for the Internet without error or delay.

A new database is created for sharing metadata and is connected to the CG system by a database link. The database is separate from the CG system for broadcast so that web casting does not have a negative effect on the system of broadcast. The timing for obtaining data differs depending on each metadata. For example, SBO and the running score are obtained every three seconds and the player lineup is obtained every 10 seconds. Score content is shown in Figure 2.

Consistency for contents creation

A handling mechanism for the consistency of each metadata is required for content creation using various metadata,



Figure 2: Score content of baseball games (In Japanese)

because the generation process and retention cycle differ for each metadata. For example, a maximum of 4 games are performed per day in this tournament. The switch timing of metadata is necessary for switching the content for each game. The switch timing depends largely on the broadcast time for TV. Metadata for the next broadcast is generated in the period between games when broadcasting does not occur. This metadata may be incomplete. A synchronization mechanism for the metadata matched to the game progress is needed. The procedure of the synchronization mechanism is shown in Figure 3.

(1) Metadata synchronization for the broadcast and the Internet
(2) End processing for a game
(3) Same as (1)
(4) Same as (2)

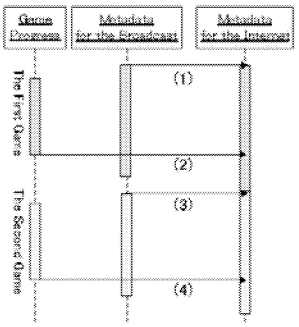


Figure 3: The procedure of metadata synchronization

On addition, the procedure for score content creation is shown in Figure 4.

(1) Game data (game date and the Nth game) are acquired
(2) Consistency check and synchronization of metadata
(3) SBO data are acquired
(4) Running score data are acquired
(5) Check for existence or non-existence of highlights
(6) Running score content is transferred
(7) Game data are acquired
(8) Consistency check and synchronization of metadata
(9) Player lineup is acquired
(10) Players' data are acquired

(11) Current player check
(12) Players' data are acquired
(13) Players' content are transferred
Here, players' data include the players' name, batting average, and earned run average.

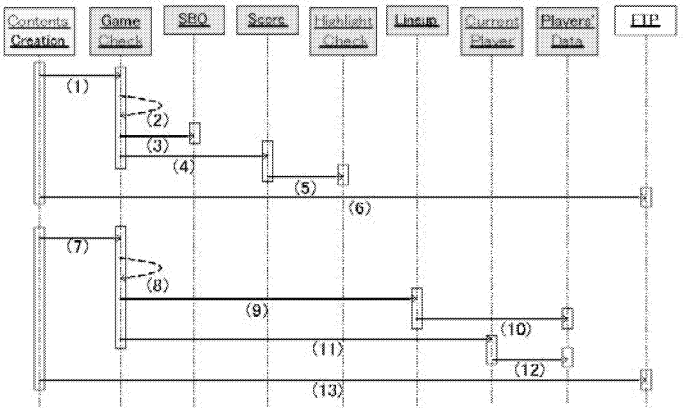


Figure 4: The procedure for score contents creation

Highlight content

Highlight content provides an important scene in a game. Highlights often include a particularly good play (a rare catch, stealing a base, etc.), a hit, or a homerun. About 20-30 highlighted scenes are provided as video-on-demand content for each game on our web site. When a play to be highlighted is found, the operator edits this scene in DV quality using a video editing system. An AVI file is made from a video file of DV quality by exporting using an editing system. AVI files may also be converted to streaming media files, such as for a PC, cellular phones or mobile devices and shared for these streaming media. This also allows for cost reduction and sharing of the operator's environment for content creation.

Highlight content is created from these streaming media files and attached metadata. Here, this metadata include the description of the scene and is created by human annotation. The procedure for highlight content in our system is shown in Figure 5.

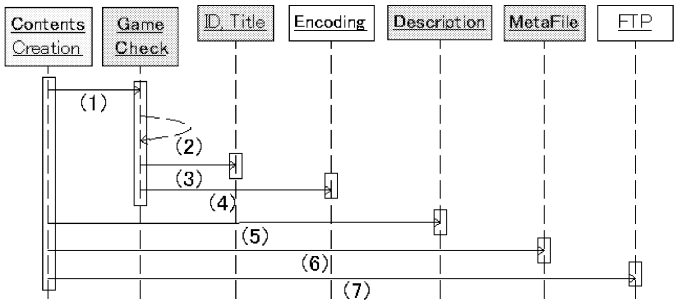


Figure 5: The procedure for creating highlight content

(1) Game data are acquired
(2) Consistency check and synchronization of metadata
(3) Tagging with a unique ID and title entry
(4) Encoding from AVI file into streaming media files (rm, wmv)
(5) Annotation for description of the highlighted scene
(6) Creating meta files for play streaming media
(7) Highlight content is transferred
We create highlight content by integrating various metadata and media files.

Retrieval for video scene
This content is what an Internet user retrieves while watching his or her favorite scenes. Highlights are only available for certain scenes, which are prepared by content providers such as broadcasting stations. However, users can select the scenes they desire by video retrieval.

For this type of content, metadata for each event regarding a baseball game must be created using a certain method. The contents of the live broadcast and the highlights are provided during a game. The video retrieval content is opened soon after a game. All processes for these contents during the game must be completed. The most important process is an effective annotation method for metadata creation in real-time. The detail of this type of content, or video retrieval content for baseball games, is considered in the next chapter.

METADATA CREATION AND APPLICATION FOR SCENE RETRIEVAL

Metadata for scene retrieval

Metadata attached video of baseball games enables users to retrieve and watch their favorite scenes via the Internet. Here, the metadata is the description of the semantics of a game's scenes. The unit size of metadata is fine enough to annotate events of the game, but a manual annotation method is required because computers cannot understand the semantics of a video scene. For example, referee judgment calls about whether or not a ball is foul or the direction of a batted ball require manual annotation.

Voice tagging system

Some trials have been performed to create metadata for baseball content in the past. One such trial is an annotation system for baseball content using an external database for input support. Common problems for such systems are the speed and accuracy of the annotations. (Akafuji et al., 2004)

We utilize a Voice Tagging System (Kobayashi 2004) for annotation of metadata. The Voice Tagging System is an annotation system using speech recognition developed by Matsushita Electric Industrial Co., Ltd. The Voice Tagging System resolves these problems with a user-friendly simple voice interface.

Speech recognition for this system uses keywords, rather than sentences or context. Therefore, the recognition rate can be made dramatically high by creating a keyword dictionary for specific contents. Such keyword dictionaries are prepared for such sport events as baseball, American football, golf, etc. This system has a linked time code on the video material and the moment of utterance. The metadata synchronized with

the event time can be attached effectively. The screen image of Voice Tagging System is shown in Figures 6.

Policy for the annotation of metadata

Content production consists of numerous factors. An empirical method needs to be introduced after taking into account real conditions, such as the necessity and cost for content production. Various viewpoints for metadata exist (i.e., batters in a baseball game as well as the defense and audience are considered targets), and these viewpoints require the making of policies depending on the properties of each content. This policy eliminates the subjectivity of an annotator and creates the metadata according to the purpose of the contents, so limiting the target for an annotation is realistic. We performed preliminary training for input metadata in consideration of these problems to create the policies.



Figure 6: The screen image of Voice Tagging System

The configurations of this training are as follows.
[Date] 2004-8-13 The 3rd game.
[Reference Video and Audio] Normal video for TV and audio is only announcers' comment before mixing.
[Annotator] Two people (information science graduate student)

The system for content using the Voice Tagging System is shown in Figure 7 and Table 1. Metadata and the media file are synchronized using an NTP server.

Target of annotation and content creation
The target of an annotation is limited to a batter's event, reporting from the director and editor in charge of sports. This policy is common to the editing of baseball content. Two types of content using metadata are created using this policy. One is video retrieval of a batter's scene as previously noted, and another is an *inning movie*. An *inning movie* shows a video of each inning for the entire game. Information regarding time in metadata, i.e., the start time and duration of each inning, is used for creating this content.

Table 1: The system configuration

Server	Software, OS
Contents Server	PostgreSQL-7.4.3, RedHat Linux 9
WEB Server	Apache-1.3.31, PHP-4.3.5, x86 Solaris 8
NTP Server	NTP-4.1.1, x86 Solaris 8
VoiceTag PC	Voice Tagging System, WindowsXP

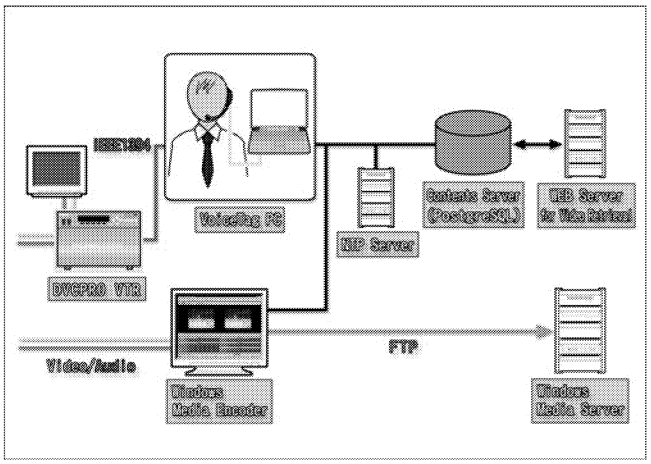


Figure 7: The whole system using Voice Tagging System

Structure of metadata for video retrieval
Metadata needs to structure each batters' information to extract various batters' scenes from the entire video. In this case, a batter's name and inning are used for the delimitation of metadata. For example, the following structured data are assumed;

```
<inning>
        <batter>
                <event/>
                <event/>
        </batter>
        <batter>
                <event/>
        <batter>
</inning>
<inning>
…….
…….
</inning>
```

Errors in speech recognition
These errors include accent of utterance and speech errors of the annotator. A simple adjustment to speech recognition is performed on the engine of the Voice Tagging System for each annotator. The Voice Tagging System can be reflected adjustments for errors of speech in a keyword dictionary, but although the results of this training indicate that practical exercises enabled improvement of the recognition rate and problem-free operation, an adjustment of the dictionary depending on the speech errors of each annotator were not performed.

Keywords and dictionary
An improvement in the accuracy of annotation data is related to the reduction of speech errors. In general, if fewer keywords are used, then it is easier for an annotator to operate this system and the system can contribute to the accuracy of the annotation data. In fact, annotators request decreasing the number of keywords in the dictionary. We have decreased the keywords from 400 to 200 by considering the necessity of video retrieval for this purpose.

Official operation and contents creation
The metadata and content using semi-final games (2004-8-21) and final game (2004-8-22) were created based on the

result of preliminary training. Annotation data are structured and stored in a database after a game. The interface for video retrieval is similar to the normal web retrieval system. The retrieval image is shown in Figure 8. The query using one or more words is input and retrieved. The retrieval system is implemented using PHP4 (Hypertext Preprocessor). This system returns results of retrieval and creates meta files for streaming media.



Figure 8: The screen of video retrieval (In Japanese)

CONSIDERATIONS

Sharing metadata for multimedia contents

We described the significance of sharing metadata for content production. Metadata in broadcasting stations are of high reliability and availability, and sharing metadata also enables cost reduction. Here, the sharing of metadata needs to be consistent with other systems, which have diverse metadata. This handling of metadata mainly requires timing of data acquisition and data retention through real operation in broadcasting stations..

However, the standardization of metadata and the specification of metadata do not resolve all problems of cost and the operation of content production for practical use. We need to implement a system for processing metadata using empirical methods, for content production. Therefore, an accumulation of knowledge through demonstration of the handling of metadata is very important.

An automatic processing for sharing metadata and controlling cost was realized, and advanced content synchronized with broadcasts was provided. A further direction of this study will not only be mere data sharing, but also semantic integration of metadata for the effective utilization of metadata. Semantic integration of metadata will allow more intelligent contents.

Video retrieval using metadata by annotation

It is important for the promotion of the use of metadata that the availability of content, i.e., the benefit of metadata, is indicated. This availability is necessary because metadata is valuable only after use in applications and content.

Basically, verification including all processes from metadata creation to its application is needed. From this point of view, an annotation that considers the property of an annotation system and various viewpoints of metadata is performed.

Our video retrieval content using this metadata was highly convenient for users. The correctness rate of metadata reached 95% by appropriate training for metadata creation and the consideration of content. The time from an annotation to the opening of the web site of video retrieval was about ten minutes. This time indicates that all processes were appropriately completed. Our trial is appropriate as a method of real-time metadata creation.

On the other hand, to evaluate the utility of our contents, web traffic of the retrieval system for the batters' scenes and the inning movie is shown in figures 9,10. Figure 9 is daily traffic. Figure 10 is hourly traffic. The time in which the semi-finals games and final game were performed is as follows;

[Semi-final 1st game] 2004-8-21 10:00-12:20
[Semi-final 2nd game] 2004-8-21 13:00-15:30
[Final game] 2004-8-22 13:00-16:00

These figures show our content was accessed by many users. This access tends to increase greatly after opening content (after the game). Such on-demand content was expected to be provided as soon as possible after a game. This content is a practical example of an application using annotation data created in real-time.

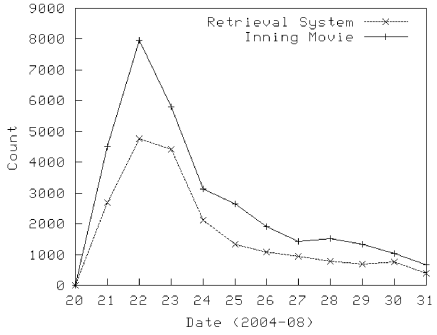


Figure 9: The daily traffic of contents using metadata

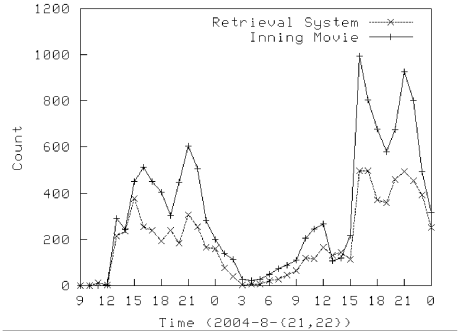


Figure 10: The hourly traffic of contents using metadata

Moreover, this content, which is a low-resolution video (such as Windows Media Format), and the retrieval system using metadata, are the proxy video system. A proxy video is used instead of the original high-resolution video for scene retrieval and content confirmation of editing support via an IP network. In other words, metadata annotated to

multimedia content allows creation of advanced content and operational support in broadcasting stations.

CONCLUSION

In this paper, a concrete case of metadata creation and application in broadcasting stations was given. Each content tends to be individually created. In content production, coordination of content creation via metadata enables construction of good circulation of metadata. Various results of the study and operation of metadata needed to be established for the utilization of metadata. A clear understanding of the problem and empirical knowledge are necessary to create one piece of content from a lot of metadata.

In this sense, our trial of content production through the creation of metadata by sharing existing data and real-time annotation has much significance. However, problems of handling metadata still need to be discussed. A further direction of this study will be to integrate metadata of various domains to create multimedia content.

REFERENCES

Society of Motion Picture and Television Engineering. "SMTPE 330M: Unique Material Identifier(UMID)". White Plains, N.Y.

W3C. "Resource Description Framework (RDF) Model and Syntax Specification". http://www.w3.org/TR/REC-rdf-syntax/

Tomohisa Akafuji, Kaname Harumoto, Keishi Kandori, Koiti Hasida, Shinji Shimojo. 2004. "An Annotation Method and Application for Video Contents Based on a Semantic Graph". Pacific-Rim Conference on Multimedia (PCM2004), part.1 ,pp. 478-486(Dec)

Masaaki Kobayashi. 2004. "Voice Tagging". Technical Report of IEICE, pp.35-39 (Jan.)(In Japanese)

BIOGRAPHY

TOMOHISA AKAFUJI is a Ph.D. student in the graduate school of Information Science and Technology, Osaka University. He received the M.E. degree from University of Tsukuba in 1996. In 1996 he joined Asahi Broadcasting Corporation, Osaka, Japan and has been engaged in the research and development of multimedia systems using metadata for broadcast.

KANAME HARUMOTO received the M.E. and Ph.D. (Eng.) degrees from Osaka University, Osaka, Japan, in 1994 and 1998, respectively. From 1994 through 1999, he was with the Department of Information Systems Engineering, Graduate School of Engineering, Osaka University. From 1999 through2004, he was an Assistant Professor in Cybermedia Center, Osaka University. Since April 2004, he has been an Associate Professor in Graduate School of Engineering, Osaka University. His research interests include data management in advanced network environments. He is a member of IEEE.

MASAAKI KOBAYASHI is an Executive Engineer of Matsushita Electric Industrial Co., Ltd., Osaka Japan. He received the M.S. degree from Shizuoka Univ. in 1972 and Ph. D. degree from Osaka Univ. in 1993. In 1972 he joined Matsushita Electric Industrial, where he has been engaged in development of VTRs and professional systems. He received BCCE Transaction Award from IEEE Consumer Electronics Society in 1980. He is a Fellow of the Institute of Image Information and Television Engineers.

MITSURU YASUKATA is a Senior Coordinator in Matsushita Electric Industrial Co., Ltd., Osaka, Japan. He went to Kumamoto University, where he studied electronic engineering. In 1981 he joined Matsushita Electric Industrial, where he has been engaged in product planning and research planning of audio and video equipment. He received the Best Paper Award of Kansai of the Institute of Image Information and Television Engineers in 2003.

YOZO YAMAMOTO is a Senior Coordinator in Matsushita Electric Industrial Co., Ltd., Osaka, Japan. He graduated from Department of Commercial Science at Doshisha University in 1989. He joined Bandai Co., Ltd. in 1989, and left in 1991. He ended the Business Administration Master's course at Ashurand University in 1995. In 1995 he joined Matsushita Electric Industrial. After joining the company, he has been engaged in game software development and video content creation.

KEISHI KANDORI is a Vice-chief of bureau, Information Technology Center, Asahi Broadcasting Corporation, Osaka, Japan. He went to Waseda University, where he studied acoustic engineering. In 1972 he joined Asahi Broadcasting Corporation, where he has been engaged in research of multimedia and network systems for broadcast. Between 1985 and 1986, he was a research affiliate of the Film & Video Group, MIT Media Laboratory in Cambridge, Massachusetts.

SHINJI SHIMOJO received the M.E. and D.E. degrees from Osaka University in 1983 and 1986, respectively. From 1986 to 1989 he was an Assistant Professor in the Department of Information and Computer Sciences, Faculty of Engineering Science, Osaka University. Since 1998 he has been a Professor in Cybermedia Center, Osaka University. His current interests cover wide diversity of multimedia applications such as News On Demand System, multimedia databases and networked virtual reality. He is a member of ACM and IEEE.

ACHNOWLEDGMENT

This work was supported in part by Research and Development Program of Ubiquitous Network Authentication and Agent (2003), The Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT), Japan.

An Icon-Based Communication Tool on a PDA*

S. Fitrianie and L.J.M. Rothkrantz
Man-Machine-Interaction Group, Delft University of Technology
Mekelweg 4
2628 CD Delft, the Netherlands
E-mail: {s.fitrianie, l.j.m.rothkrantz}@ewi.tudelft.nl

KEYWORDS
Iconic interface, natural language processing, PDA application.

ABSTRACT

To encounter limited options of user interactions provided by PDAs, we propose a new interaction paradigm using icons to represent concepts or ideas. As icons offer a potential across language barriers, we developed two experimental icon-based communication interfaces that are language independent, as proofs of our concept. They were applied on a tool for travellers. Users can create an iconic message as realization of their concepts or ideas in mind. The tool is able to interpret and convert the message to (natural language) text and speech. Corpus-based approach was used to have more insight of how humans express their concepts or ideas using this type of messages. This tool also provides an icon prediction to have faster interaction in next icon selections. Our user test results confirmed that using provided icons our target users could express their concepts and ideas solely using a spatial arrangement of icons

INTRODUCTION

At this moment, many people use a Personal Digital Assistant (PDA) for their personal information management. The advantage of a PDA is its portability and facility for wireless connection. However, the user interaction options for a PDA are quite limited. Like a traditional workstation, interaction with a PDA can be realized with pointing device on a display. Most PDAs also contain a few physical buttons and a few of them have a small-sized keyboard. Recent research has been done on adding multimodal capabilities to a PDA, e.g. (Comerford et.al 2001)(Dusan et.al 2003) (Wahlster et. al 2001). However, for optimal speech recognition, the environment in which the technology is used should be the same as the training environment of the system (Bousquet-Vernhettes 2003), whereas PDAs are often used in various environments under various conditions. The result is misrecognition of commands, which is frustrating to the user. Since the current technology makes speech input less suitable for mobile activities, we aimed at a natural interaction style based on the strength of the Graphical User Interface.

* The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024

Humans communicate to share facts, feelings, and ideas among each other. This communication involves the use of concepts. In interacting with their environment, humans form concepts as internal mental models to represent themselves, the outside world and of things with which they are interacting (Perlovsky 1999). An icon is understood as a representation of a concept, i.e. an object, an action, or a relation. An icon functions as communication means by virtue of a resemblance between the icon and the object or the movement it stands for. By this way, icons also offer a direct method for conversion to other modalities.

An icon can be interpreted by its perceivable form (syntax), by the relation between its form and what it means (semantics), and by its used (pragmatics) (Chandler 2001). By this way, icons also form a language, where each sentence is formed by a spatial arrangement of icons (Chang et. al 1994). Each icon constructs meanings. Since icons are representations of models or concepts, with which humans are actually interacting, we expect this language is easy to learn. Once a set of iconic representations is established, increased usages can lead to more stylized and ultimately abstract representation, as has occurred in the evolution of writing systems, e.g. the Chinese fonts (Corballis 2000).

Some other benefits that are offered by icons as representations of concepts or ideas are: (1) images can be recognized quickly and committed to memory with surprising persistence (Frutiger 1989); (2) an icon can replace a word(s) or phrase(s). The meaning of individual icons represents a word or a phrase created according to the metaphors appropriate for the context of this type of languages; (3) icons offer a potential across language barriers (Perlovsky 1999), which creates a human natural language independent like Esperanto; and (4) icons can evoke a readiness to responds for a fast exchange of information and a fast action as a result (Littlejohn 1996). A direct manipulation on the icons on a GUI allows us to have a fast interaction. A research showed that direct manipulation with a pointer is better time performance than form filling with Soft Input Panels or handwriting recognition (Kjeldskov 2002).

RELATED WORK

Icons have been used as early as the middle ages complex iconic system have been used, e.g. to denote systems of astrological signs. It may even be argued that ancient Egyptians were using icons as a language because they did

communicate using graphics. In modern time, icons still appear in human life as communication means from icons that are used to operate devices, icons that appear in public places to communicate about rules and directions until icons that are created for handicap people. These icons intended to provide the same information are often substantially different visually speaking across languages.

Nowadays, icons form important part in almost every GUI-based computer software as a small graphical representation of a program, resource, state, option or window. With the development in computing, people would be able to communicate with an iconic language without the need to draw pictures themselves, since they could choose these pictures from the screen. Recent attempts in computer-based iconic communication, e.g. the Hotel Booking System that allows communication on a restricted domain (Mealing & Yazdabi 1996), CD-Icon that is designed as pure person-to-person communication system (Beardon 1992), and the Elephant's memory, a computer iconic environment that allows the user to build a visual message by combining symbols from its vocabulary (Housz 1994). However, most of these iconic communication systems are too complex to learn. The languages are either based on too complex linguistics theories or the icons are not intuitive and self-explanatory.

A deep research has been done by (Leemans 2001) on the use of iconic language as a universal language. They developed Visual Inter Language (VIL), an iconic visual interlingua based on the notion of simplified speech. This allows its complexity to be significant reduced; for example, it has no inflection, no number, gender, or tense markers, no article and no linear order. This is possible because it was designed as a visual language in contrast to written languages, where sequencing and ordering is critical. By this way, VIL emphasized simplicity and ease of use.

This paper describes our proposal of a user interaction paradigm offered by two iconic interfaces on PDAs: (1) an interface that plays as a language phrase book for travellers; and (2) an interface for describing a direction on a map. They were applied on a language tool for travellers. The use of icons to represent concepts or ideas makes the interaction particularly suitable for language-independent contexts. The use of PDAs created new constraints and requirements of the user interaction and the interface. Therefore, on one hand, many principles were adopted from VIL such as the language simplicity, self-explanatory icons and ease to use interfaces. On the other hand, both interfaces were designed for a mobile use context. Using the developed language tool, we performed a laboratory user test. It assessed the design concepts and also addressed usability issues of the iconic interface.

COMMUNICATIONS WITH ICON STRINGS

Our first iconic interface was designed as a digital phrase book on a PDA. A user can select a sequence of icons to create an iconic sentence as a realization of his/her concepts or ideas in mind (see examples in figure 1). The iconic interface is able to interpret the sequence and convert it into a human natural language text and speech.

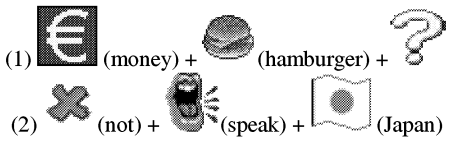


Figure 1. Two iconic sentences: (1) "How much money is the hamburger?" and (2) "I don't speak Japanese"

An iconic sentence with more than just one icon can still be represented by the same icons, but it turn out to be difficult to determine the meaning. The only thing that can be automatically derived from the semantics of the icons in an iconic sentence is a fixed word or phrase belonging to these icons (e.g. "not", "speak", "Japan"). The problem is that individual icons provide only a portion of the semantics of the sentence. The meaning of the iconic sentence is derived as a result of the combination of these icons, and cannot be detected without a global semantic analysis of the sentence. It is unlikely that constructing an iconic message with a one to one matching of icons to words would be appropriate. Humans, on the other hand, have innate ability to see the structure underlying a string of symbols. We are equipped with a number of innate rules (principles) that determine the possible shapes of human languages.

A workshop, involving eight participants and one as a pilot, has been conducted to acquire knowledge on formulating iconic messages from sentences (Fitrianie 2004). The participants were asked to select icon strings from groups of icons to represents six different texts. As comparison, large number corpora of iconic messages were analyzed to have more insight of how humans express their concepts or ideas using this type of message. Although to represent a natural language sentence may use different sets of icon strings, both studies had a similar result: all participants tended to use a small number of icons to represent their concept and ideas. These icons represent important keywords of the message.

Based on these studies, we developed grammar rules using Backus Naur Form and English grammars. The first step in defining a grammar is to define terminal symbols: a lexicon or list of allowable vocabulary icons. The icons are grouped into the categories or parts of speech familiar to dictionary users, such as: nouns, pronouns, proper-nouns, verbs, adjectives, adverbs, prepositions and quantifiers. The grouping depends on what an icon represents, for example:

-  represents "not" and is a negation,
-  represents "speak" and is a verb-transitive,
-  represents "Japan" and is a proper-noun.

The next step is to combine the icons into phrases as non-terminal symbols. Examples of these phrases are: Sentence, Noun Phrase, Verb Phrase, Prepositional Phrase, and

Negation Sentence. As examples using the corpora on figure 1, we have grammar rules:

```
Noun-Phrase     noun | noun Noun-Phrase
Verb-Phrase     verb-transitive Noun-Phrase
Sentence     Noun-Phrase + question-sign |
             negation + Verb-Phrase
```

A parser takes a user input and extracts each icon against the grammar rules. If the inputted iconic sentence is syntactically correct, the parser creates seven slots: prefix, subject, infix, verb, object, preposition, and suffix slot. The prefix slot is a container for question words; the infix slot is a container for a to-be or an auxiliary and a negation ("not"); and the suffix slot is a container for a question mark, an exclamation mark, or an archaic word, such as "please". The position of the slots may not in a sequential order. It depends on the type of a sentence, for example: for a question sentence, the infix slot may be located between the prefix and the subject slot. A slot may be empty and may contain more than one word. For example, using the second corpus in figure 1 and the grammar rules above, the seven slots are:

```
Prefix = [empty]
Subject = [empty]
Infix = [negation ("not")]
Verb = [verb ("speak")]
Object = [proper-noun ("Japan")]
Preposition = [empty]
Suffix = [empty]
```

To complete the sentence, some extra rules are necessary to be fired. The rules derive the meanings associated with iconic sentences from the meaning associated with the individual icons forming the sentence. Based on the seven slots above, some examples of rules that are required to be fired are those for adding a subject (e.g. "I", "you"), for an auxiliary verb (i.e. "do", "does") and for changing a proper-noun's format (e.g. "Japan" into "Japanese"). The resulted text for this example is:

```
Subject + Infix + Verb + Object
I do not speak Japanese
```

To develop these rules, we used a corpus based approach.

COMMUNICATIONS WITH ICON MAPS

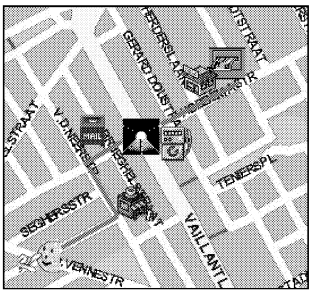


Figure 2 Communicate with icon on a map

Our second idea was designed as a travel guide that describes how to find location on a map. A user can select icons and place them on a map to describe things, places, or situations that the user will view from his/her current position in the direction to the final destination (see figure 2). The iconic interface displays a line drawing to show the closest route from the starting point to the final destination. It also creates a scenario and converts it to a text or speech. The user can save his/her profiles and preferences, which will be adapted by the interface in the next interaction.

The interface lists all inputted icons in a tuple that consists of four elements: icon, position (in global world coordinates), timestamp, and pointer to the next selected icon. The global world coordinates are calculated based on the local coordinates of a map. The input list is composed in XML schemes. The following are some tuples from the example in figure 2:

```
<icon name="coffee shop" x="30" y="385"
        time="000403" next="post office"/>
<icon name="post office" x="25" y="201"
        time="000447" next="tunnel"/>
...
```

The pointer to the next icon is required because a user may place icons not in a sequential order. Besides adding a new icon on a map, a user can also delete an icon from the map or move it to another coordinate. Therefore, a knowledge base was developed to update and reason inputted facts on the map. This knowledge base also has default facts that are defined impossible locations of icons on a map. For example: it is not possible to place an icon "supermarket" on a highway. The interface will only inform the user about it.

The knowledge base contains three sets of rules. First, it has a set of rules for detecting the beginning and ending of a street and detecting an intersection. It uses a database that contains global world coordinates of every street. Second, a set of rules were defined for determining a pointer value for a tuple in the input list. An example of rules for determining the pointer value is:

```
If time[i] < time[i+1] and
   length(x[i],y[i]) < length(x[i+1],y[i+1])
Then next[i] =  name[i+1]
```

where *i* is an icon's index in the list; the value of time, x, y, name and next are attributes of the i^{th} element in the input list. The length is determined by calculated the length of the closest route of the position of an icon (x,y) from the starting point. As default, the starting point is the user's current position received from a global positioning system, except if the user puts an icon that functions to mark her/his position, i.e.  ("I"), as the starting point.

Finally, another set of rules were also developed for developing a scenario. Using the results of the first set of rules in the knowledge base, the third set of rules defines two actions: (1) move (go forward, left or right) until or on <object>; and (2) see <object> (in front, left-side or right-side), for every tuple in the input list. An example of a scenario using the map and the input list above:

```
<scenario action="move">
    <dir>forward</dir>
    <object type="flag">eo street</object>
 </scenario>
 <scenario action="see">
    <dir>front</dir>
    <object type="icon">coffee shop</object>
 </scenario>
...
```

To develop all rules, we analyzed different arrangements to place icons on a map.

A user can save an input list as his/her profile. The user can also add special icons such as, his/her hotel, conference building, or favourite restaurant. This information is saved in XML schemes, for example:

```
<profile name="post office" x="25" y="201"
        map_id="M355"/>
<profile name="my hotel" x="110" y="400"
        map_id="M355"/>
...
```

Besides containing global world coordinates of streets, the developed database also contains information about the location of public places on a map, such as stations, restaurants, museums, hospitals, etc. The interface gives options to its users regarding which groups of icons are preferable to be displayed on the map and the maximum number of icons can be displayed on the map automatically. The next time a map is displayed; the interface will display the user's profiles that were assigned for this map and those icons that fit with the user's preferences. These icons are included into the input list automatically to create a relevant scenario. All elements in the list will be checked if a fact is added or changed.

A natural language processing (NLP) module processes the scenario and converts it into natural language text. It uses default phrases, such as: "go forward until ..", "you'll see a .. in your left-side", "on the third street go right", etc. The following is the resulted text of the scenario above:

```
Go forward until the end of the street.
You'll see a coffee shop in front of you.
...
```

USABLE ICONIC INTERFACE

In term of communication means, both interfaces provide a large vocabulary of icons. Our design concept supports users to create iconic messages as small number of steps as possible. We approached it by three ways: (1) designing good icons; (2) designing a usable interface; and (3) providing ways for fast interactions.

Design "Good" Icons

According to Macintosh Human Interface Guidelines 1992, poorly designed icons can have disadvantages: (1) ambiguity; (2) personal (prior-) knowledge dependency; (3) incapable to replace words completely; and (4) linguistics and/or culturally bias. Therefore, we took some guidelines and standards to guide us during design process i.e. Macintosh Human Interface Guidelines, ISO/IEC 2000, (Horton 1994) (Schneiderman 1992), and (Smith & Mosier 1986). Most of them are from interface design guidelines and books.

As pictorial signs, icons could also be analyzed using semiotics: the study of sign (Chandler 2001). This concerns how signs obtain their meaning and how they convey them. Based on Peirce's triadic view of sign (figure 3), the degree to which the user interaction exists depends on how close the interpretation in the user's mind is to the object of an icon (Theleffsen 2000). Peirce also defined ten sign types; three of them are particularly interesting for our project: (1) iconic signs that occur when the representation relates to the object through a resemblance; (2) indexical signs that occurs when the representation relates to the object through causation; and (3) symbolic signs in which the representation relates to the object purely through convention. Our approach is not to design icons that cause the user to recognize what they represent immediately, but can be used to help the user to remember them. This is one reason why we prefer indexical and iconic signs.

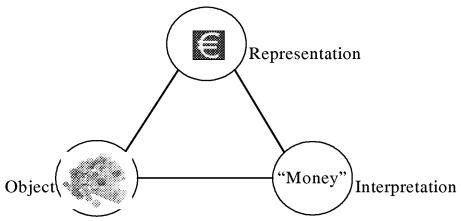


Figure 3. Peirce's triadic view of sign.

The semiotic approach provides us a good structure for a common sense evaluation on icons. Since this interpretation relation is a subjective matter, in particular, the approach shows that we must be aware of all aspects. Therefore, we performed a user test on every new icon design. The test was designed based on (Horton 1994)'s following equation:

$$\text{Icon}_i + \text{Context}_j + \text{Viewer}_k \quad \text{Meaning}_{ijk}$$

We tested each icon in the context of other icons. The test participants were selected to include different nationalities to solve problems of linguistics and culturally bias of the interpretation of an icon. If the interpretations of different viewers to an icon were not the same as the intended meaning of its designer, it means that the icon was required to be redesigned.

Design Usable Iconic Interface on a PDA

Squeezing complex functionalities into the interface of a pocket size computer and putting it on the hand of a mobile user requires additional knowledge about human-computer interaction in a mobile use context to be obtained. Mobile devices are different from the conventional systems, on one hand, due to its constrained screen size, limited options of input devices, and restricted power and performance. On the other hand, these devices are designed to support people operating it anywhere and everywhere. Therefore, in designing an iconic interface for a PDA, we also took some guidelines and standards, i.e. Pocket PC UI Guidelines 2004, (Kärkkäinen & Laarni 2002), (Ostrem 2003), and (Vanderdonckt et. al 2001). They describe requirements and recommendations to cope with PDA's hardware constraints and mobile user characteristics.

Using our iconic interface, users will interact with a large number of icons. To help users to find their way around icons, we approached this by grouping related icons with the same set of concepts. Besides supporting a compact

interface, this grouping is powerful way to hint where an icon can be found because the meaning of icons almost exists in the context of other icons (Horton 1994).

Since it deals with a large database of icons, the interfaces also support faster interaction by three approaches. First, both interfaces provide a next icon prediction tool, which predicts and ranks next possible icons (see next subsection). Second, the interfaces also provide a search engine based on a keyword what an icon represents. Finally, the interface for creating icon strings in particular provides a real time distinctive appearance of which icons can be selected next according to syntactical rules. The interface is designed to prevent the users to make a syntax error sentence. An icon cannot be selected if it results a grammatically incorrect sentence. All icons are always displayed, however, only active icons can be selected. By this way, the interface could convince its users by offering accurate and reliable interpretation of inputted icon strings since mobile users cannot devote their full attention to operate the application.

Considering the PDA's display size, the interfaces were designed to display only most important features. Interface elements that are used more frequently for creating iconic messages are displayed around the input area. Therefore, any action on the elements can be percept by the user almost directly. This could increase users' attention for the coinciding action-perception space. At the same time, distinctive visual cues for a selected element from the rest unselected elements are provided.

The results of any user interaction are provided as direct as possible as the users intended. Thereby, although some icons are still unknown, the users can learn them on trials. Both interfaces provide ways to its users to return easily to the previous interaction state (e.g. using a delete or clear action), to the previous selected of icons, and to other groups of icons.

N-Gram Next Icon Prediction

Due to the high number of presented icons, scrolling on the interface is inevitable. Extra steps to find intended icons are needed. To generate a smoother dialog especially for a language tool, an icon prediction system could help the user to eliminate these steps. Besides for improving the input speed, in creating iconic sentences particularly, the prediction also improves the quality of syntax. It predicts which icons are most likely to follow a given segment of an icon string that is grammatically correct. As a result, the number of icon look-up processes that necessary to generate the string will be reduced.

We adapted a word prediction technique, which was frequently used in writing support device. The prediction operates by generating a list of suggestions for possible icons after the first icon is selected. A user either chooses one of the suggestions or continues entering until the intended arrangements of icons appear. A selected suggestion is automatically inserted into the input area.

We used statistical approach that was derived from a probabilistic language model. The probability of an icon string is estimated with the use of Bayes rule as the product of conditional probabilities:

$$P(s) = P(w_1, w_2, ..., w_n) = \prod_{i=1}^{n} P(w_i \mid w_1, ..., w_{i-1}) = \prod_{i=1}^{n} P(w_i \mid h_i)$$

where h_i is the relevant history when predicting w_i. To predict the most likely icon in a given context, a global estimation of the icon string probability is derived which is computed by estimating the probability of each icon given its local context (history). Here, we used estimating conditional probabilities of n-grams type features. Our icon prediction use trigram language model to propose the most likelihood icon by knowing the previous selected one, using the equation:

$$P(w_{n+1} \mid w_n w_{n-1}) = \frac{C(w_{n+1})}{N} + \frac{C(w_{n+1} w_n)}{C(w_{n+1})} + \frac{C(w_{n+1} w_n w_{n-1})}{C(w_{n+1} w_n)}$$

where C = the frequency of the user selection, w_n = the n^{th} position of inputted icons, N = the total number of the selection, and $C(w_{n+1}\ w_n)$ and $C(w_{n+1}\ w_n\ w_{n-1})$ = the frequency of all pairs. This approach needs a large amount data to compute multi-grams model. Our developed iconic interfaces collect the data from user selections during the interaction.

A TOOL APPLICATION FOR TRAVELLERS

During travelling to other countries, travellers may need to communicate with local citizens or natives. As language is fundamental to human activities, proficiency in other languages becomes important. With the introduction of computerized mobile devices (i.e. PDAs), new opportunities for communication in other language arose. Our iconic interfaces also offer such opportunity.

Our users can use the language tool to act as a communication bridge with others. The icon map interface in particular, the interface can be used to describe a direction as if we explain it by sketching on a piece of paper when a verbal explanation is considered insufficient or impossible. Besides supporting a fast interaction, the language tool also can be combined with any language translator. It provides a speech synthesizer to read aloud the resulted translations with their correct pronunciations.

Our language tool was dedicated for travellers. The purpose of the travels may for either business and/or pleasure. A questionnaire involving forty travellers from different nationalities was conducted (Fitrianie 2004). The questionnaire addressed a deeper insight into the experiences during travelling. We also intended to determine user requirements for our language tool. The results from this study indicated that people have already used iconic representation to express concepts or ideas in sign language for survival purposes, especially when verbal communication is not feasible.

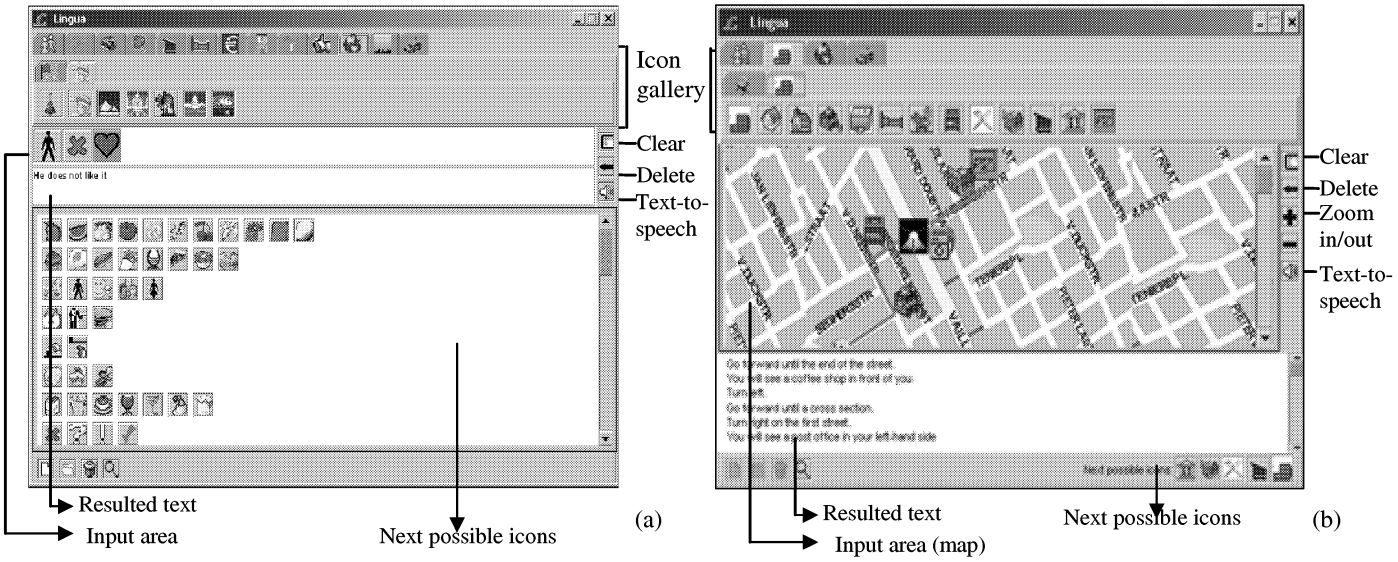


Figure 4. Our experimental iconic interfaces: (a) a digital phrase book and (b) an icon

From the results, we also gathered data about the topics and the situations in which the users need to use the language tool. From the questionnaire, we found that tasks during travelling are relatively the same for all travellers. According to more than 60% of the participants, they needed to speak a local language mostly for greetings to others, asking for directions, in a station (buying ticket, asking time), in a market (buying and paying goods), and in a restaurant (ordering foods and drinks). As comparison, we also surveyed pocket-phrase books for travellers (Kosmos Taalgids 1997), We found that these books contained relatively the same topics. The topics are typically organized into sections, such as: greetings, self-introduction, at a hotel, etc.

Based on more than 50% of the participants, some important requirements for a language aid were: (1) to provide a search engine based on a keyword; (2) to display resulted texts both in the user's mother tongue and in the local language; (3) to read with correct pronunciations; and (4) to provide a list of pictures for various situations, topics, places and objects, to point at. Some participants mentioned that it was not needed to speak the local language with a perfect grammar. The natives would appreciate them and more helpful, if the travellers had already tried to speak their language. This encouraged the designers for developing the language tool incrementally, since it was difficult to complete the iconic database at once. However, correct pronunciations are essential to have meaningful communications.

Figure 4 shows two interfaces of the current version of our language tool. They were developed to cope with the screen size of the Sharp Zaurus Personal Mobile Tool (640x480 pixels). Both interfaces were developed in Java SDK 1.3 and their knowledge base was developed using Jess (Java Expert System Shell).

The phase book interface has five hundreds icons in its vocabulary. The icon map interface has a hundred icons. They are divided into topics. In the icon gallery, icons are grouped based on their concept. These icons are displayed in hierarchy that comprise three levels. The top level presents the concept icons and will incorporate compound icons. Querying any concept will initiate a move to the second level, at which can contain sub-concept icons or base icons. A further level contains only base icons, which users can choose to select.

The input area (or the map) is a container of selected base icons. Each time an icon is selected, the NLP will process the conversion of the input into text. The user can delete or clear his/her selection using some buttons right next to the input area. A deletion action will remove only one selected icon. As default, the last insertion will be deleted. A clear action will remove all inputted icons.

On the icon map interface, we added two menus for zooming in and zooming out to prevent placing icons on the wrong location. The zooming actions affect the map's size, not the icons' size.

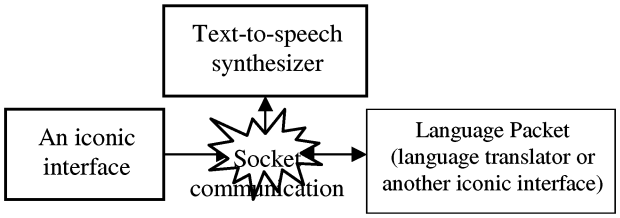


Figure 5. Communication an iconic interface with other applications or other interfaces

The next prediction tool works on both interfaces and gives suggestions for the next possible selected icons. Particularly in the digital phrase book, the suggestions are ranked, listed, and grouped based on their concept. Any changing on the input area or the map, the resulted text will be then refreshed.

The current version of the language tool application does not have any language translator. However, through a socket network, both interfaces are able to create a communication environment by TCP/IP connections (see figure 5). The socket allows the interface to communicate to other language tools (e.g. a text to speech synthesizer, a language translator, or a language dictionary) and to other applications (e.g. a web application or other communication interfaces). These language tools and applications can be produced by third-parties.

As an experiment purposes, we used Carnegie Mellon University's Flite (Black & Lenzo 2003) for synthesizing the resulted text to speech. The text and speech are in English. Via the speech synthesizer, the tool provides handy common utterances with their correct pronunciations. The user can click the speaker icon to activate this action.

EXPERIMENTAL EVALUATION

A user test assessed whether or not users were capable to express their concepts in mind solely using a spatial arrangement of icons. This test also addressed usability issues on interacting with both interfaces. Eight people took part in the test and one person as a pilot. They were selected in the range age of 25-50 years old (based on AvantGo survey 2003: 69% responses are a PDA user and in the range age of 25-50 years old) and to include different nationalities.

We run the experiment using Thinking Aloud method (Boren & Ramey 2000). Each participant had five tasks. The tasks were created real situation for travellers using cartoon-like stories which the participants should have used the system to communicate with others. While performing the tasks, the participants were asked to think aloud. There were not incorrect answers, except if the participant did not answer the question at all. All activities were recorded on a tape recorder and all user interactions on the tool were logged, to be analyzed afterward.

Smith's measurement of the sense of being lost in hypermedia (Smith 1996) was adapted to measure disorientation of being lost in an iconic interface. As the term "hypermedia" is used to denote a hypertext document that is made up of interlinked pages or nodes, our icon space also was made up interlinked icons. Since all information was displayed only by icons, we thought that an iconic interface also might give cognitive overload and disorientation to its users. The problem referred to "users cannot find what they are looking for". For a perfect search, the lostness rating should have been equal to 0.

From the results, it appeared that our target users were able to express their concepts and ideas in mind using icons. All participants could compose iconic messages for the given situations.

The experimental results also indicated that our target users still had problems in finding their desired icon from the interface. This was indicated by the high number of the lostness rating of five sequence tasks (see figure 6). Some of the reasons were referring to some problems in recognizing some icons, such as: their size and colour contrast between elements in an icon and between icons and their background. Other reasons were: (1) the participants needed an adaptation time with the tool; (2) The time was also needed to find more relevant concept to represent their message; and (3) they should have rethought another concept that could fit with the problem domain due to limited vocabulary.

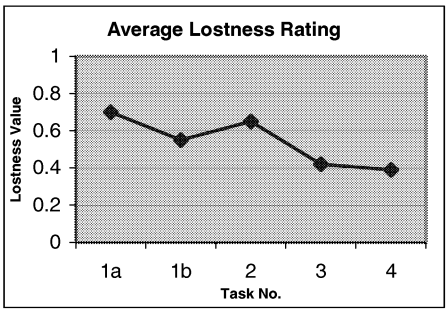


Figure 6. Average lostness rating

Apart from these problems, we viewed decreasing lostness rating in terms of improvement of user performances. Our test users had taken benefits provided by both interfaces, e.g. the next icon prediction tool and the searching engine, in creating iconic messages. We concluded that our test users only needed a small period of time to adapt to both interfaces.

CONCLUSION

Two experimental icon-based communication interfaces have been developed, applied on a language tool for travellers. The first prototype is an iconic interface for creating iconic sentences and plays as a language phrase book. The second prototype is an iconic interface for describing a direction on a map.

NLP has provided a method to interpret and convert iconic messages. We solved the problems of ambiguity and missing information that resulted by this type of messages using a usable icon design approach and rule-based approach. We also developed rules to create more natural sentence. N-grams language model was used to help the users in creating iconic sentence fast. Our experimental results showed that icon interfaces could serve as communication mediator. However, we also found that users generated some arrangement of icons, which were out of domain. Therefore, future work is necessary to analyze more corpora of dialogs from various situations and topics.

As language independent interfaces, both interfaces are designed with a socket network to communicate with other application or other interfaces. Thereby, the interface opens opportunities for the development of other applications in different languages and domains. One might think of those where verbal communication is not possible, for example a communicator media to help people with speech disability. However, the interface requires more research for the specific target users.

Furthermore, a field test is necessary to gather data about how people might use an iconic interface in their normal situation during travelling and how they experience this. Therefore, the interface design can cover more user requirements in mobile context use.

REFERENCES

Apple Computer (1992), Inc. Staff, Macintosh Human Interface Guidelines, Addison-Wesley Publishing Company.

AvantGo Inc. (2003), Research Summary: 2003 Mobile Lifestyle Survey, http://www.avantGo.com.

Beardon C. (1992), CD-Icon: an Iconic Language-Based on Conceptual Dependency, Intelligent Tutoring Media, 3(4).

Black A.W. & Lenzo A.K. (2003), Flite: a Small, Fast Speech Synthesis Engine, System Documentation Edition 1.2, for Flite version 1.2, USA.

Boren T. & Ramey J. (2000), Thinking Aloud: Reconciling Theory and Practice, IEEE Transactions on Professional Communication, 43-3.

Bousquet-Vernhettes C., Privat R., & Vigouroux N. (2003), Error Handling in Spoken Dialogue Systems: Toward Corrective Dialogue, *Proc. Of ISCA'03*, USA.

Chandler D. (2001), Semiotics: the Basic, Routledge.

Chang S.K., Polese G., Orefice S., & Tucci M (1994),. A Methodology and Interactive Environment for Iconic Language Design, International Journal of Human Computer Studies, 41: 683-716.

Comerford L., Frank D., Gopalakrishnan P., Gopnanth R., & Sedivy J. (2001), The IBM Personal Speech Assistant, *Proc. of the ICASSP 2001,* USA.

Corballis M.C. (2000), Did language Evolve from Manual Gestures?, *3rd Conference of The Evolution of Language'00,* France, 2000.

Dusan S., Gadbois G.J., & Flanagan J. (2003), Multimodal Interaction on PDA's integrating Speech and Pen Inputs, *Proc of. EUROSPEECH'03*, Switzerland.

Fitrianie S. (2004), An Icon-Based Communication Tool on a PDA, Postgraduate Thesis, TU Eindhoven, the Netherlands.

Frutiger A. (1989), Sign & Symbols: Their Design and Meaning, New York: van Nostrand Reinholt.

Horton W. (1994), The Icon Book, New York, John Wiley.

Housz T.I. (1994-1996), The Elepahant's Memory.

ISO/IEC (2000), International Standard, Information Technology – User System Interfaces and Symbols – Icon Symbols and Functions, 1st Edition, ISO/IEC 11581-1:2000(E) to ISO/IEC 11581-6:2000(E).

Kärkkäinen L. & Laarni J. (2002), Designing for Small Display Screens, *Proc. of 2nd Nordic HCI'02*, ACM Series.

Kjeldskov J. & Kolbe N. (2002), Interaction Design for Handheld Computers, *Proc. of the 5th APCHI'02*, Science Press, China.

Kosmos Taalgids (1997), Wat & Hoe: Samengesteld Door VAN DALE Lexicografie.

Leemans N.E.M.P. (2001), VIL: A Visual Inter Lingua, Doctoral Dissertation, Worcester Polytechnic Institute, USA.

Littlejohn S.W. (1996), Theories of human communication, 5th Edition, Wadsworth.

Mealing S. & Yazdabi M. (1992), Communicating Through Pictures, Department of Computer Science, University of Exeter, England.

Microsoft (2004), MSDN: UI Guidelines (Pocket PC User Interface Guidelines), http://msdn.microsoft.com.

Ostrem J. (2003), PalmSource: Palm OS User Interface Guidelines, http://www.palmos.com.

Perlovsky L.I. (1999), Emotions, Learning and Control, Proc. of International Symposium: Intelligent Control, Intelligent Systems and Semiotics, 131-137.

Schneiderman B. (1992), Designing the User Interface: Strategies for Effective Human-Computer Interaction, 2nd Edition, Addison-Wesley Publishing, 1992.

Smith P. (1996), Toward a Practical Measure of Hypertext Usability, Interacting with Computers, 8-4:365-381, Elsevier Science Ltd B.V.

Smith S.L. & Mosier J.N. (1986), Guidelines for Designing User-Interface Software, ESD-TR-86-278, Bedford, MA 01730: The MITRE Corporation, 1986.

Theleffsen T. (2000), Firstness and Thirdness Displacement: Epistemology of Pierce's Sign Trichotomies, AS/SA 10(2), Applied Semiotics, ISSN 1204-6140, 2000.

Vanderdonckt J., Florins M., Oger F., & Macq B. (2001), Model-Based Design of Mobile User Interfaces, Proc. of 3rd Mobile HCI'01, M.D. Dunlop and S.A. Brewster (eds.), Lille, 135-140.

Wahlster W., Rethinger N., & Blocher A. (2001), SmartKom: Multimodal Communication with a Life-Like Character, *Proc. Of EUROSPEECH'01*, Denmark.

AUTHOR BIOGRAPHY

SISKA FITRIANIE was born in Bandung, Indonesia and went to Delft University of Technology, the Netherlands, where she studied Technical Informatics and obtained her master degree in 2002. After doing her two years post-graduate programme at Eindhoven University of Technology, she involves in the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, since 2004 as her PhD project at Delft University of Technology. Her project aims at designing and developing models of a computer-human interaction system.
Email: s.fitrianie@ewi.tudelft.nl
Webaddress: http://mmi.tudelft.nl/~siska

LEON J.M ROTHKRANTZ received the M.Sc. degree in mathematics from the University of Utrecht, Utrecht, The Netherlands, in 1971, the Ph.D. degree in mathematics from the University of Amsterdam, Amsterdam, The Netherlands, in 1980, and the M.Sc. degree in psychology from the University of Leiden, Leiden, The Netherlands, in 1990. He is currently an Associate Professor with the Data and Knowledge Systems Group, Mediamatics Department, Delft University of Technology, Delft, The Netherlands, in 1992. His current research focuseson a wide range of the related issues, including lip reading, speech recognition and synthesis, facial expression analysis and synthesis, multimodal information fusion, natural dialogue management, and human affective feedback recognition. The long-range goal of his research is the design and development of natural, context-aware, multimodal man–machine interfaces. Drs. Dr. Rothkrantz is one of the Program Committee for EUROSIS.
Email: l.j.m.rothkrantz@ewi.tudelft.nl
Webaddress: http://mmi.tudelft.nl/~leon

VIDEO BASED SYSTEMS

COMPLEXITY ANALYSIS OF INTERPOLATION FILTERS FOR SCALABLE VIDEO CODING

Koen De Wolf, Yves Dhondt, Jan De Cock, and Rik Van de Walle
Ghent University – IBBT, Department of Electronics and Information Systems – Multimedia Lab,
Sint-Pietersnieuwstraat 41,
B-9000 Ghent, Belgium
+ (32) 9 264 89 29
{koen.dewolf, yves.dhondt, jan.decock and rik.vandewalle}@ugent.be

KEYWORDS

Scalable video coding, multi-layer motion prediction, interpolation filters

ABSTRACT

Three low complexity algorithms that allow spatial scalability are presented in this paper. We discussed the feasibility of reusing motion and residual texture information of the base layer in the enhancement layer. The prediction errors and the complexity of the discussed filters and schemes are evaluated. The presented 6-tap and bicubic filters perform better than the bilinear and nearest neighbor filters for the prediction of the high resolution pictures. In contrast, the bilinear and nearest neighbor filters perform better for the interpolation of residual texture information.

INTRODUCTION

The Moving Picture Experts Group (MPEG, ISO/IEC JTC 1/SC 29/WG 11) recently started exploring the field of Scalable Video Coding (SVC). The proposed technologies are mainly focusing on temporal, spatial and quality scalability. At the moment, several indications exist that the high complexity of the algorithms might be a problem for (real-time) content creation (Wul 2002; Landge 2004).

Motion estimation is typically the most complex and time-consuming part of the encoder. If we presume that we have three levels of spatial scalability, we need to estimate the motion on these three levels for every motion compensated picture in order to achieve maximum coding efficiency. In scenarios where live content is streamed in a scalable manner it may not be possible to determine the ideal motion vectors for all spatial resolution levels due to the complexity of SVC. In this paper, we discuss the possibility of reusing motion and residual texture information gathered at the lowest spatial resolution (generally called base layer) for the prediction of higher resolution versions (enhancement layers). This evaluation will be done using three different prediction schemes.

In these schemes interpolation filters will be used for the prediction of high resolution pictures. Therefore, an important issue will be the choice of interpolation filters that must be used. More in particular, the question is whether an increase of the number of filter taps - and therefore also an increase of the complexity of the filter - lead to a better prediction? In this paper, we will compare four interpolation filters with a different number of filter taps based on one decimation filter. The latter filter is used for the generation of the lower resolution pictures in the base layer.

PREDICTION SCHEMES

In this paper three prediction schemes are used for the comparison of the interpolation filters. In scheme 1, we predict all high resolution pictures in the enhancement layer by interpolation of the corresponding decoded pictures of the base layer. The obtained interpolated pictures are then used for the prediction of the original high resolution picture. In scheme 2 the high resolution pictures are predicted using motion information of the base layer. These motion vectors, together with the macroblock partitioning, are used for the motion compensation on the high resolution pictures. Finally, in scheme 3 motion information is also reused but now the residual texture information of the base layer is interpolated and added to the motion compensated pictures of the enhancement layer. This means that in scheme 2, no interpolation is done; in scheme 3, residual pictures are interpolated. For the three described schemes, I-pictures of the enhancement layer are also predicted by interpolating the low resolution equivalent. Other techniques such as intra macroblok prediction (ITU-T 2003; Wiegand 2003) can also be used.

The motion vectors remain the same for both spatial layers; i.e. motion vectors have ¼ pixel accuracy for the base layer and ½ pixel accuracy for the enhancement layer. Motion compensation in the enhancement layer is done on macroblock partitions which are doubled in size compared to the corresponding macroblock partitions of the base layer. Schemes 2 and 3 are using uni-directional prediction.

The difference between the original high resolution pictures and the predicted pictures (using one of the discussed schemes) still needs to be coded. The same concepts, techniques, and partition modes as specified in the H.264/AVC standard (ITU-T 2003; Wiegand 2003) can be used. To be more specific: integer transformations (DCT and Hadamard), scaling, and context adaptive entropy coding. Doing so allows the reuse of already available components at the encoder.

To allow quality scalability, the obtained residual pictures can be coded using a Fine-Granular Scalability (FGS) based entropy coding mechanism. This is not discussed in this paper. The effect on temporal scalability is discussed in (Schwarz 2004a).

FILTERS

We compared four symmetric filters for the interpolation of the reference pictures and the interpolation of the residual texture information: nearest neighbor, bilinear (1, 1)/2, bicubic (-1, 5, 5, -1)/8 and the 6-tap filter defined in the H.264/AVC specification (1, -5, 20, 20, -5, 1)/32. The latter is defined in H.264/AVC for the purpose of half-sample interpolation and used for motion estimation with sub-pixel accuracy.

In order to compare the performance of the described schemes and filters, we encoded several sequences with different motion characteristics (ranging from low to high motion activity). All macroblocks of P-pictures were inter-coded using only one reference picture; no B-pictures were used. The pictures of the low resolution sequence were generated using an 11-tap downsampling filter (1, 0, -5, 0, 20, 32, 20, 0, -5, 0, 1)/64 (Schwarz 2004).

Complexity

The filters described above can all be implemented using bit shifts, additions and subtractions. Figure 1 illustrates a possible 1-dimensional implementation of the given filters. In this figure, we see that for the nearest neighbor filter no computation has to be done. Thus, this filter can be implemented using move instructions. The bilinear filter acts as an averaging filter and can be implemented using 2 additions and a right shift over 1 bit. Adding 1 to the sum is for compensating the possible loss of accuracy introduced by the shift operator.

The implementation of the bicubic filter shown in Figure 1 illustrates that no extra memory locations are required during the application of this filter. However 6 additions/subtractions and 2 bit shifts are needed. When implementing the 6-tap filter, intermediate results can be stored for subsequent use; this is shown by underlining the intermediate values that can be stored. This allows us to minimize the number of calculations. Still 7 additions/subtractions and 3 bit shifts are needed.
Table 1 summarizes the number of computational operations. Please note that these values are only valid for a 1-dimensional implementation.

Table 1: Number of computational operations

Filter	# additions / subtractions	# bit shifts	# memory locations
nearest neighbor	0	0	0
bilinear	2	1	0
bicubic	6	2	0
6-tap	7	3	4

RESULTS

Test set-up

We selected five sequences with different motion characteristics ranging from low motion activity to complex scenes with a lot of motion. I-pictures are inserted every 32 or 300 pictures. An I-picture in the context of H.264/AVC is a picture consisting entirely of intra predicted slices. Furthermore, 2 values for the Quantization Parameter (QP) for the base layer and 5 values for the enhancement layer are compared. The exact numbers are shown in Table 2: Test conditions.

Table 2: Test conditions

Sequence	Bus, Crew, Foreman, Mother & Daughter, Stefan
Intra period	32, 300
QP base layer	16, 32
QP enhancement layer	0, 12, 24, 36, 48

Evaluation

The Mean of Absolute Differences (MAD) between the obtained predicted pictures and the original spatial high resolution pictures is determined. These values are used for the evaluation of the prediction schemes. This gives us the average deviation per pixel.

The most interesting observations are given in the next section.

Interpretation

In Figure 2, we see that for sequences with a homogeneous type of motion there is no significant difference between the bicubic and the 6-tap filter when using scheme 1. In Figure 3 the deviation of the MAD compared with the 6-tap filter is plotted for the Mother & Daughter sequence. The impact of the interpolation filters is only visible in the first pictures following an I-picture. The deviation of the MAD stabilizes for the subsequent pictures. This means that for sequences with a low amount of motion the nearest neighbor algorithm and the complex 6-tap filter produce comparable results for

scheme 2.
The fact that motion compensation is done on the decoded pictures of the enhancement layer results in rather small prediction errors. This is especially true for sequences with low to moderate motion activity for scheme 2. Adding the interpolated residual of the base layer – as described in scheme 3 - gives even lower prediction errors for sequences with motion activity.

In sequences that contain a lot of motion (e.g. crew) however, the nearest neighbor filter and the bilinear interpolation filter used for the interpolation of the residual texture information perform on average respectively 0.44% and 1.07% better than the 6-tap filter (Figure 4).

CONCLUSION

In this paper, we have described three low complexity algorithms that allow spatial scalability. We discussed the feasibility of reusing motion and residual texture information of the base layer.

The bicubic and 6-tap filters outperform the nearest neighbor and the bilinear filters for scheme 1 and the former filters have a similar performance in terms of SAD. The choice for the interpolation filter for the interpolation of the residual data as available in the base layer is more difficult. We see however that when motion activity is high the nearest neighbor and bilinear filter perform slightly better.

The latter indicates that the choice of interpolation filter and prediction scheme is highly content dependent.

ACKNOWLEDGEMENTS

The research activities that have been described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC), and the European Union.

REFERENCES

Horowitz, M., A. Joch, F. Kossentini, and A. Hallapuro. "H.264/AVC Baseline Profile Decoder Complexity Analysis". IEEE Trans. CSVT, vol.13, no. 7, pp. 704- 716, July 2003.

ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC. 2003. "Advanced Video Coding for Generic Audiovisual Services", (version 1: 2003, versions 2: 2004) version 3: 2005.

Landge, G., M. van der Schaar, and V. Akella. 2004. "Complexity analysis of scalable motion-compensated wavelet video decoders". Proc. SPIE Int. Soc. Opt. Eng. 5558, 444 (2004)

Schwarz, H., D. Marpe, and T. Wiegand. 2004. "Scalable extension of H.264/AVC". ISO/IEC JTC 1/SC 29/WG11 MPEG2004/M10569/S03, Munchen, Germany, March 2004

Schwarz, H., D. Marpe, and T. Wiegand. 2004a. "MCTF and Scalability Extension of H.264/AVC". Proc. PCS'04, San Francisco, CA, USA, December 15.-17. 2004.

Wiegand, T., G. J. Sullivan, G. Bjøntegaard, and A. Luthra. 2003. "Overview of the H.264/AVC video coding standard", IEEE Trans. CSVT, vol. 13, no. 7, pp. 560-576, July 2003.

Wu1 F., S. Li1, R. Yan, X. Sun, and Y.-Q. Zhang. 2002. "Efficient and universal scalable video coding". IEEE International Conference on Image Processing (ICIP), vol. 2, pp 37-40, Rochester, September 2002

$$(1)\ \begin{cases} x_{2n} = x'_n \\ x_{2n+1} = x'_n \end{cases}$$

$$(2)\ \begin{cases} x_{2n} = x'_n \\ x_{2n+1} = \left(x'_n + x'_{n+1} + 1\right) >> 1 \end{cases}$$

$$(3)\ \begin{cases} x_{2n} = x'_n \\ x_{2n+1} = \left(\left[\left(x'_n + x'_{n+1}\right) << 2\right] + x'_n + x'_{n+1} - x'_{n-1} - x'_{n+2} + 4\right) >> 3 \end{cases}$$

$$(4)\ \begin{cases} x_{2n} = x'_n \\ x_{2n+1} = \left[\begin{array}{l} \left(\left[\underline{\left(x'_n << 2 + x'_n\right)} + \underline{\left(x'_{n+1} << 2 + x'_{n+1}\right)}\right] << 2\right) - \\ \underline{\left(x'_{n-1} << 2 + x'_{n-1}\right)} - \underline{\left(x'_{n+2} << 2 + x'_{n+2}\right)} + x'_{n-2} + x'_{n+3} + 16 \end{array}\right] >> 5 \end{cases}$$

Figure 1: Possible implementation of the described interpolation filters: (1) nearest neighbor, (2) bilinear, (3) bicubic, and (4) 6-tap

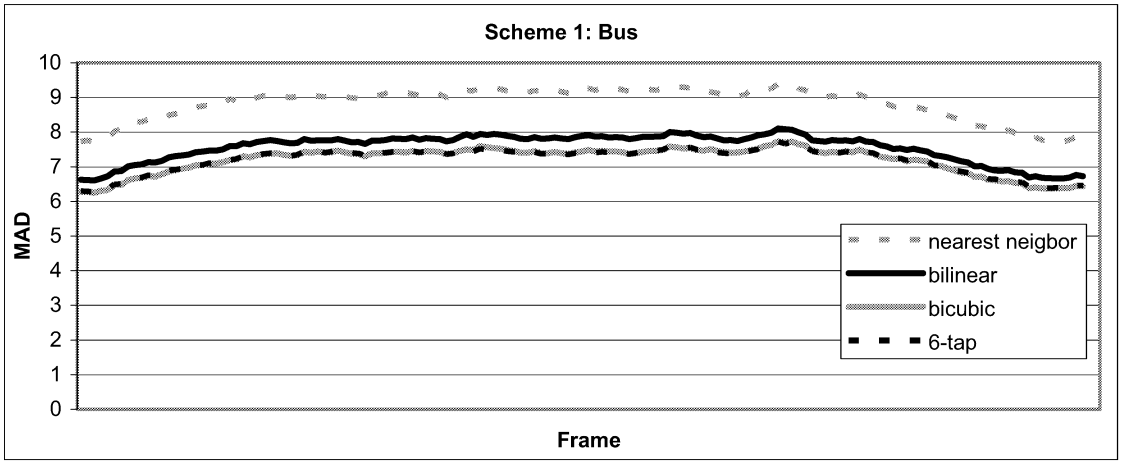


Figure 2: Scheme 1: MAD, sequence: Bus, QP base layer: 16

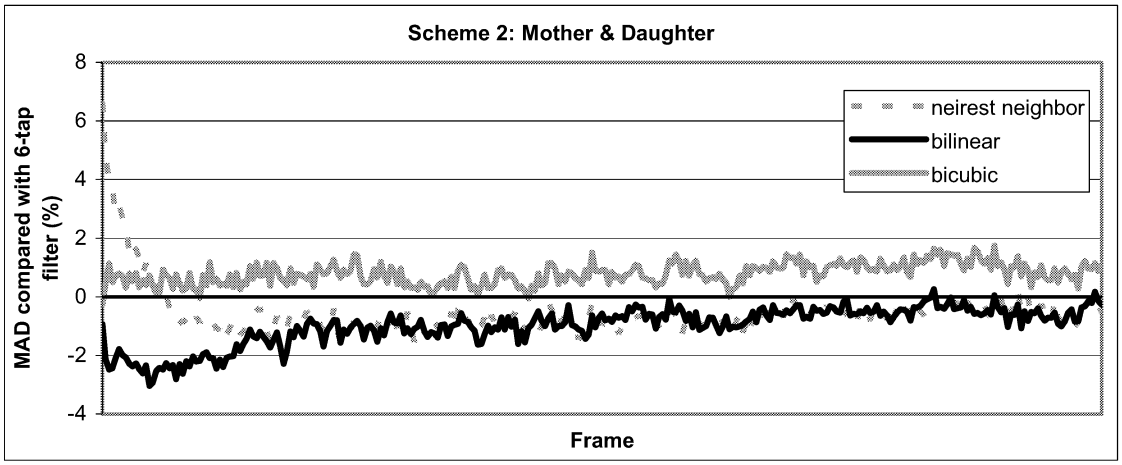


Figure 3: Scheme 2: MAD per picture plotted against the 6-tap filter (%),
sequence: Mother & Daughter, QP base layer: 16, QP enhancement layer: 24

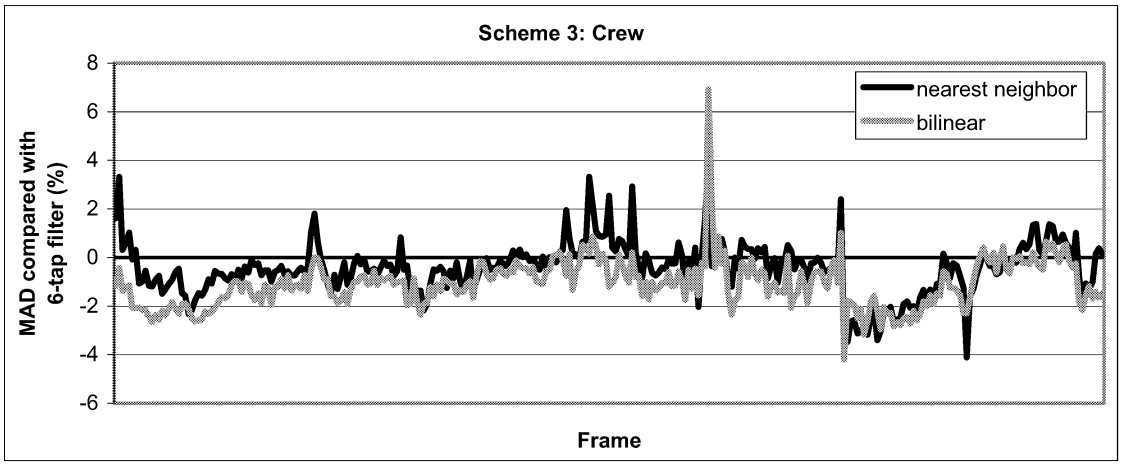


Figure 4: Scheme 3: MAD per picture plotted against the 6-tap filter (%),
sequence Crew, QP base layer: 32, QP enhancement layer: 36, interpolation base layer: bicubic

PROVIDING ACCESSIBLE AND PORTABLE VIDEO LECTURING USING CONTENT TRANSCODING

Silvia Mirri
Paola Salomoni
Department of Computer Science
University of Bologna
Mura Anteo Zamboni, 7 40127 Bologna, Italy
{mirri, salomoni}@cs.unibo.it

KEYWORDS
Multimodal interface, Usability interface and designs, E-learning, T-Learning Accessibility, Content Transcoding.

ABSTRACT

E-learning has the potential to reach students worldwide, making education a resource available to every user, anywhere, anytime, and from any network and device. However, many people will be excluded from the benefits of Web-based education if constrains to guarantee so wide an access are not considered in the design stage. Offering universal access to E-learning and T-learning does not mean sacrificing interactivity or appeal by producing text-based lectures. Furthermore, as technologies develop, rich media are becoming common resources for e-learning activities and a set of transcoding mechanisms could effectively improve portability and accessibility of on-line lectures. In this context, we present an application designed to support accessibility of on-line education that automatically produces accessible and portable versions of lectures, to be experienced on different networks, applications and devices.

1. INTRODUCTION

Wide access has long been regarded as one of the major benefits of Web-based education, but this is growingly proving to be the unfulfilled promise of these technologies. E-learning materials are frequently designed to be used with a specific client device, technology or configuration and this makes them less available to those who have limited access capabilities and use non standard equipment (AA.VV. 2003). There are two main reasons which encourage taking advantage of technology in whatever way possible to widen access to education:

i. Accessibility (*e-learning included*): learners with disabilities can benefit a great deal from e-learning, not just because it allows distance and flexible learning activities, but also because it helps impaired students to access resources which would otherwise present significant barriers to them. In particular, learners with reading difficulties, like people with visual disabilities, can greatly benefit from the integration of digital materials into teaching practice. Across the world, laws are in place or under definition to ensure that interactive/on-line services, and sometimes specifically e-learning, are made accessible to people who use non-standard access modalities, and in particular to people with disabilities (U.S. Rehabilitation Act 1998, Information Society European Commission 2002, Innovation and Technologies Italian Ministry Lucio Stanca 2004).
ii. Universality (*device-independence*): students can need or simply decide to use a specific access technology. This ranges from students using open source software on standard hardware platforms to people who are using mobile devices with non standard capabilities, or learners who want to access educational contents by using Interactive TV (*t-learning*). ITV is quite a recent and still unexplored development, which could facilitate e-learning adding value to this practice. The digitization that iTV offers is essential to any e-learning and therefore t-learning application. High quality sounds and pictures are prerequisites to this kind of services. Furthermore, the interactivity attribute of iTV allows to interact with other actors in the learning system, such as other learners (Lytras et al. 2004).

In order to meet these requirements two main strategies combined together are usually exploited:

- The use of technologies, compliant to world wide standards, to enhance interoperability and portability (W3C 2001, W3C 1999a).
- The use of alternative versions designed to accommodate different users, i.e. students accessing by using limited input device and/or students with sensorial impairments (W3C 1999a, W3C 1999b, IMS Global Learning Consortium 2002).

A large use of multimedia components is actually made to improve Web-based teaching efficacy, ranging from Flash animations to pre-recorded video-lectures.

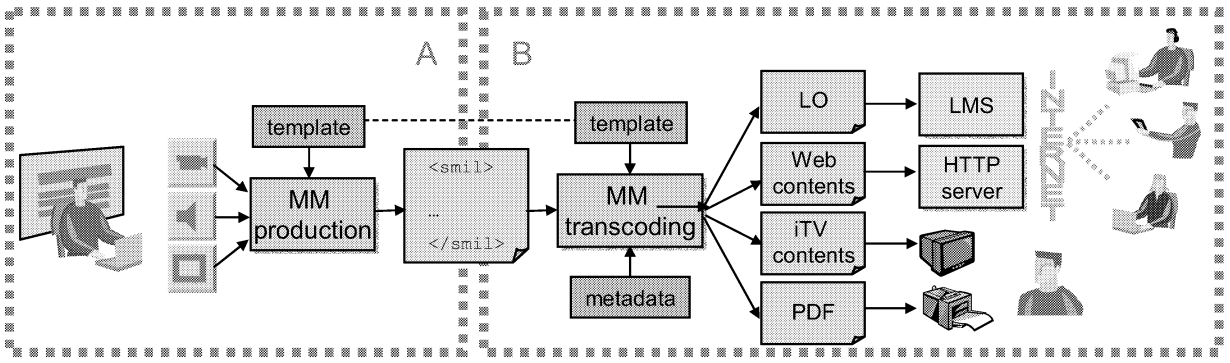


Figure 1: System Architecture

These rich media are often a useful addition to a Web-based course, used for example to describe a complex concept or to show a process flow. Sometimes audio and video elements are just a means to add a human touch to courses taught at distance. Finally, real lectures are frequently recorded to integrate or enrich on line materials. In all of the above mentioned situations the provision of fully accessible and portable multimedia contents is usually considered a daunting and time-consuming task (Burgstahler 2002). On the contrary, taking into account wide access requirements during the design process and using correct recording and transcoding mechanisms could help teachers to easily produce accessible and portable video lectures.

In this paper we present a system that supports on-line lecture accessibility and automatically produces different accessible/portable versions of audio/video educational materials. The on-line lecture consists in a rich media which is transcoded to provide equivalent content and improve computer accessibility to students equipped with non-standard hardware and software devices, including mobile terminals (such as PDA and phones), Interactive TV and platforms used by people with disabilities.

The remainder of the paper is organized as follows. Section 2 introduces the system offering an overview of its functionalities and architecture. Section 3 describes the transcoding process and Section 4 finally concludes the paper and briefly introduces future work.

2. ARCHITECTURE

Generally computer scientists use electronic slides to reinforce their lecture and offer students a simple support to review the lesson. From a technical point of view, a digital video lecture has to synchronize three main information flows:

i. The video track, showing the instructor.
ii. The audio track, playing the instructor talking.
iii. The slide sequence, simply considered as a sequence of images with text, tables or other graphical elements.

To ensure portability and accessibility, two information flows have to be added, maintaining synchronicity:

iv. The caption sequence; captioning is essential for deaf and hard-of-hearing students and for foreign students. Captioning is necessary also for students that are using devices without audio output.
v. The description sequence of slide contents (one description for each slide).

Synchronicity of all these flows is automatically detected by the presentation application, which captures all the events generated by the teacher to produce the time scheduling of all continuous media. As depicted in Figure 1, two phases emerge: a first one (A) allows the teacher to automatically produce a rich version of the lecture, synchronizing flows 1 to 5 into an audio/video presentation-based multimedia structure. A natural format for this synchronous multimedia object is SMIL (Synchronized Multimedia Integration Language), a standard markup language which allows media composition and synchronization. The SMIL multimedia presentation is realized in conformance with the SMIL accessibility features specified by W3C (SMIL 1999). The teacher produces and shows the slide sequence by using a personal productivity tool, such as Microsoft Power Point or OpenOffice Presenter. A speech recognition application can also be integrated with the above mentioned tool to create text captioning without human intervention. A second phase (B) starts from the SMIL description of the lecture and automatically produces a set of alternative contents, which are used to enhance portability and accessibility, such as:

1. *Web-based materials*, to be delivered through the Internet. Two versions of the lecture are produced, one to be distributed on-line on the Web and another encapsulated as a Learning Object (LO, compliant to SCORM 1.2 or 1.3). To enhance accessibility, both formats of the original content include a SMIL-based lecture and different XHTML-based transcoding, as described in the following Section 3.
2. *ITV materials*, i.e. MPEG movies that can be delivered through iTV channels. This continuous version of the lecture is post-produced with or without captions, as described in Section 3.

3. *Printed materials in PDF format*, obtained from the SMIL lecture by using XSL Formatting Objects (XSL-FO). This procedure is described in the following Section 3.

3. TRANSCODING

The different versions of contents delivered to users are computed by combining different transcoding processes, implicitly driven by the SMIL code that defines the contents structure. An example of SMIL code describing a lecture is reproduced in the following Figure 2.

```
...
<par>
  <video src="video/video.rm"
  region="region_video"/>
  <seq>
    ...
    <par>
      <img region="region_slide"
      src="img/2.gif" dur="60s"
      height="361" width="480"
      alt="Perche studiare questa
      tecnologia, 1"
      longdesc="img/2.htm"/>
      <audio region="region_audio"
      src="audio/2.mp3"/>
      <textstream src="caption/1.rt"
      region="region_subtitle"
      width="640" system-captions="on"/>
    </par>
    ...
  </seq>
</par>
...
<img region="region_slide" src="img/2.gif"
dur="60s" height="361" width="480"
alt="Fondamenti di elaborazione
multimediale" longdesc="img/2.htm"/>
<audio region="region_audio"
src="audio/2.mp3"/>
<textstream src="caption/1.rt"
region="region_subtitle" width="640" system-
captions="on"/>
</par>
...
```

Figure 2: A SMIL Code Fragment

Transcoding of single media is generally based on two different types of transformations (Salomoni 2004):

- *Translation*, which involves the conversion from one modality to another. Some translations can be automatically computed by using a transcoding function such for example *text to speech* (TTS) or *animation to image*. Other translations need an explicit declaration of content equivalence, as for example *image to text*.
- *Scaling,* which involves transcoding, recoding and compression and has effect in term of reduction of size, quality and data rate. Examples of scaling include image and video resizing and audio recoding and compression.

The translation process transforms the original SMIL lecture following two trends:

- *Reconstruct media continuity*: from the synchronized SMIL flow, the system has to produce a video containing all the information. This composition reconstructs continuity of the lecture that is completely driven by time.
- *Discretization*: from a synchronous SMIL source, different discrete or semi-discrete versions of the lecture have to be derived (fully accessible XHTML lectures and a printable PDF documents).

All the transcoding processes are based on a set of templates that are used to define layous and graphical aspects of the automatically produced lectures, as described in Figure 1 above. The reminder of this Section is devoted to describe the transcoding process by enumerating the different delivery mechanisms.

a. Web (mobile and accessible contents): SMIL multimedia composition is not accessible enough, especially for students that use assistive technologies such as screen readers or magnifying devices. Two main problems arise from this multimedia format:
   - SMIL uses rigid layout structures that are not adaptable to low resolutions (since they may need to be magnified for students with low vision) or to visualization on small devices (for students with non standard or mobile terminals).
   - The above described presentation produces a cognitive overload to blind users. Two (synchronous) audio tracks are technically available: the main audio track with the teacher's voice and another track containing a synthetized voice reading slide contents. Usually assistive technologies do not read dynamically changing textual contents, so the second track, and the information provided by it, is lost.

User testing phase has confirmed the need to provide alternative versions of the on-line lecture. A first alternative consists of an XHTML hypertext with textual slide contents (alternative for image-based slides) linked to the correspondent audio explanation. Blind students can use this version of the video lecture by reading slide text (with voice synthesis or Braille display) and, after or before that, listening to the teacher explanation. By alternating speech synthesis and human voice the entire content of the video lecture is provided as an auditory flow. Another alternative version is needed to meet requirements of students who need captioning and have difficulties reading synchronously changing text.. This could happen for example to deaf or hard of hearing students who consider teacher's speech too fast to be fully understood.

b. ITV (continuous contents): A video lecture that includes synchronized slides and captioning is easily derived from the SMIL playout.
c. *Printable contents*: printed materials in PDF format are obtained from the SMIL lecture using Formatting Objects Processor (The Apache XML Project 2005). The Formatting Objects Processor (FOP) uses standard XSL-FO file format as input, lays the content out into pages,

then renders it to the PDF output.

4. PRELIMINARY RESULTS

The system has been used for an e-learning project, called A^3, Accessible Learning Environment, in Italian Ambiente Accessibile d'Apprendimento (University of Bologna 2005), which provides contents and services to support Information Technology skills. A^3 offers different educational modules in blended e-learning. Educational contents are provided trough a LMS (Learning Management System) and for each module a brief video-lecture is supplied to introduce main goals and outline contents. In order to meet the accessibility requirements of the A^3 project, the video lecture has to be provided to users in a fully accessible form. The lecture is recorded and post-produced by an editorial equipe that uses a speech recognition application to automatically obtain captions.
The editorial equipe produces the SMIL version compliant to SMIL accessibility guidelines (SMIL 1999), as the one shown in Figure 2. The editorial equipe is also responsible to fill in the meta-information form, used to collect and structure SCORM metadata. Figure 3 shows a SMIL video-lecture and in particular a fragment that corresponds to the code showed in Figure 2 above.



Figure 3: An Online Synchronized Multimedia Lecture

It is worth noticing that it is essential to supply textual slide content in XHTML format with a transcoding mechanism. Both the original Power Point presentation and the HTML version that could be obtained by using a built-in "*Save as Web Page*" tool (Web Accessibility in Mind 2005) are not accessible. In particular, the automatically produced HTML is not compliant to accessibility guidelines and law requirements. Figure 4 shows the two different versions of the lecture implemented in XHTML, the first one with audio and textual slides and the second one with captions and textual slides. Finally, Figure 5 reproduces a frame of the MPEG lecture, produced for ITV delivery and Figure 6 depicts a page of the printable lecture in PDF format.

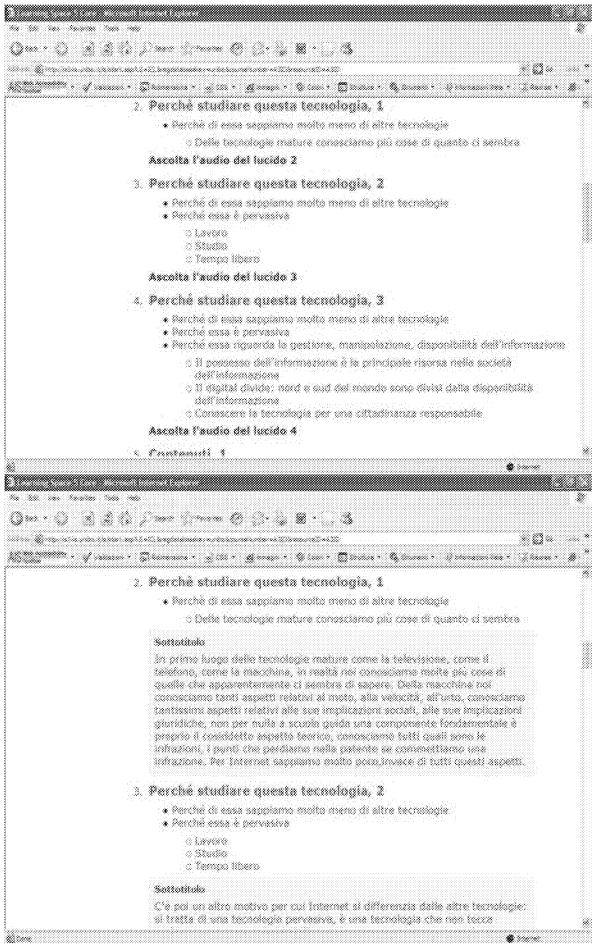


Figure 4: Alternative Versions

5. CONCLUSIONS

This paper presents a software application for the production of accessible and portable video lectures. Some improvements are still under development and in particular we are enhancing transcoding mechanisms to offer on-demand SMIL presentations, constraining screen dimensions to user needs. Other improvements will consider studies on how to provide teachers with open source presentation applications effectively integrated with Italian speech recognition system.
It is important to stress the importance of iTV as a means to involve excluded groups in learning activities by providing courses in an informal learning context. A television is usually placed in the social center of people's homes so, as well as taking learning directly to individuals, this medium allows education to become also a social activity. Unfortunately, currently iTV is not an accessible device.

Unlike computers, where adaptive technologies such as speech synthesizers, screen readers and Braille displays can be used to assist in screen interaction, television relies mainly on magnifying devices, such as monocular or screen enlargers.



Figure 5: The MPEG Lecture

Different perceptual challenges for the visually impaired are raised by other aspects of viewing that rely on a user's ability to navigate through channels, read text, identify visual characters, interpret moving objects and distinguish colors and shapes.



Figure 6: The printable lecture

The main accessibility features that may be problematic (highlighted by a recent study) include: ease of switching attention between screen and handsets, suitability of text and font sizes for on-screen reading, awareness of pauses and non-verbal actions, recognition of function keys on remote control handsets, appropriate feedback to selected iTV programmes and availability of audio descriptive services (Rice 2003). The layout of this version of the e-learning materials is currently still in a design phase, considering these accessibility problems.

REFERENCES

AA.VV. 2003. "Reusing Online Resources: a Sustainable Approach to eLearning". Allison Littlejohn & Simon Buchingham Shum.

Burgstahler B. 2002. "Distance Learning Universal Design, Universal Access". In *Educational Technology Review*, Vol. 10, No. 1.

Information Society European Commission 2002. "Web Content lic Sites in Europe". http://europa.eu.int/information_society/topics/citizens/accessibility/web/wai_2002/text_en.htm.

Innovation and Technologies Italian Ministry Lucio Stanca 2004. Law nr. 4 – 01/09/2004 (Official Journal nr. 13 – 01/17/2004).

IMS Global Learning Consortium 2002. "IMS Guidelines for Developing Accessible Learning Applications". http://www.imsproject.org/accessibility/accessiblevers/index.html.

Lytras M.; Lougos C.; Chozos P.; and Pouloudi A. 2004. "Interactive tTelevision and e-learning convergence: examining the potential of t-learning," *in proceedings of the 3rd European Conference on eLearning* (ECEL).

Rice M.D. 2003 "Television and visual impairment: prospects for the accessibility of interactive television". *In proceedings of the HCI International conference*.

Salomoni P. and Mirri S. 2004. "A multimedia broker for ubiquitous and accessible rich media content transcoding". *In proceeding of 1st IEEE International Workshop on Networking Issues in Multimedia Entertainment – NIME 2004* (Globecom 2004 Satellite Workshop).

The Apache XML Project 2005. "Formatting Objects Processor (FOP)". http://xml.apache.org/fop/.

University of Bologna 2005. "Ambiente Accessibile d'Apprendimento". https://a3.unibo.it.

U.S. Rehabilitation Act Amendments of 1998. "Section 508". http://www.webaim.org/standards/508/checklist.

Web Accessibility In Mind (WebAIM) 2005. "PowerPoint Accessibility Techniques". http://www.webaim.org/techniques/powerpoint/.

World Wide Web Consortium 1999a. "Web Content Accessibility Guidelines 1.0". http://www.w3.org/TR/WCAG10/.

World Wide Web Consortium 1999b. "Accessibility Features of SMIL". http://www.w3.org/TR/SMIL-access/.

World Wide Web Consortium 2001. "Synchronized Multimedia Integration Language 2.0". http://www.w3.org/TR/smil20/.

RECOGNITION SYSTEMS

MACHINE LEARNING TECHNIQUES FOR FACE ANALYSIS

D. Datcu and L.J.M. Rothkrantz
Man Machine Interaction Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands
E-mail: D.Datcu@ewi.tudelft.nl

KEYWORDS
Machine learning, pattern recognition, classifiers, face detection, facial expression recognition.

ABSTRACT

Facial related analysis represented milestones in the fields of computer vision for many decades. Lots of methods have been designed and implemented so as to solve the specific requirements. In the current paper we present three different classification algorithms that we use to fulfill the tasks concerning face detection and facial expression recognition. One of the methods, Relevance Vector Machines (RVM) stands for a novel supervised learning technique that is based on a probabilistic approach of Support Vector Machines. The mathematical base of the models is presented. The data for testing were selected from the Cohn-Kanade Facial Expression Database. We report recognition rates for six universal expressions based on a range of experiments. Some discussions on the comparison of different classification methods are included.

INTRODUCTION

Human computer interaction stands for a major step towards making machines have an even more important role to play in the human life. It is based on interdisciplinary researches that aim to implement knowledge from behavioral and social sciences into machines. It is the human nature that we can estimate a person's psychological state following the observation on his face. Nonverbal communication channels are typically set during common interpersonal relations and visual messages are processed in a transparent manner. The general tendency is to construct robotic systems that are able to understand the environmental world and to interact with the existent actors. Human-computer interfaces play an essential role in the perception and feedback the system is capable of. In this context, the advantage of making machines to read human facial expressions is tremendous. Facial expressions reveal internal characteristics of the expresser. To address the problem of facial expression recognition, in our approach we extract parametric information with high discrimination power from facial feature space and use it in a data-driven classification environment. The current paper primarily focuses on the aspects related to the classification methods for facial expression recognition. Secondly, techniques related to vision have to be involved for processing the video signal for detection of faces.
The classifiers are aimed at solving the universal problem of classification. We begin from Support Vector Machines SVM (Vapnik 1995) that is based on a solid mathematical foundation. The Naive Bayes classifier (Langley et al. 1992) is also introduced and practical aspects on the performance are presented. Then the novel Relevance Vector Machines RVM (Tipping 2000) is introduced as an alternative to SVM. The difficulty of the automatic analysis of facial expressions (Pantic and Rothkrantz 2000) resides in the variety of characteristic appearance with respect to both individuality and face anatomic dynamics. The inner complexity makes from the processes of feature detection and feature oriented expression recognition difficult tasks. To our knowledge, this is the first research that involves Relevance Vector Machines for facial expression recognition.

RELATED WORK

The recognition of facial expressions implies finding solutions to three distinct types of problems. The first one relates to detection of faces in the image. Once the face location is known, the second problem is the detection of the salient features within the facial areas. The final analysis consists in using any classification model and the extracted facial features for identifying the correct facial expression. For each of the processing steps described, there have been developed lots of methods to tackle the issues and specific requirements. Depending on the method used, the facial feature detection stage involves global or local analysis.
The internal representation of the human face can be either 2D or 3D. In the case of global analysis, the connection with certain facial expressions is made through features determined by processing the entire face.
The efficiency of methods as Artificial Neural Networks or Principal Component Analysis is greatly affected by head rotation and special procedures are needed to compensate the effects of that. On the other hand, local analysis performs encoding of some specific feature points and use them for recognition. The method is actually used in the current paper. However, other approaches have been also used at this layer.
One method for the analysis is the internal representation of facial expressions based on collections of Action Units (AU) as defined in Facial Action Coding System (FACS) (Bartlett et al. 2004; Ekman and Friesen 1978) . It is one of the most efficient and commonly used methodology to handle facial expressions. Some attempts to automatically detect the

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

salient facial features implied computing descriptors such as scale-normalized Gaussian derivatives at each pixel of the facial image and performing some linear-combinations on their values. It was found that a single cluster of Gaussian derivative responses leads to a high robustness of detection given the pose, illumination and identity (Gourier et al. 2004). A representation based on topological labels is proposed in (Yin et al. 2004). It assumes that the facial expression is dependent on the change of facial texture and that its variation is reflected by the modification of the facial topographical deformation. The classification is done by comparing facial features with those of the neutral face in terms of the topographic facial surface and the expressive regions.

A robust face detection technique was developed in (Viola and Jones 2001) based on a cascading classifier that include a set of so called 'weak' classifiers. The features stand for values of difference between the sums of pixel intensities computed in different areas in the image. Some approaches firstly model the facial features and then use the parameters as data for further analysis such as expression recognition. The system proposed by (Moriyama et al. 2004) is based on a 2D generative eye model that implements encoding of the motion and fine structures of the eye and is used for tracking the eye motion in a sequence. As concerning the classification methods, various algorithms have been developed (Pantic and Rothkrantz 2000), adapted and used during time. Neural networks have been used for face detection and facial expression recognition (Stathopoulou and Tsihrintzis 2004; deJong and Rothkrantz 2004). The second reference directs to a system called Facial Expression Dictionary (FED) (deJong and Rothkrantz 2004) that was a first attempt to create an online nonverbal dictionary. Other classifiers included Bayesian Belief Networks (BBN) (Datcu and Rothkrantz 2004), Expert Systems (Pantic and Rothkrantz 2000) or Support Vector Machines (SVM) (Bartlett et al. 2004). Other approaches have been oriented on the analysis of data gathered from distinct multi-modal channels. They combined multiple methods for processing and applied fusion techniques to get to the recognition stage (Fox and Reilly 2004).

VISUAL FEATURE MODEL

Although local analysis is sensitive to identity and partial occlusions, we overcome that by increasing the variability of data for training the model and by increasing the parameter redundancy. The variability is handled by the face characteristics in the Cohn-Kanade database (Kanade et al. 2000). The redundancy assumes the use of feature parameters on an asymmetric model, i.e. in case of occlusion or low visibility for one eye, the recognition is done by taking into account the data from the other eye. The task prior to classification stage is aimed at preparing the feature data associated with the input face. Depending on the type of the classifier involved in the next step, the data have to be transformed to a certain format.

In the current approach a transform $\Gamma : \theta \to \vartheta$ converts the facial feature image θ to some parameters $p_i \in \vartheta, i = 1,...,L$ of an intermediate model. The parameterization of the facial features has the advantage of providing the classifier with data that encode the most important aspects of the facial expressions. Furthermore, it acts as a dimensionality reduction procedure since the dimension of the feature space is lower than the dimension of the image space.

An advantage of the model is that it also can handle a certain degree of asymmetry by using some parameters for both left and right sides of the face. Each facial feature $\kappa \in \theta$, θ={Left/Right eye, Left/Right Eyebrow, Mouth, Chin} is extracted at the previous processing stage by distinct processing channels.

The transform Γ firstly extracts the location of each FCP from the input facial feature κ. Eventually the feature parameters p_i are computed as the values of some angles and/or Euclidean distances between key points assumed to reflect the location of the facial features.

The symmetry of the model is assumed to make the recognition process of facial expressions robust to occlusion or poor illumination i.e. if the left eye area is not directly visible do not use related information.

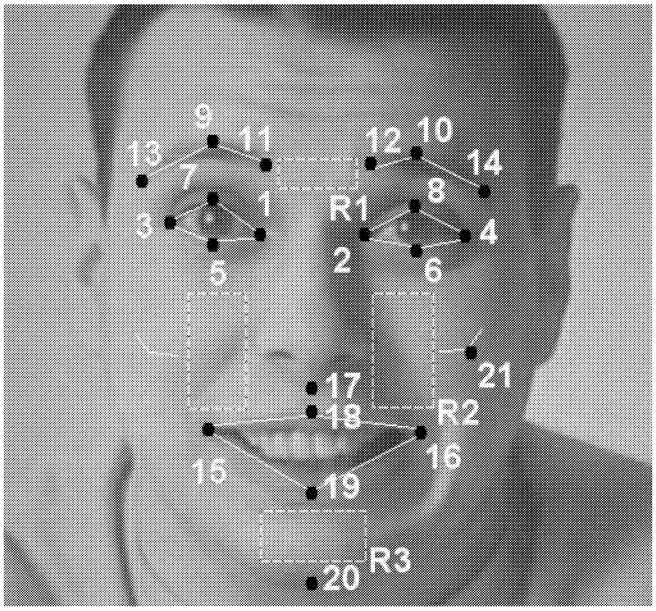


Figure 1: FCP set

The key points are defined as Facial Characteristic Points (FCPs) and the FCP-set (Figure 1) is based on an extension of Kobayashi & Hara model (Kobayashi and Hara 1972).

The final step of preprocessing was related to scale all the distances so as to be invariant to the size of the image. For the face detection, a procedure is run on each frame in the video sequence. First a 19 × 19 sliding window is defined. The processing for one window implies different tasks, according to the model used. If the image bitmap is used as an unique feature, then the set of pixel intensities is provided directly to the face/non-face classifier. Otherwise additional processing are required i.e. basic operations on pixel intensities (Viola and Jones 2001).

The cases when the dimension of an existent face is different than that of the window, a multiresolution pyramid is computed and analysis take place on each level. The process ends with the list of face locations in the current frame. An alternative consists in using feature detection and based on the findings to determine the location of any faces.

CLASSIFIERS

The current section is aimed at presenting the theoretical background of the classification techniques used in the research.

If X denotes the space of input variables representing the face images and T the space of output variables i.e. the facial expression label, then f is the associated deterministic function and $t_n = f(x_n) + \varepsilon_n$ represents the possible overlapping target values. The training database is:

$$Z = \{(x_i, t_i) \in X \times T \mid i = 1,..., M\}$$

The learning step implies the use of the training database Z together with any other prior knowledge for finding a function $\hat{f}$ out of a class of functions F, that encodes the estimated dependency. The process can be seen as a transform Ψ of the initial data ν into some internal knowledge ν' of the interest phenomena and so $\Psi : \nu \to \nu'$. From the notations used above, $Z \subseteq \nu$ and $\hat{f} \subseteq \nu'$. The result function $\hat{f}$ is assumed to produce an efficient classification of the facial expression label t given a new feature vector x.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) algorithm has been successfully used in classification related problems since it was introduced by Vapnik (Vapnik 1995) in the late 1970's. The idea was that given the collection of input-target pairs Z with $X \in R^n$ and $T = \{-1,+1\}$, a hyperplane $f \in F_h$, $F_h = \{f : X \to R \mid (w^T x) + b\}$ with the maximal margin has to be found as a solution of an optimization problem. The distribution of the two classes is such that they are linearly separable. The constraints aim at determining the model parameters $\{w, b\}$ that fit the training data and minimize the complexity of the decision function in the same time. The result is a classifier with a certain level of robustness to over-fitting. The margin represents a measure of class separation efficiency and is defined as the Euclidean distance between the data and the separating hyperplane.

Non-linearity that is specific to facial expression representation is handled through kernel methods (non-linear SVM) that first preprocess the data by non-linear mapping $\phi : R^N \to \varepsilon$ and then apply the linear algorithm in the image space ε. The image space is a vector space of functions $\varepsilon = \{f \mid f : X \to R\}$. The positive definite kernel function $k : X \times X \to \phi$ acts as a dot-product over ϕ and the mapping is expressed as $\phi(x) = k(\cdot, x)$.

Naive Bayesian Classifier (NB)

The Naive Bayesian (NB) is a probabilistic classifier based on the assumption that the set of model parameters are independent given the class parameter. An example E is presented as a tuple of attribute feature values $\langle x_1, x_2,..., x_n \rangle$. The parameter C represents the classification variable and c_j is a value of C. The probability of a sample being class c_j is:

$$p(c_j \mid E) = \frac{p(E \mid c_j) p(c_j)}{p(E)}$$

The Bayesian approach to classifying a new instance relates to assigning the most probable target value, given the attribute values $\langle x_1, x_2,..., x_n \rangle$ describing the instance.

$$f_b(E) = \arg\max_{c_j} \quad p(c_j \mid x_1, x_2,..., x_n)$$
$$\cong \arg\max_{c_j} \quad p(c_j \mid x_1, x_2,..., x_n) p(c_j)$$

where $f_b(E)$ is called a Bayesian classifier. Considering that all the attributes are independent given the value of the class variable, then:

$$p(E \mid c_j) = p(x_1, x_2,..., x_n \mid c_j) = \prod_{i=1}^{n} p(x_i \mid c_j)$$

By rewriting, the previous formula of the classifier becomes:

$$p(E \mid c_j) = \arg\max_{c_j} \quad p(c_j) \prod_{i=1}^{n} p(x_i \mid c_j)$$

The function $f_{nb}(E)$ is called the Naive Bayesian classifier or naive Bayes (NB). Naive Bayesian is the simplest case of Bayesian network. The assumption of conditional independence that is the independence of attribute variables given the class variable is not properly fulfilled for the real case. For facial expression recognition, some of the parameters $p_i \in \vartheta, i = 1,..., L$ do not follow the independence assumption because they measure the behavior of related components of the same facial features. Moreover, the parametric changes of different facial features are correlated due to the dynamics of facial expressions. Though, the naive Bayes classifier performs very well even in these cases. An explanation given for the performance of the binary NB classifier is presented in (Zhang 2004).

Relevance Vector Machines (RVM)

Tipping introduced the Relevance Vector Machine (RVM) (Tipping 2000) as a probabilistic sparse kernel model based on the support vector machine theory. Each of the model's weights has associated a prior that is characterized by a set of hyperparameters whose values are determined during the learning process. Following the above notations, $p(y \mid x)$ is assumed to be Gaussian $N(y \mid f(x), \sigma^2)$ and the mean of the distribution is computed as specified for the SVM. The overfitting effect that occurs while determining the parameter values of the likelihood $p(y \mid w, \sigma^2)$ is confined by including an ARD Gaussian prior over the weights w_i as

$$p(w \mid \alpha) = \prod_{i=0}^{N} N(w_i \mid 0, \alpha_i^{-1})$$

The dataset likelihood is expressed by using logistic functions, in the form:

$$p(w \mid w) = \prod_{n=1}^{N} \sigma\{f(x_n)\}^{y_n} [1 - \sigma\{f(x_n)\}]^{1-y_n}$$

The parameters α_i and w_i of the model are computed through an iteration procedure until convergence is achieved. One advantage over SVM is that for comparable generalization performance, it uses fewer kernel functions. This determines less memory and time for processing and so making possible for the usage of RVM as a real-time classifier. By using RVM, the relevance vectors stand for representative training samples of the emotional classes rather than data points closer to the separation hyperplane as in SVM model.

EXPERIMENTS

For testing the face detection by using different classifiers, the CBCL (the Center for Biological and Computational Learning, MIT) face database was used. It contains 19×19 image samples. The training set contains 2429 faces and 4548 non-faces and the testing set contains 472 faces, 23573 non-faces. The classification process implied the use of all the image map as a unique feature. Testing face detection with RVM classifier for 1000 black-white samples in the database and using 5-fold cross validation, we achieved a detection rate of 99.20%. The number of relevance vectors was 392.

Ongoing work aims at using the method presented in (Viola and Jones 2001) for determining a set of representative features to be used for classification. More exactly, each feature would stand for one parameter computed as a difference between the sums of pixel intensities over some interest areas. The idea is to have a set of such features that together would help for discriminating between a face or non-face.

Table 1: Recognition mismatch rate for facial expression recognition by using SVM classifier

Expression	Mismatch rate	Number of support vectors
Surprise	14.32 ± 1.80%	63
Sadness	3.06 ± 2.78%	23
Anger	5.91 ± 1.55%	46
Happy	3.16 ± 2.47%	38
Disgust	9.54 ± 1.97%	34
Fear	24.97 ± 2.07%	72

An initial process of automatically generating features is run in the image space. Then AdaBoost (Adaptive Boosting) technique (Freund and Schapire 1995) is used to select only the most expressive features from the initial generated set. The final 'strong' classifier is expressed as a sequence of 'weak' classifiers, each classifier relying on only one certain feature from the selected set. The final cascading detector is created using a weighted voting scheme so as the weight of each classifier depends on its performance on the training set.

The data used for training the models for the facial expression recognition experiments have been processed from the Cohn-Kanade database. The database contains approximately 2000 image sequences from 200 subjects ranged in age from 18 to 30 years. Sixty-five percent were female, 15 percent were African-American and three percent were Asian or Latino.

In the first step, only 485 images have been used for the experiments, each image representing a sample data for the model. Then the classification was improved by adding new samples. In the case of Naive Bayes, the error dropped from 33% to 26.89%.

Table 2: Recognition mismatch rate for facial expression recognition by using Naive Bayes classifier

Expression	Mismatch rate
Surprise	11.66 ± 7.29%
Sadness	22.35 ± 5.98%
Anger	31.01 ± 12.60%
Happy	17.76 ± 7.79%
Disgust	38.82 ± 12.85%
Fear	39.75 ± 16.95%

Some specific steps were passed to extract and prepare the data to comply with the requirements of the classification process. The algorithms are set to perform classification on the 6 universal expressions (Happy, Anger, Sad, Surprise, Disgust, Fear). The classification results of the facial expression recognition can be analyzed by looking at the mismatch rate. The method used for computing the error was leave-5-out cross validation. The distribution of samples in the database for testing was as follows:

N(Surprise, Sadness, Anger,Happy,Disgust, Fear) = (108, 92, 30, 110, 59, 86).

Table 3: Recognition mismatch rate for facial expression recognition by using RVM classifier

Expression	Mismatch rate	Number of relevance vectors
Surprise	6.25 ± 1.51%	21
Sadness	12.29 ± 2.57%	34
Anger	5.00 ± 1.87%	15
Happy	7.92 ± 1.71%	23
Disgust	8.54 ± 1.91%	25
Fear	15.00 ± 2.38%	38

As it can be seen (Table 1 and Table 3), the error rate in the case of RVM (9.16%) is comparable to that of SVM (10.15%) classifier. One important aspect is that in case of RVM classifier the number of relevance vectors (276) is greater than the number of support vectors (156) of SVM. The effect is a decrease of the number of kernel functions and so of the complexity of the model. In terms of practical characteristics that means less processing time and also less memory for using this type of classifier.

The Naive Bayes classifier stands for an easy way to predict the class of an unseen sample by means of integrating

probability knowledge computed from the known examples. However, the performance of NB is based on the training data. The training stage implies the computation of the dependencies of each feature given the sample class. From this, the system learns about the dependency distributions of the data. If the training set does not contain enough data for learning, the result would be an approximation of the distributions that does not exactly fit the real ones. The degree of error on the real distributions greatly affects the classification results. For testing NB classifier for facial expression recognition, the cross fold validation procedure was used, where the number of folds was 10. It also implied an increased number of samples as N(Surprise, Sadness, Anger,Happy,Disgust, Fear) = (108, 206, 105, 110, 108, 86). Because of that, the comparison with the other techniques may be improper. The facial recognition general mismatch rate for NB is 26.89% (Table 2). Nevertheless, the analysis of facial expressions in static images has its own limitations. That can be mainly explained through the dynamics characteristics of the salient features involved in facial expressions' structure. An important improvement for the recognition system may include also the encoding and usage of the knowledge over these elements (Datcu and Rothkrantz 2004).

CONCLUSION

Facial expression recognition has scientifically been considered a real challenging problem in the fields of pattern recognition or robotic vision. The current research aims at proposing Relevance Vector Machines (RVM) as a novel classification technique for the recognition of facial expressions in static images. The results presented highlight the potential of the Relevance Vector Machines as a facial expression classifier and for face detection. The exemplifications start from the idea of the Support Vector Machines and addresses the issues concerning the use of two types of classifiers in the context of facial expression recognition problem. The RVM is a relatively new classification method and this work is the first one that uses the technique as a recognition engine for facial expressions. The fundamental aspects are described on both theoretical and practical sides. Each classifier model presents certain advantages and limitations and have been designed so as to perform prediction on the static images. The results for RVM show that it is suitable for facial expression classification in static images and it leads to a decrease of complexity comparing to SVM. The still image analysis is very restrictive with respect to the subtle dynamics of the facial features.

Additional research has been conducted to encode temporal behavior in the classification models so as to make possible the use of the recognition systems to run on image sequences. Another idea for increasing the capabilities and efficiency is to make use of fusion techniques to handle multiple modalities.

REFERENCES

Bartlett, M.S.; G.Littlewort; C.Lainscsek; I.Fasel; and J.Movellan. 2004. "Machine learning methods for fully automatic recognition of facial expressions and facial actions". Proceedings of IEEE SMC. 592–597.

Datcu, D.; and L.J.M.Rothkrantz. 2004. "Automatic recognition of facial expressions using bayesian belief networks". Proceedings of IEEE SMC. 2209–2214.

Ekman, P.; and W.V.Friesen. 1978. "Facial action coding system:investigator's guide". Consulting Psychologists Press.

Fox, N.A.; and R.B.Reilly. 2004. "Robust multi-modal person indentification with tolerance of facial expression". Proceedings of IEEE SMC. 580–585.

Freund Y.; and R.E.Schapire.1995. "A decision theoretic generalization of on-line learning and an application of boosting". Proceedings of the Second European Conference on Computational Learning Theory. Springer-Verlag. 23–37.

Gourier, N.; D.Hall; and J.L.Crowley. 2004. "Facial feature detection robust to pose, illumination and identity". Proceedings of IEEE SMC. 617–622.

Kanade, T.; J.Cohn; and Y.Tian. 2000. "Comprehensive database for facial expression analysis". Proc. IEEE Int'l Conf. Face and Gesture Recognition. 46–53.

Kobayashi, H.; and F.Hara. 1972. "Recognition of mixed facial expressions by neural network". IEEE International workshop on Robot and Human Communication. 381–386.

Langley, P.; W.Iba; and K.Thompson. 1992. "An analysis of bayesian classifiers". Proceedings of Tenth National Conference on Artificial Intelligence. AAAI Press and MIT Press. 223–228.

Moriyama, T.; J.Xiao; J.F.Cohn; and T.Kanade. 2004. "Meticulouslydetailed eye model and its application to analysis of facial image". Proceedings of IEEE SMC. 580–585.

Pantic, M.; L.J.M.Rothkrantz. 2000. "Self-adaptive expert system for facial expression analysis". Proceedings of IEEE SMC. 73–79.

Pantic, M.; L.J.M.Rothkrantz. 2000. "Automatic analysis of facial expressions: the state of the art". IEEE Trans. PAMI. 22(12).

Stathopoulou, I.O.; and G.A.Tsihrintzis. 2004. "An improved neural-network-based face detection and facial expression classification system". Proceedings of IEEE SMC. 666–671. deJong, E.J.; and L.J.M.Rothkrantz. 2004. "Fed - an online facial expression dictionary". Proceedings of Euromedia. 115–119.

Tipping, M. E. 2000. "The relevance vector machine". In Sara A Solla, Todd K Leen, and Klaus-Robert Müller, editors. Advances in Neural Information Processing Systems 12. Cambridge. Mass: MIT Press.

Vapnik, V. 1995. "The nature of statistical learning". Springer, New York.

Viola, P.; and M.Jones. 2001. "Robust real-time object detection". Workshop on Statistical and Computational Theories of Vision-Modeling Learning, Computing and Sampling.

Yin, L.; J.Loi; and W.Xiong. 2004 "Facial expression analysis based on enhanced texture and topographical structure". Proceedings of IEEE SMC. 586–591.

Zhang, H. 2004. "The optimality of naive bayesian".

A WEARABLE GESTURE RECOGNITION SYSTEM FOR NATURAL NAVIGATION INTERFACES

Elisabetta Farella, Luca Benini, Bruno Riccò
DEIS - University of Bologna
40136 Bologna, ITALY
efarella,lbenini,bricco@deis.unibo.it

Andrea Acquaviva
ISTI - University of Urbino
61029 Urbino, ITALY
acquaviva@sti.uniurb.it

KEYWORDS
HCI, Inertial sensors, gesture recognition.

ABSTRACT

Ambient intelligence applications pose the challenge of enabling real-time interaction with virtual environments in a natural manner. Gesture recognition based on body-mounted accelerometers has been proposed as a viable solution to translate patterns of movements that are associated with user commands, thus substituting point-and-click methods or other cumbersome input devices. These wearable systems are a promising approach to gesture recognition because of their low cost and size, high robustness and minimal constraints on user location and freedom of movements.
In this work we present a gesture interface system based on low-cost integrated accelerometers for navigation in virtual spaces. Qualitative and quantitative assessments are presented using a 3D game application as a test-bed to evaluate the effectiveness of the interface.

INTRODUCTION

The aim of Ambient Intelligence (AmI) technology is to provide personalized services tailored to user context. Where possible, this target should be achieved through understanding of user needs without an explicit request. Where a request is mandatory, this should be as close to human-to-human interaction as possible. For this reason, one of the enabling technology is motion and gesture recognition. An intelligent space may allow the user to experience a virtual environment or playing immersive game. These applications critically require natural human-computer interfaces (HCI), which reduce the gap between virtual-world and real-world interaction. In many applications that require navigation in virtual environments for learning or gaming, the quality of the user experience depends on the rendering quality but also on the interaction with the system. A successful HCI system must guarantee real-time feedback to the user and must be neither intrusive nor cumbersome (Cheok et al. 2002). Furthermore, to allow deployment on commercial-off-the-shelf computer systems, it must be designed with low cost components.
In this work, we describe a novel low cost and wearable gesture recognition system for natural navigation interfaces. We present also the qualitative and quantitative performance assessment methodology we used to evaluate its effectiveness. The system integrates hardware and software components. The hardware architecture is completely based on commodity components (Barbieri et al. 2004). Movements are detected by body mounted integrated accelerometers. Data transmitted by sensing units are sampled and collected by an acquisition/control board that communicates to the machine running a virtual world application chosen as a test-bed by mean of a serial wired or wireless link. The limited size and weight of the acquisition board makes it suitable to be installed on a palmtop computer or PDA, providing unconstrained mobility. Communication between sensing units and the acquisition board takes place by means of a dedicated wired link.
Cost of the interface hardware is lower than $100 (even in small volume prototyping). The proposed solution is wearable and flexible, as the acceleration sensors can be placed anywhere on the body. This is in contrast with systems targeting specific body parts, such as the well known glove applications (Perng et al. 1999) (Hernandez-Rebollar et al. 2002). The software part of the interface is responsible for filtering and translating accelerometer samples into commands for the virtual environment using an object oriented approach. In addition, it takes care of the initial calibration phase.
Performance assessment has been carried out in two different ways. First, the performance of the interface in terms of response time and movement loss has been tested. Then, a 3D game was used as a test-bed to evaluate usability and user experiences. The contribution of this work is thus twofold. First, we describe the hardware and software design of our interface system. Second, we provide extensive analysis of its usability, demonstrating its effectiveness in real applications.

RELATED WORK

Both vision-based and accelerometer-based techniques have been proposed for gesture recognition (Cheok et al. 2002), (Oakley et al. 2003), (Perng et al. 1999) (Pantic and Rothkrantz 2003). One general weakness of optical tracking systems is that there must be line-of-sight between the camera and the tracked object. (Regenbrecht et al. 2001) (Tanrivedi and Jacob 2000) (Headon and Curwen 2002).
Accelerometer based solutions, even though less accurate than optical tracking, are a promising alternative for gesture recognition in mobile environments. Many types of input hardware have been proposed. Perng et al. 1999 present a system interface dedicated to hand movement recognition was presented to allow mouse-like input using an accelerometer glove. In this solution, the recognition task is dedicated to a particular body part. In other solutions accelerometers are mounted on the wearable device to recognize particular commands such as scrolling or page change (Oakley et al. 2003). Another dedicated solution is presented in (Bang et al. 2003). It consists of a pen system based on accelerometers and

gyroscopes. The system is aimed at recognizing handwriting motion in 2D and 3D spaces. One major problem addressed in this work is the compensation for inaccuracies in position and orientation detection due to the integration process. A compensation technique called Zero-Velocity (ZVC) is exploited. This technique however cannot be easily exported to a more general purpose context.

Closest to our work, a micro-accelerometer based interface for interaction with mixed reality in wearable computers has been presented by Cheok et al. 2002. Accelerometers are mounted on various kinds of input devices: a pen and a tilt and gesture pad. They provide 3D information about motion of the body part on which they are worn. Compared to this, our work differs in two aspects. First, we do not perform integration on the accelerometer data, instead relying on a geometrical approach. Our approach eliminates the need for compensating for the error introduced by the integration step used in (Cheok et al. 2002): this is a critical advantage since noise cancellation from double integration is an extremely challenging task, for which no general solution is known.

The second distinction concerns qualitative assessment. In our work, we targeted the evaluation of gesture recognition effectiveness for interaction to virtual reality applications such as 3D games. Instead of just focusing on accelerometer accuracy, we provide a full evaluation of user experience by measuring the number of recognized movements for different user behaviors.

SYSTEM ARCHITECTURE

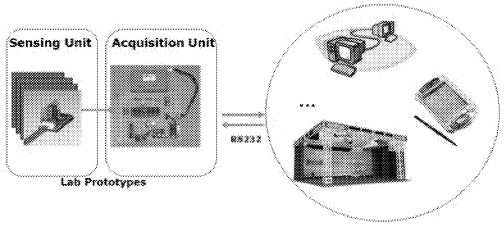


Figure 1. The Gesture Recognition Interface can Interact with Different Hardware Targets

The general system architecture is composed of a fixed system back-end and a flexible and reconfigurable front-end.

The system back-end consists of the MOCA prototype (one acquisition board and up to four sensing units), the MOCA interface and the implementation of the tracking algorithm. It is responsible for collecting and interpreting data (from basic driver operation, to position recognition or complex processing such as gesture recognition). The front-end is the target application, which can take many forms from simple visualization of the data collected to sophisticated interaction with a virtual environment.

MOCA2 Back-End

Sensing/Acquisition Board. MOCA2, the sensing and data acquisition hardware, is an evolution of MOCA, a MOtion CApture system (Barbieri et al. 2004), built entirely with low-cost commodity components. MOCA2 is based on two boards: (i) sensing units and (ii) acquisition board. The system prototype is shown in Figure 2. The acquisition board is equipped with a RISC microcontroller with integrated analog-to-digital converter (ADC) and UART controller, a serial port and an external ADC. Each sensing unit contains two different MEMS (Micro Electronic Mechanical System) accelerometers. In order to collect 3D acceleration vector, a mono-axial sensor is mounted orthogonally with respect to bi-axial accelerometer.

The acquisition board is responsible for data sampling, conversion from a set of sensing units, processing and transmission to the gesture recognition software (on a wearable computer or a PC) through a serial link. Acquisition takes place exploiting both internal and external ADC in an optimized way. The processing step involves the arrangement of data samples into packets with marking stamps indicating the beginning and end of one acquisition step. For each step, data from all accelerometers are collected.

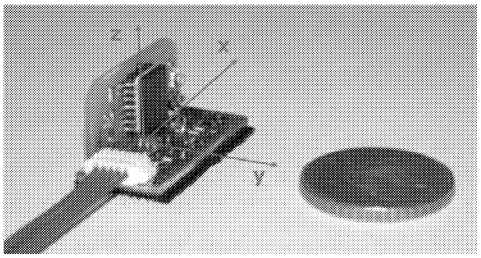


Figure 2. The Sensing Unit

The acquisition process accounts for a time interval short enough (about 350μs from start of sampling to the end of the serial transmission) with respect to data sampling rate (10ms for 100Hz accelerometer bandwidth), to leave the rest of the time for local processing.

A single acquisition board supports up to 4 sensing units in a time-division multiplexed pattern. The power consumption of a single acceleration unit is 12 mA, while the acquisition unit consumes 46 mA. Thus a single 6V battery can support the system in continuous operation for at least 12 hours. The total component cost for the two boards is around $100, where each sensing unit accounts for around $10.

MOCA2 improves upon MOCA in two directions. First and foremost, the sensing units (Figure 2) have been aggressively miniaturized (the new board is a factor of five smaller than the first version). Miniaturization implies much better wearability: the back of the board is glued to a velcro strip and can be easily tied, with a ribbon, on fingers, arms, or any other part of the body. Another improvement is on power dissipation of both boards, which has been cut by approximately a factor of 2, thanks to clock gating and shutdown of idle components.

MOCA2 interface. We implemented two alternative drivers as support for interfacing with the prototype: a Linux driver and a small set of Java classes.

Both drivers or interface layers are responsible for controlling the acquisition and process data packets received from the sensing/acquisition board and extract the payload. A specific task is implementing the handshaking protocol required to communicate with the MOCA2 board. The handshaking

protocol consists in sending a start character to the microcontroller to trigger the beginning of the acquisition and sending a stop character to cause the board to go in waiting state or completely stop the acquisition.
After payload extraction, the driver performs some minimal processing to convert input data, which are expressed as voltage values, into acceleration values. Finally, the driver configures some parameters such as gain and offset with default values. These parameters are needed to tune the above mentioned conversion process.
The Java solution is motivated by the need to guarantee platform independence. The Linux driver implemented offers a standard application level interface to communicate with and handle the prototype. Developing a specific Linux driver guarantees better efficiency than the JAVA based solution, as the processor does not experience the workload generated by the JAVA virtual machine.

MOCA2 Front-End

The front-end of the MOCA gesture recognition system perform recognition of gestures. Once detected, an interface layer translates them into application specific commands. The application we target in this work is the navigation in a virtual environment. Our aim was to test the interaction system, thus we tried a subset of the many possible interaction metaphors. For the same reason, more attention was given on the easy and fast deployment of a test-bed virtual environment than on the performance of the graphical libraries. Future work will implements a more complete set of controls and care will be given to the implementation of the target application.

Gesture recognition framework. The gesture recognition framework relies on the Java MOCA driver to obtain acceleration values. Information on the users' movements is derived exploiting the recognition algorithm described later. The system is flexible, in that it allows the specification of many different patterns that will be recognized as specific gestures. For the sake of clarity we describe a practical example, namely a simple gesture recognizer for single arm gestures. In this setup, there is a single sensing unit, tied to the user's wrist. Arm gestures are defined to control navigation in the virtual world described in the following subsection. We defined 5 gestures, namely "rest" position, "turn right", "turn up", "turn left", "turn down". This is only a small subset of possible gesture and since the system can distinguish between at least 20 different inclinations, and consequently gestures, in a range of 180° degree around a single axis, as demonstrated in the experiments section. Relying on tilt provides to the recognition system independence from sensor location and to recover the gesture performed we do not need complex human body models.
The system translates acceleration data in inclinations by computing the components of the gravity vector g along three orthogonal axes and avoids the use of integration. Integration should be avoided whenever possible in accelerometer-based systems, because it leads to accumulation of measurement and sensor noise. Integration-based systems require frequent re-calibration phases (Cheok et al. 2002) and they suffer from significant time drift.
We based the recognition on an algorithm specialized for 2 Degrees of Freedom (DoF) systems (Lee and Ha 1999). In this case, the acceleration value in one point is exploited to obtain inclination information. The system is assumed to be in steady state, i.e. the hypothesis made is that the gravitational acceleration is the main acceleration component. The particular angle which identifies a position is derived as a composition of trigonometric functions of the accelerations in one or more points.
Figures 3 shows the case of measuring tilt from two acceleration values, where the two angles Ω and θ determinate the inclination of $\overline{g}$ [1], the gravitational vector, with respect to the plane π defined by the axis $\hat{x}_1$ and $\hat{y}_1$ of two accelerometers. Let us call a_x and a_y the accelerometers outputs, and $g_\Pi = \sqrt{a_x^2 + a_y^2}$. The geometric interpretation of g_Π is the module of the projection $\overline{g}_\Pi$ of $\overline{g}$ onto π. Thus, we obtain Ω and θ as follows:

$\cos\Omega = a_x / g_\Pi$ (or in alternative $sen\Omega = a_y / g_\Pi$)

$\cos\theta = g_\Pi / g$

where g (9.80665 ms^{-2}) ax and ay are known.

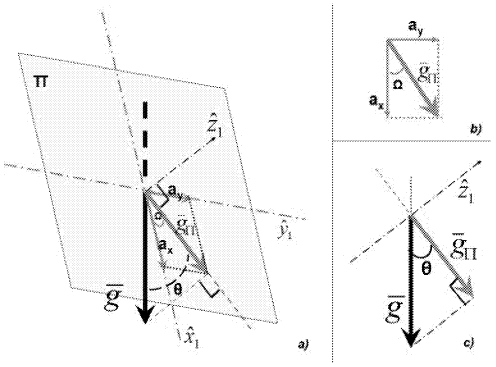


Figure 3. Measuring Tilt from Two Acceleration Values

The accuracy of the inclination becomes smaller as the angle between the axis of measurement and the direction of g decreases. This is due to the way accelerometers are used to compute inclinations (Perng et al. 1999). To overcome this problem, a simple solution consists in placing two accelerometers in such a way that measurement axes are orthogonal each other. As a consequence, one of them is always in an error-free inclination (Barbieri et al. 2004).
After normalization, orientation values are first averaged and then translated into gesture. To decrease sensitivity due to movement inaccuracies, a range of valid acceleration values has been defined for each gesture, instead of identifying with a single value. A so called *recognition window* corresponds to the extension of this range.

[1] In general $\overline{v}$ is a vector, v its module with sign and $\hat{v}$ its versor, so that $\overline{v} = v \cdot \hat{v}$.

The interface system defines also another type of window that will be referred as *time window*. The time window size is identified by the sampling period times the number of samples, say *m*, (equal for all acceleration components) used to average sampled data. Since the accelerometer bandwidth is 100Hz, 10 ms elapse among two different sampling, so that time window is proportional to 10 ms through *m*.
The application interface layer derives control events from the gesture recognition process and it maps them in understandable commands for the end-user application.



Figure 4. Screen Captures of the Virtual World, as Seen from an LCD while a User is Interfacing with it Through MOCA2

Navigation Environment. The navigation application chosen as a case study is a Java3D Virtual World enabling multi-client interaction. On the application side, commands corresponding to recognized gestures are converted by an input interface layer responsible into events known by the rendering engine (Figure 4). Use of Java 3D helps in rapid deployment of an application prototype executable on a number of different hardware targets. The application creates a virtual world based on a 2-dimensional map input with additional annotations to describe textures and various objects placed in the virtual world. Each user is represented as simple avatar (3-D geometric shapes). The application supports multiple viewpoints for each user and it is capable of being rendered on multiple, distributed visualization systems.

EXPERIMENTS

The test setup for our system is composed by two sensing units, each using a combination of a MEMS biaxial (ADLX202) and a MEMS mono-axial (ADXL105) accelerometer by Analog Devices. The range for the ADLX202 is ±2g and for ADXL105 is ±5g (where g is the gravitational acceleration). The acquisition board is based on an ATMEL AVR AT90S8535 microcontroller, working at 8 MHz, equipped with an internal 10 bit A/D converter and a UART module; on the board we placed also a 16 bit, 200kSPS A/D converter (AD974 by Analog Devices).
Tests were performed using the MOCA2 board with two sensing units, a laptop PC equipped with a Pentium 4, Windows XP© Operating System, and a stepper motor to perform test at different controlled speed.
Two groups of tests were performed. The first evaluated the user-perceived quality of interaction. The second quantitatively measured the recognition loss with the help of a stepper motor.

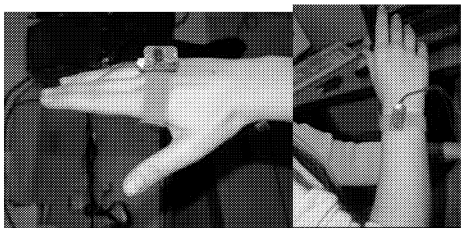


Figure 5. Wearable Setup Used in Human-Based Tests

The recognition window was the same size for each gesture and was kept constant during the tests. The performance of the system was evaluated as a function of various parameters, namely speed of movement, settling time and time window. Settling time refers to the time a user remains in a certain position.
For both groups of tests, a two step calibration process was performed before starting the measurements. The calibration process is fast and it is needed only once after placing the sensing unit along the body the first time. During the first step, the sensing unit was positioned so that the biaxial accelerometer was orthogonal to gravity. In the second step the same procedure was repeated for the mono-axial accelerometer. This was used to center the voltage range with respect to the horizontal position.
*User Experience***.** A group of ten users (aged between 23 and 30) were chosen. Quality of user interaction was evaluated using two metrics, namely gesture loss and general user satisfaction.
Users were asked to wear one of the sensing units on their wrist or on the back of their hand, depending on their personal preference. The positions are illustrated in Figure 5. They were asked then to perform a sequence of gestures such as moving their hand up, down, right and left. To increase the range of movements tested, users were requested to perform the same sequence once at a natural pace (defined as 'spontaneous' in the tables and plots), once fast and once slowly. Visual feedback was provided to indicate the result of the recognition process. The test was repeated for each user with different sizes of time window.
Results of the first test are shown in Table 2 and plotted in Figure 6. For each sequence under test the percentage of successful recognitions was determined. Both not recognized and not correctly recognized gestures were considered as errors.
As expected, when the user moves fastest (right-most column of Table 2) system performance is at its worst. Clearly, for a fixed time window (rows in Table 2) the number of samples falling in the recognition window increases as the movement speed decreases, thus helping the recognition. Looking at each column in Table 2, we observe that lower values of time windows provide a better performance, since the system is less sensitive to previous values caused by movements. In fact, to perform the recognition, only more recent samples are considered.
The minimum time window represented in Table 2 is 50ms. Smaller values are not represented since almost all recognitions failed. This is due to the noise introduced by the user stopping in incorrect positions at the end of the movement. This noise is not averaged out for very small time windows.

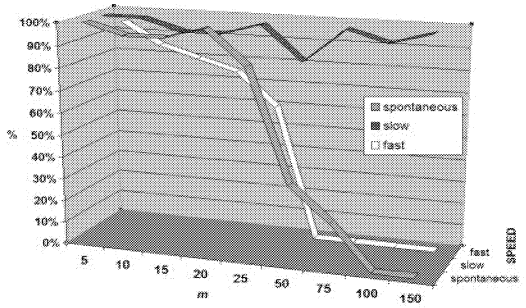


Figure 6. Successful Recognition Percentage as Function of the Recognition Window and of Speed of Movement

Table 2: Human- Based Test Results

*m**10ms	Slow	Spontaneous	Fast
50 ms	100%	100%	95%
100 ms	100%	95%	85%
150 ms	95%	95%	80%
200 ms	95%	100%	75%
250 ms	100%	85%	60%
500 ms	85%	35%	0%
750 ms	100%	20%	0%
1 s	95%	0%	0%
1,5 s	100%	0%	0%

It must be observed that the percentage of successful recognitions decreases rapidly as a function of the movement speed. This is due to the contribution of the time users remained in a given orientation (t_{settle}). In fact, this settling time was not forced (the user had no indication on how long she/he should stay in a certain position), but almost all users waited in the final position until feedback was provided by the system, showing some kind of adaptation to t_{settle} to achieve a successful recognition.

A final test was performed to assess general user satisfaction by letting users to freely navigate the virtual environment. The time window value was set to that on which each user has produced the lowest errors in the previous study: 50ms in the majority of cases. All the users expressed satisfaction and a couple of them also were positively impressed by the usability of the system and the fast learning curve. In some cases, a little latency was experienced. We believe that this was due to the Java3D based rendering. One of the users lamented the muscular effort required to use the system.

These experiments clearly suggest that improvements to the system can be obtained by a configurable recognition window. In fact, the recognition window must be established depending on both the characteristics of user gestures and the position of the sensing unit on his or her body. Some of the errors where clearly caused by the fact that some gestures where less natural than others for a particular user. For example, placing the sensing unit on the right hand causes the gesture of turning left to be more difficult than turning right, indicating the need to extend the recognition window for the left position.

Synthetic Experience. The second group of tests was performed by securing the sensing unit to a bar on a revolving support. The bar was moved by a stepper motor whose rotation speed and settling time (t_{settle}) could be pre-determinated. These experiments eliminated the noise introduced by the users and allowed us to assess the system's basic recognition capabilities.

We performed two sets of experiments. The first one was performed for a constant t_{settle} value. The second set was performed with a constant speed value.

Results of the first test, shown in Figure 7 for t_{settle} of 900ms, confirm those obtained in the user experience evaluation. It can be observed that the number of successful recognition is in general higher than human-based tests because of the improved synthetic movement precision.

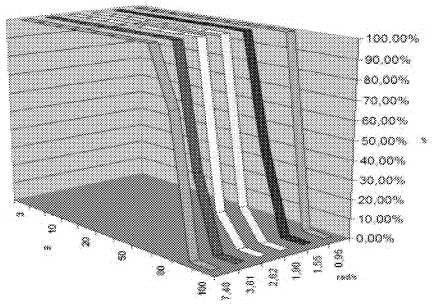


Figure 7. Successful Recognition Percentage as Function of Time Window (m) and Speed of Movement for t_{settle} =900ms

The second set of experiments shown in Figure 8, outlines the effect of t_{settle} on the recognition process for various time window sizes. Results suggest that small time windows improve the system robustness, especially for short t_{settle}. It must be noted that, compared to the human-based experiments, smaller t_{settle} values are permitted, due again to the higher movement precision of the stepper motor.

To summarize, by combining human-based and mechanical tests, we can conclude that low inertial recognition (i.e. small time windows) provides the best recognition performance. The lower bound for the time window should be determined by the level of noise introduced by lack of precision in the human movement. In any case, both user and synthetic experimental results demonstrate the effectiveness of the proposed gesture interface system.

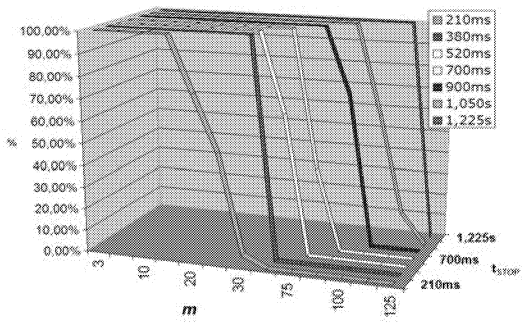


Figure 8. Successful Recognition Percentage (Function of t_{settle}) and Time Window (m) For Movement Speed 7,48 rad/sec

REFERENCES

Bang W., W. Chang, K. Kang, E. C. Potanin, A.D. Kim 2003 "Self-contained spatial input device for wearable computers", *in Proceedings of the Seventh International Symposium on Wearable Computers*, October 21 - 23, 2003, White Plains, NY, USA. IEEE 26- 34

Barbieri R., E. Farella, A. Acquaviva, L. Benini, B. Riccò, 2004 "A Low-Power Motion Capture System with Integrated Accelerometers," *in Proceedings of 2004 First Consumer Communications And Networking Conference* 5-8 Jan. 2004, Las Vegas, NV, USA. IEEE, 418 - 423

Cheok A. D., K. G. Kumar, S. Prince, 2002 "Micro-Accelerometer based Hardware Interfaces for Wearable Computer Mixed Reality Applications," *Proceedings of IEEE International Symposium on Wearable Computers*, 2002.

Headon R., R. Curwen, 2002 "Movement Awareness for Ubiquitous Game Control," *in Personal and Ubiquitous Computing*, 2002. Springer-Verlag London Ltd Vol. 6, Issue 5 - 6, (Dec) 2002, 407 - 415

Hernandez-Rebollar J. L., R. W. Lindeman, N. Kyriakopoulos, 2002 "A Multi-Class Pattern Recognition System for Practical Finger Spelling Translation," *Proceedings of International Conference on Multimodal Interfaces,* IEEE 2002, 185-190

Hinckley K., J. Pierce, M. Sinclair, E. Horvitz, 2000 "Sensing Techniques for Mobile Interaction," *Proceedings of User Interface Software and Technology*, ACM 2000, 139-148.

Lee J., I. Ha, 1999 "Sensor Fusion and Calibration for Motion Captures using Accelerometers", *Proceedings of International Conference on Robotics and Automation*, 1999. IEEE, (10-15 May) vol. 3, 1954 - 1959

Oakley I., J. Angesleva, S. Hughes, S. O'Modhrain, 2003"Tilt and Fill: Scrolling with Vibrotactile Display", *Proceedings of EuroHaptics* 2003.

Paiva A., R. Prada, R. Chaves, M. Vala, A. Bullock, G. Andersson, C. Hook, 2003 "Towards Tangibility in Gameplay: Building a Tangible Affective Interface for a Computer Game," *Proceedings of International Conference on Multimodal Interfaces,* 2003, IEEE 303-304

Pantic M., L. J. M. Rothkrantz, 2003 "Toward an Affect-Sensitive Multimodal Human-Computer Interaction," *Proceedings of* the IEEE 2003. vol. 91, no. 9, 1370-1390

Perng J.K., B.Fisher, S.Hollar, K.S.J. Pister, 1999 "Acceleration Sensing Glove (ASG)," *Proceedings of 3th International Symposium on Wearable Computers* 1999, pp. 178-180.

Regenbrecht H., G. Baratoff, I. Pouopyrev, M. Billinghust, "A Cable-less Interaction Device for AR and VR Environments," *Proceedings of 5th International Symposium on Wearable Computers* 2001.

Ribo M., P. Lang, H. Ganster, M. Brandner, C. Stock, A. Pintz, 2002 "Hybrid Tracking for Outdoor Augmented Reality Applications," *IEEE Computer Graphics and Application*s, Vol. 22, Issue 6, pp. 54-63, Nov.-Dec. 2002.

Tanrivedi V., R. Jacob, 2000 "Interacting with Eye Movements in Virtual Environments," The Future is Here, *CHI Letters*, 2000.

APTEC

A NETWORKED TELEPRESENCE SYSTEM FOR CARS AND ITS TRANSMISSION ISSUE

Kazutoshi Fujikawa[1] Kazumasa Yamazawa[1] Tomoya Ishikawa[1] Yuichi Butsuen[1] Tomokazu Sato[1]
Yutaka Nakaramura[2] Hideki Sunahara[1] Naokazu Yokoya[1]
[1] Graduate School of Information Science
Nara Institute of Science and Technology
Ikoma, Nara 630-0192, Japan
{fujikawa,suna}@itc.naist.jp, {yamazawa,tomoya-i,yuui-b,tomoka-s,yokoya}@is.naist.jp
[2] Information Science Center
Kyushu Institute of Technology
Kitakyushu, Fukuoka 804-8550, Japan
yutaka-n@isc.kyutech.ac.jp

KEYWORDS
Multicast, Wireless communication, Mobile communication, Video streaming

ABSTRACT

Recently, the telepresence which allows us to experience a remote site through a virtualized real world has been investigated. We have implemented a networked telepresence system using omni-directional images which enable multiple users to share a virtualized dynamic real scene in the distance and to see any direction of the scene they want. By adopting multicast transmission, our telepresence system will provide a scene for a lot of users without increasing network loads. In this paper, we introduce our networked telepresence system and discuss its network communication issue when using multicast as a transmission protocol.

INTRODUCTION

Recently, there are many researches of telepresence that acquires a dynamic real world into a virtual world and enables a user to be immersed in the remote environment (Moezzi, 1997). The telepresence can be applied to various fields such as entertainment, medical service, and education. We have already developed a telepresence system which uses an omni-directional camera and enables a user to look around the scene so that he/she increases the presence in telepresence (Ikeda et al. 2004; Onoe et al. 1998). Our telepresence system acquires and transfers the remote omni-directional scene by an omni-directional camera, and shows a user the view-dependent images in real time. The telepresence system also enables a user to look around the remote scene widely.

On the other hand, the progress of the wireless network technology, a car can have the Internet connection at all times. If our telepresence system applies to the car with the Internet connectivity using wireless communication channels, a variety of applications can be realized. For instance, a lot of remote users can experience having a drive and share the scene among them, if the omni-directional video stream is transmitted by multicast. Besides, each user can see any direction of the scene he/she wants. Generally, for the telepresence system, the video stream should be transmitted constantly. Therefore, a robust transmission mechanism is required, since a car always moves and the condition of the wireless communication channel might be unstable.

In this paper, we introduce our networked telepresence system, which multiple users on the Internet to look around a remote scene captured by omni-directional cameras, and discuss its network issue.

Section describes the overview of the proposed system. Section discusses the transmission issue and Section describes the related work. Finally, this paper concludes in Section .

OVERVIEW OF OUR TELEPRESENCE SYSTEM

Figure 1 briefly shows the architecure of our telepresence system. Our telepresence system consists of a car which mounts omni-directional camera "HyperOmni Vision (Yamazawa et al. 1993)," multicast relay server for video stream on the car, and the omni-directional video viewer. The omni-directional video stream is transmitted to the campus LAN from the car via a wireless communication channel of IEEE802.11a or IEEE802.11g. A user can receive the omni-directional video contents using a desktop PC, a laptop PC, or even an immersive client with a head mounted display (HMD).

Hardware components on the car

Figure 2 shows the hardware configuration on the car of our telepresence system. In the car, there are three PCs. One PC records GPS data including the positioning data and the time. Another PC is for the omni-directional camera. The omni-directional camera acquires omni-directional progressive video surrounding the car (see Figure 3). The acquired omni-directional video is transferred to the PC through the FireWire channel. The PC encodes the omni-directional video (640x480 pixels, 30 fps) to Windows Media Format

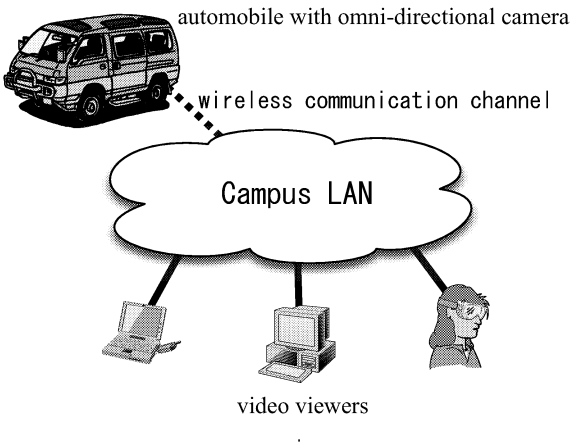


Figure 1: Overview of the proposed telepresence system

(1Mbps) by Windows Media Encoder. The other PC is for the wireless communication. This PC has both IEEE802.11a interface and IEEE802.11g interface. The GPS data and the omni-directional progressive video are transferred to the campus LAN through this PC. Table 1 describes the detail specification of the hardware equipped on the car.

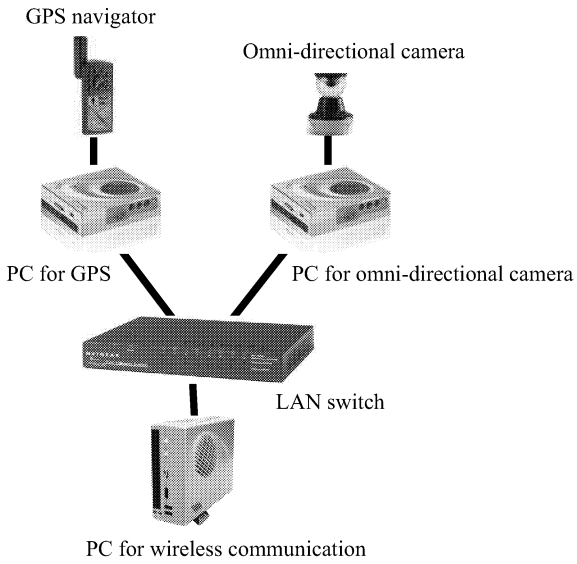


Figure 2: Hardware comfiguration on the car of our telepresence system

Omni-directional video viewer

The omni-directional video viewer converts the omni-directional video (left image of Figure 4) to the common perspective video (right image of Figure 4) to show the user the common perspective video. Our telepresence system uses a texture mapping function of GPU (graphic processor unit) for converting video images in real-time by the method of the literature (Onoe et al. 1998).

A user on a desktop PC or a laptop PC can watch the



Figure 3: Omni-directional camera acquiring the scenery surrounding the car

Table 1: Hardware configuration of our telepresence system

Omni-directional camera	SONY DCR-TRV900 + Hyperboloidal mirror
GPS navigator	GARMIN GPS V
wireless network interface	IEEE802.11a, IEEE802.11g
PC for omni-directional camera	CPU: Intel Pentium4 2.53GHz OS: Windows XP
PC for GPS	CPU: Intel Celeron 1.0GHz OS: FreeBSD
PC for wireless communication	CPU: Intel Pentium4 1.7GHz OS: FreeBSD
Multicast relay server	CPU: Pentium4 2.5GHz OS: Windows 2003 Server

omni-directional video contents through the viewer invoked by a web browser. We have been implementing an omni-directional video viewer using Active-X as a plug-in software of a web browser (see Figure 5).

When the user access the web page which provides an omni-directional video content, the omni-directional video viewer implemented by Active-X is installed automatically. The user can easily see an omni-directional video content without care of an omni-directional camera type, parameters of camera, file path of the content, and so on, because a content provider embeds them in an HTML file. The user can look around the omni-directional video by using a mouse-drag operation.

Omni-directional video contents

There are two kinds of omni-directional video contents in the present implementation: stored video contents encoded in advance and live video contents encoded in real-time. Note that stored video contents can be provided as an on-demand-service.

The live video contents are used for simultaneously pro-

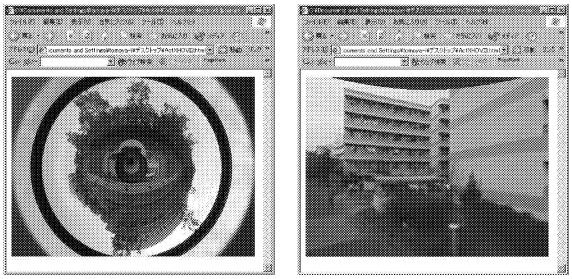


Figure 4: Omni-directional image and common perspective image

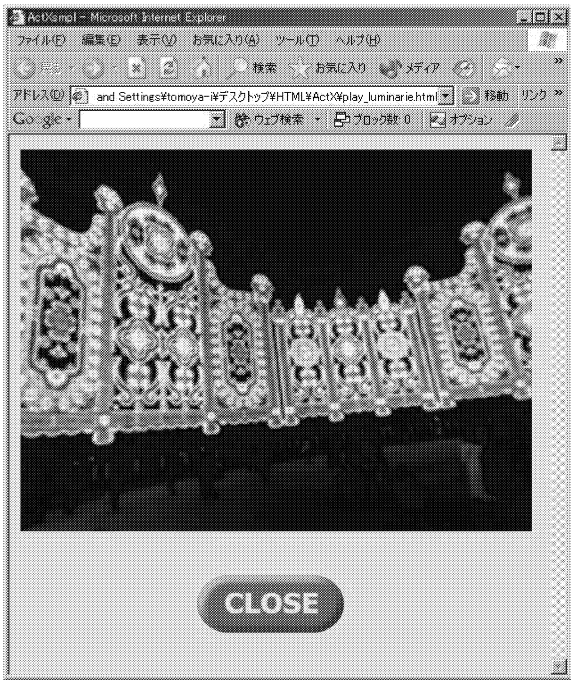


Figure 5: Web based omni-directional video viewer

viding multiple users with the same contents similar to TV broadcasting. The user can see the live video contents acquired by the omni-directional camera. The live video contents can be transferred to multiple users by multicast without increasing the network load.

TRANSMISSION ISSUE

Experiment on campus network

In our telepresence system, the communication channel between the car and the campus LAN uses wireless communication facility of IEEE802.11a and IEEEE802.11b/g.

We have implemented the prototype system of the proposed telepresence system to clarify the problem of the multicast transmission using wireless communication channels. Figure 6 shows the experimental environment.

Now, we discuss the behavior of the network communication of the prototype system. In our experiment, we used "Foundry FirstIron WorkGroup" which has 24 ports of 10/100 Mbps as the LAN switch and "Avaya Access Point III" as the wireless access point. Wireless access point (1) and (2) connected to port 2 and 1 of the LAN switch, respectively. A desktop PC for the omni-directional viewer connected to port 7 of the LAN switch, and we used three laptop PCs which have IEEE802.11a interface and IEEE802.11g interface.

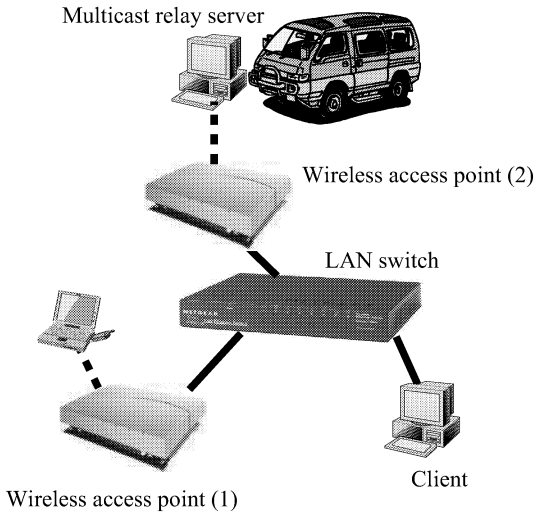


Figure 6: Experimental environment

In this experiment, each PC repeatedly did join/leave operation to the multicast group for the omni-directional video stream. Each laptop PC moved between two wireless access points (1) and (2). Moreover, in the second experiment, the car moved around these two wireless access points, that is, the multicast relay server move between both wireless access points. In the experiments, we used four PCs as the omni-directional video receivers. Through the omni-directional viewer, each user on the PC could look around the scene in any arbitrary direction which he/she want to see.

We measured the throughput of the mulicast traffic in the following cases under the network condition that the wireless communication channel between the car and the campus LAN was stable using IEEE802.11b/g, and the car did not move. If a PC is connected to a different wireless access point from the one that the car connected, the throughput is about 1Mbps. That is, the video stream is displayed on the omni-directional viewer at 30 frames/sec, since the bit rate of Windows Media Format (WMT9) is 1Mbps. If a PC is connected to the same wireless access point that the car connected, the throughput is about 800kbps. If several PCs are connected to a different wireless access point from the one that the car connected, the throughput is about 600kbps. Due to the multicast transmission the network load did not increase, although the number of received omni-directional video viewer increased.

We also did another experiment under the condition that the car moved around in the campus. That is, the access point that the car would be connected was changed according to

the position of the car. In the experiment, we measured the packet loss rate using the video of WMT9 whose bit rate is 700kbps. We found the results depend on the packet size. When we used the default packet size of WMT9 which is 10600 bytes, the packet loss rate was 25-35%. The such high packet loss rate is caused by the fragmentation of packet. When we adopted 1432 bytes as the packet size that is less than the fragmentation size (1516 bytes), the rate was about 2%.

In the above experiments, the time delay between the acquisition and the presentation omni-directional video took about 10 seconds. It is due to the video encoding and the buffering between the acquisition process and the sending process.

In the above experiments, the IP address of the multicast relay server did not change even though the car moved around in the campus, because the wireless network in the campus consists of a single LAN segment. In this case, the multicast tree management is not difficult for network elements such as switches or routers. In general, if the multicast sender moves and connects to the different wireless access point, the sender's IP address of the multicast group changes. In this case, network elements have to reconfigure the multicast tree. However, most multicast-enabled network elements do not suppose the such situation. Additionally, if no PCs joined to the multicast group, the multicast traffic sent to the all ports of the LAN switch which caused the multicast storm. And basically, the condition of wireless communication channel is not always stable.

Therefore, some robust multicast transport mechanism is required, which can have redundant paths in a single multicast tree. By such multicast transport mechanism, we can use multiple wireless communication channels simultaneously.

New multicast protocol

To cope with the above problem, we propose a new multicast transport protocol, which consists of Protocol Independent Multicast-Sparse Mode/MultiPath support extension (PIM-SM/MP) and Protocol Independent Multicast-Location Independent support extension (PIM-LI). Figure 7 shows the architecture of the proposed multicast protocol. The proposed multicast protocol is based on Protocol Independent Multicast-Sparse Mode (PIM-SM) (Fenner et al. 2004).

PIM-SM/MP can discover the candidate router which can establish a redundant path in the creation process of the multicast distribution tree, and then choose a router which actually establishes a redundant path from the candidates (see Figure 8). PIM-LI can identify the multicast sender which moves among the different network segments, absorb the change of sender's IP address, and repress the reconstruction of the multicast distribution tree. Repressing the reconstruction can be archived by establishing a redundant path according to the new location of the multicast sender. In this case, PIM-LI uses the information about the candidate routers which can establish a redundant path from PIM-SM/MP. PIM-LI selects an appropriate redundant path, and requests PIM-SM/MP to establish the path.

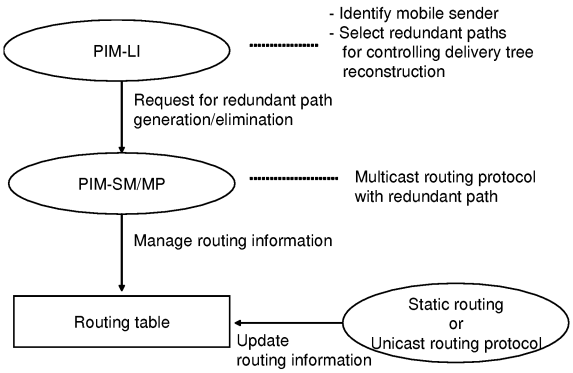


Figure 7: Architecture of our proposed multicast protocol

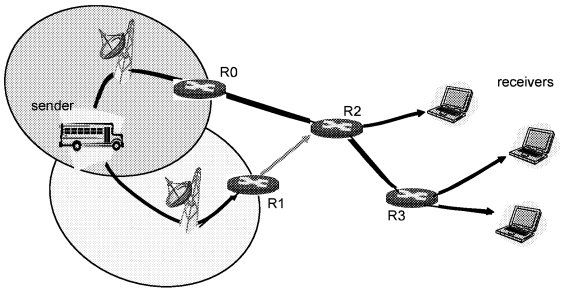


Figure 8: Overview of PIM-SM/MP

RELATED WORK

In most telepresence systems (Krishnan and Ahuja, 1996; Moezzi, 1997; Szeliski, 1996; Zheng and Tsuji, 1990), active cameras such as panning and tilting cameras are used to present a user to view-dependent images of a scene in the distance. This often suffers from the time delay from the change of viewing direction to the change of displayed image. The time delay is mainly caused by both the communication between an observation site and an observed world and the control of cameras, as well as the time to follow the user's viewing direction. Especially, the former factor depends on the actual distance between the observation and observed sites.

Moreover, in the existing telepresence systems, if a user want to change the view direction of a scene, he/she have to control a camera. Therefore, it is difficult that multiple users share the same telepresence system.

In the IPv6 technology, some multicast protocols which consider the mobility of the node have been proposed (Johnson et al. 2004; Suh et al. 2004). These multicast protocols can handle the mobility of only the multicast receivers. Source Mobility Support Multicast (SMM) (Sato et al. 2004) proposed a multicast protocol considering the mobility of the sender. Since this multicast protocol supposes to be used by the Intelligent Transport System (ITS), it uses the Cellular IP (Valkó, 1999) as the underlying network protocol between the car and the access point. This multicast protocol can not be applied to the other underlying network protocol except the Cellular IP. Therefore, the more general multicast proto-

col considering the mobility of the sender is required.

CONCLUSIONS

We have developed a networked telepresence system which easily enables multiple users to look around a remote scene with the omni-directional camera. The system uses a web-browser, and enables users to see omni-directional video such as common video. In the experiment, the omni-directional video was distributed through wireless and wired network by multicast protocol, and multiple users could look around the scene in arbitrary directions in real time. The delay between the acquisition and the presentation of omni-directional video is 10 seconds. In future work, we should reduce the delay in transmitting an omni-directional video stream. Also, we have to provide some robust transport mechanism which can handle multiple different wireless communication channels simultaneously. Moreover, we have designed and developed a new multicast protocol which has redundant multicast path considering the mobility of the sender.

REFERENCES

FENNER, B., M. HANDLEY, AND H. HOLBROOK (2004): *Protocol Independent Multicast-Sparse Mode*RFC 3775.

IKEDA, S., T. SATO, M. KANBARA, AND N. YOKOYA (2004): "Immersive telepresence system using high-resolution omnidirectional movies and a locomotion Interface," in *Proceedings of SPIE Electronic Imaging*, vol. 5291.

JOHNSON, D., C. PERKINS, AND J. ARKKO (2004): *Mobility Support in IPv6*draft-ietf-pim-sm-v2-new-11.txt.

KRISHNAN, A., AND N. AHUJA (1996): "Panoramic image acquisition," in *Proceedings of 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 379–384.

MOEZZI, S. (1997): "Special issue on immersive telepresence," *IEEE MultiMedia*, 4(1), 17–56.

ONOE, Y., K. YAMAZAWA, H. TAKEMURA, AND N. YOKOYA (1998): "Telepresence by real-time view-dependent image generation from omnidirectional video streams," *Computer Vision and Image Understanding*, 71(2), 154–165.

SATO, K., M. KATSUMOTO, AND T. MIKI (2004): "Source Mobility Support Multicast (SMM)," *IPSJ Journal*, 45(2), 412–425.

SUH, K., D.-H. KWON, Y.-J. SUH, AND Y. PARK (2004): *Fast Multicast Protocol for Mobile IPv6*draft-suh-mipshop-fmcast-mip6-00.txt.

SZELISKI, R. (1996): "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, 16(2), 22–30.

VALKÓ, A. G. (1999): "Cellular IP: a new approach to Internet host mobility," *ACM Computer Communication Review*, 29(1), 50–65.

YAMAZAWA, K., Y. YAGI, AND M. YACHIDA (1993): "Omnidirectional imaging with hyperboloidal projection," in *Proceedings of International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1029–1034.

ZHENG, J. Y., AND S. TSUJI (1990): "Panoramic representation of scenes for route understanding," in *Proceedings of 10th International Conference on Pattern Recognition*, pp. 161–167.

SUPERASSIST: PERSONAL ASSISTANTS FOR COOPERATIVE HEALTHCARE TREATMENT

Geert de Haan
Charles A.P.G. van der Mast
Man-Machine Interaction Group
Delft University of Technology,
Mekelweg 4, 2628 CD,
Delft, the Netherlands
E-mail: {geertdehaan|c.a.p.g.vandermast}@ewi.tudelft.nl

Olivier Blanson Henkemans
Mark A. Neerincx
TNO Human Factors
P.O. Box 23, 3769 ZG,
Soesterberg, the Netherlands
E-mail: {blanson|neerincx}@tm.tno.nl

KEYWORDS

Human-computer interaction, tele-medicine, supervision, personalization, ambient intelligence, health care, personal assistants, cooperative problem-solving.

ABSTRACT

This paper describes the research plans in the SuperAssist project, introducing personal assistants in the care of diabetes patients, assisting the patients themselves, the medical specialists looking after the patients' healthcare, and the technical specialists responsible for maintaining the health of the devices involved. The paper discusses the issues of trust and cooperation and the proposed scenario-based approach to design a future of agent-based healthcare to increase the self-help abilities of individual patients.

INTRODUCTION

TNO Human Factors (TNO), Delft University of Technology - Man-Machine Interaction group (TUDelft) and Leiden University Medical Centre - Medical Informatics (LUMC) are developing guidelines, models and methods for joint user-"electronic assistant" supervision of critical equipment and information.
The aim is to establish effective and efficient distributed supervision of networked information compilations and technical equipment, which is trustworthy for the user and takes place in a socially approved manner. Specific innovative project results are: communication and interaction model for these assistants; methods for joint human-computer supervision; improved test methods, tools and criteria for systematic assessment of user experience; "best practice" implementation-method and guidelines; and a "proof of concept" in the transmural health care domain (e.g. diabetes).

For the medical application domain, the SuperAssist framework aims to reduce the costs by improving the local, self-care capacity of people by efficient employment of remote, distributed expertise. Figure 1 provides a possible setting for the SuperAssist framework in which both, a patient, the medical specialists who treat the patient, and a technical specialist who looks after the proper functioning of the medical equipment involved are all provided with personal assistants to facilitate their own tasks as well as to support cooperative tasks which might occur, for instance, with the introduction of new drugs and therapies or with the inevitable malfunctioning or erroneous use of the medical equipment.

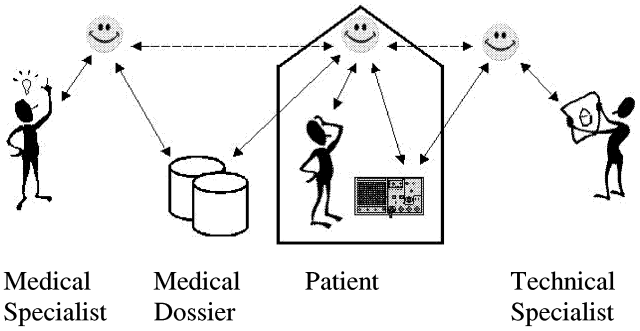


Figure 1: Users, Personal Assistants and Information Flow

The project's business partners - Science & Technology, Philips Research, Pemstar and Sigmax PDA Solutions - bring in their technology and contribute to the development and validation of SuperAssist elements. The joint activities are included in cognitive engineering cycles, in which the foundation, specification and demos of the SuperAssist concept are being refined and extended. The SuperAssist research takes place in the medical domain, but it aims at a generic solution for the distributed supervision of complex environments.

A USER SCENARIO

Ms. Brown is a vital 72 year old who, after loosing her husband several years ago, refound her balance and takes a keen interest in bicycling and reading. Two years ago she was diagnosed with type II diabetes for which a diet and medication were prescribed. Although she is sincere in following up the medical directions, sometimes her memory fails her and she simply forgets things like when she has last taken her pills or that she should only buy low glucose products
Since ms. Brown is well overweight, she has been asked to participate in a therapeutic program whose purpose is to attempt to let participants lose some of that weight in a responsible yet playful manner. As a result, every day before breakfast and before bedtime Ms. Brown consults her personal diabetes assistant, provided to her as part of the program, to check her blood glucose levels and to help her remind whether she has taken her pills or not. To

address her deteriorating eyesight, the assistant is equipped with an extra-large extra-bright screen, a few large and clearly labelled buttons. In addition, the form and content of the dialogue have been adapted for ms. Brown according to her preferences and to the way she uses the diabetes assistant.

On three consecutive days, ms. Brown's blood glucose level has been slightly higher than normal and today it is rather high. An abnormal reading is not necessarily a cause for alarm. One possibility is that there is nothing wrong with ms. Brown but rather that the technique let her down. The reading of the blood glucose level is derived from the light passing through the finger that ms. Brown offered for testing, which is much more comfortable but also less reliable then blood sampling. Ms. Brown may call the district-nurse to do blood tests but she prefers not to, because he is busy enough with the "really sick people". Finally, the measurement might be genuine. It may be that ms. Brown has forgotten to take her night time pills. However, the display of the assistant clearly shows that she has confirmed taking her pills until this morning and besides, the compartments of her pill-box are empty. It may be -again- that she has accidentally taken a regular instead of a low sugar desert.

A little alarmed, ms. Brown at first doesn't know what to do but then she sees the help-button on the diabetes assistant and remembers from the instruction session that "if you would like to ask a question, first try the help-button". This she does and a friendly voice assures her that there is only a little worry but not an acute problem. Furthermore, the assistant suggests her to redo the measurement, this time using the little finger of her other hand. Using her other little finger, ms. Brown now learns that her blood glucose level is only slightly higher than normal and her assistant asks her to take her pills, including an extra TZD "you know, the big blue one" just to be on the safe side.

Behind the Scenario

By this time, also behind the scenes some activity has taken place. Because of the measurement error, a maintenance module inside the assistant increased the problem count of the measurement module, reset it and performed a self evaluation test in order to exclude a range of technical problems. Also a temporary wireless network is set-up for a number of routine and problem reports. When the assistants' problem count exceeds a certain level of serious measurement or transmission problems, a technical expert is informed for further maintenance or for a replacement of the device.
In addition, when serious problems occur, the district nurse will be informed of the existence of a possibly faulty device, so that he can check out to take over or provide a replacement diabetes assistant. In this case, the measurements remained within the safety boundaries, there was no need to alarm the specialist diabetes team, although, as a matter of routine, both blood glucose readings and the measurement problem are fed into ms. Browns medical dossier.
Finally, because there has been a change in ms. Brown's stable blood glucose level, a reminder is send to the district nurse. The personal patient-visiting assistant agent will remind the nurse the next time he goes on a periodical visit to ms. Brown. Even though the diabetes assistant is "allowed" within certain boundaries to change a patient's medication, only medical specialists (or in their place the district nurses) are allowed to make any enduring or substantial changes.

Diabetes Type II

Diabetes type II or "old-age diabetes", occurs when the body doesn't use the available insulin effectively. Because of the low insulin level, the blood glucose level may build up, causing a sick feeling. Type II diabetes is generally caused by a life-long combination of a bad diet, lack of exercise and overweight. The treatment of diabetes II is focused on keeping a healthy blood-glucose level which should neither be too high not too low. If the glucose level drops too low, there is the possibility of passing out, coma and eventually death. A minority of diabetes II patients has to inject themselves with insulin (about 15%) but for most patients the treatment consists of a diet, in combination with medicines to boost the production and the effectiveness of insulin.
Diabetes II is not deadly by itself but its symptoms, in the form of a deterioration of small blood vessels may lead to blindness, numbness, inflammations, etc. which may eventually cause death, although, in principle, diabetes II is not difficult to treat

CRITICAL ISSUES IN COOPERATIVE HEALTHCARE

In the SuperAssist project we intent to follow a scenario-based approach (Carroll 1995) in that the project consists of a number of phases, each of which is approached iteratively, and guided by hypothetical and real scenarios of would-be situations (Rizzo and Bacigalupo 2004). In dealing with the novel situation of having multiple types of users communicate and cooperate with multiple types of agents, the domain is relatively new and, according to Erickson (1995), a scenario-based approach is the preferred method. In addition to such standard HCI practices, it is a "sine qua non" that users should participate in the development of a personal service or consumer product that SuperAssist aims at (Kemp and van Gelderen 1996).

Even though the project targets at a novel type of application in a new area of multi-user multi-agent (MUMA) interaction, it does not start from scratch. Apart from being concerned with software agents which assist human beings in the area of ubiquitous computing (Maes 1998; Weiser 1991), SuperAssist attempts to build further upon the PALS project, a Dutch project investigating the requirements to meet changing user needs and usage contexts with respect to web-services (see: PALS; Herder 2003; Neerincx and Streefkerk 2003).

From this and other projects, two factors turned out to be of critical importance for the success of human-agent interaction: trust and cooperation. Trust in the sense that people need a basic level of trust in a system or personal assistant to "do business" with them, and cooperation in the

sense that, like in human-human cooperation, participants rely on images, models or ideas about the communication and cooperation abilities of their discussion partners in order to set appropriate expectations.

Trust

The trust of a patient, a medical specialist or a nurse in their personal assistant is a prerequisite to delegate part of their responsibilities to the electronic device. To enable trust and delegation between people and electronic assistants, it is a first necessity that the form and content of the interaction dialogue are acceptable to the human user, including a possibly frightened patient or an excessively busy medical worker.
Here, it may be possible to adapt the dialogue to the current context of use including aspects like fear and stress (Neerincx and Streefkerk 2003).
Apart from a humanly usable dialogue for solving problems, it is necessary that there is a shared view on the problem domain. This may be achieved by creating a representation of the problem space in which rules exist or may be derived to guide the behaviour of the assistants. Note that, even in a well-delimited problem domain as diabetes, this is a major undertaking due to influences from outside the system, such as the nurses' other responsibilities or the complex (side) effects between things like medication, stress and diet.

Cooperation

Cooperation for problem solving between people and their assistants and the cooperation between assistants among themselves requires some kind of model which at least describes 'what to expect from whom' in terms of questions, actions, etc. In (software) multi-agent architectures like e.g. FIPA (see: FIPA) the question 'what to expect from whom' is specified in agent-communication models, either by way of static models, specified beforehand, or in on-request models, used in dynamically changing environments.

Also at a much higher level, the concept of cooperation is important. For example, in the discussions about the introduction of an electronic medical passport and in several projects in which the cooperation between medical staff and patients goes beyond exchanging complaints, pills and advise. In the Telemedicine-Diabetes project at the University of Oxford (Tarassenko et al. 2004) for example, diabetes patients do not merely monitor their blood glucose level to treat themselves, but by additionally phoning-in the measurement results to a diabetes treatment centre, these can be stored in a database and used for personal advice.

From these and related investigations, a general "lesson learned" is that patients are generally able and willing to take participate in such mutual medical specialist-patient cooperation project, but only if the personal benefits to the patient are clear, and when advice and treatment are personalised to the patients context and circumstances. As a result, a main research idea in the SuperAssist project is to use a PDA-like device to present diabetes patients with an existing (work)book "the diabetes manual" (Stewart 2004), to be translated in Dutch) in a personalised, adaptable and context-sensitive way to help patients cope with their disease and the often associated overweight.

SCENARIOS, PROTOTYPES AND DESIGN

Within SuperAssist, scenario-based design (Carroll 1995) will be the main technique to develop research ideas in the project. Scenario-based design will be used in combination with prototyping as the projects main technique to test the applicability and the usefulness of the ideas in more-or-less tangible form. In combination, scenario's and prototypes and mock-ups form cognitive engineering cycles in which they are connected by different levels of specification. Since the project is in an early stage, the main focus is currently on the use of scenarios for two different purposes.
First, scenarios are used to collect 'wild ideas' to envision the outcomes of the SuperAssist project in the form of integrated real-world situations. Scenario-envisioning is thus used to provide some -however idiosyncratic and erroneous- idea of what the project is going to deliver eventually.
In addition, by doing so, the scenarios also provide a means to create a common as well as shared understanding among the project members. The first scenario that was discussed in this paper and figure 1 present the use of scenario's for envisioning purposes.

Cognitive Engineering and Scenario-based Design

Research and design in the SuperAssist project will make use of cognitive engineering cycles, an engineering approach to HCI design with a focus on the human abilities and limitations as the most critical of all design aspects. In relation to other HCI design approaches this means that design is user-centred, involves user participation and takes the need for design iterations for granted. This is not all too different from the common HCI approach to design, except perhaps that an engineering stance is taken in that both, a particular problem has to be solved and that the solution should be systematically applicable to a whole class of similar problems.
The cognitive engineering framework in one in which practical theories, communication and interaction models, and prototypes are being developed in a structured iterative process (Lindenberg, Nagata and Neerincx 2003). A practical theory includes accepted features of cognition and affection, distinguishes different task contexts within a specific application domain, provides descriptions of computer behavior, and predicts the interaction effects between these features, contexts and computer behavior. An interaction model formalizes and quantifies these interaction effects and is used to design a prototype of an “intelligent” user interface, which adapts its behavior to the present goals state, and to the cognitive and affective processes of the user. In addition to the individual user-assistant cooperation model, proposed in Lindenberg et al., the SuperAssist project will develop a co-operation model for the overall, multi-actor supervision, which formalizes the information exchange between all of the actors involved, both agents and men.

The research in SuperAssist is further organized around three main experiments or rather three series of experiments

to investigate the main research questions in the project: task allocation among agents and men, the organization of joint anomaly management or problem solving, and finally, the “look & feel” of the assistants, respectively. In addition, a number of flanking experiments and scenario exercises is planned, intended to delimit the design space for the main experiments by excluding in an early stage the less- or irrelevant factors and to find solutions for design details for which a full-size experimental setup is overkill.

Apart from empirical validation of models and theories, the three experiments also serve to validate and improve the design of the prototypes which are designed to run these experiments. The prototypes are used as vehicles to develop both the required knowledge about human-agent cooperation as well as drive the actual design of the specific solution to the problem at hand. As such, in the course of the project, the design of the prototypes being tested changes and becomes more refined and detailed. Whereas the initial experiments will take place using mock-ups, at the end of the project the aim is to present and test a demonstrator in the form of a hafiz prototype. It should be noted, however, that the primary aim of SuperAssist is to acquire and validate knowledge to design MUMA applications. As such, on the one hand, the creation of mock-ups, demonstrators and prototypes is a vehicle to facilitate research; not a 'deliverable' as in many EC research projects. On the other hand, the prototypes play a dual role, to validate the underlying theoretical knowledge as well as that they themselves are tested as solutions to the specific design problem.

Scenarios in SuperAssist

How scenario's are used in each stage of design and investigating is exemplified by Rizzo and Bacigalupo (2004) who describe a full range of using scenario's from early design stages, like familiarization and discovery of the main issues to later stages involving validating of ideas and design proposals with stakeholders, creating specifications for design and implementation, and validation of design solutions. According to Rizzo, particularly in new design contexts, it is important not to get lost in the false security of formal specifications. Instead they propose to use scenarios throughout design for envisioning, specification and validation.

In the early stages, the design process starts with observing the relevant aspects of the 'old' design situation, laid down in a number of scenarios or storyboards. Such descriptive scenario's are subsequently used to discuss and develop a common view with stakeholders what the essential problems and principled solutions may be. Note that this stage, like any of the following may require a number of iterations to arrive at a suitable set of working scenarios. In the project, interviews and observation sessions are currently held with both support engineers, patients and medical personal.
Next, the chosen scenario's are further analyzed to derive a number of so-called micro-scenarios in which for each part of a scenario, a number of essential elements is identified. This step is most similar to classical task-analysis but instead of aiming at fully describing the stakeholders' tasks, the aim is to arrive at a level of description which allows one to develop and feed-back or validate partial design solutions. The analysis step mostly involves desk work, but to evaluate the feasibility of each of the partial design proposals required cooperation with the stakeholders who actively have to decide which solutions will in practice make sense and which do not. It is at this step that the experiment about task-allocation will be planned and executed.
After analysis follows a synthetic step in which a number of design solutions is taken together as proposals to solve the underlying design problem. The results of this step are called macro-scenarios which describe a number newly proposed task-scenarios in combination together with the supporting technology. Describing design proposals in scenario form rather then in some formal way, makes it easier for stakeholders to be involved. Also here, stakeholders involvement is required, not as individual stakeholders but rather as a community of stakeholders. In such planned feedback sessions, including the planned experiment on joint human-agent problem solving, the role of stakeholders is both to evaluate and select (some) overall design solutions as well as to propose improvements to the selected proposals.
Finally, after the subsequent creation and implementation of one or more of the proposed design solutions, one or more concluding evaluation sessions are held, guided high-level scenario's from the early stages, to validate the overall design solutions and, where necessary, to improve them at a more detailed level. This step results in some form of a final design solution which can be evaluated in the last experiment, concerning the "look & feel" of the design.

It may be clear from the preceding description that scenario's seem particularly useful vehicles to enable user- or stakeholder participation during all the design steps. In addition, scenario-based design approaches also seem particularly useful in situations in which design does not merely involve the creation of a computer application but rather the design of solutions which involve both men and machines -or rather- men, agents and machines with each specific roles and responsibilities to fulfil. What is not yet clear at this stage of the project is how exactly to fit in the planned experiments into a scenario-based approach and to establish a 'best practice' for experiment-backed scenario-based design.

RESULTS AND OUTLOOK

The SuperAssist project attempts to set up an integrated healthcare service in the area of diabetes treatment, assisted by electronic devices and software agents. From this research we expect to be able to derive models, architectures, guidelines and general design knowledge to introduce human-agent systems that enable people to help themselves to a greater extend than they are able to do, now. As such, the SuperAssist case is intended as a test case for a range or healthcare and other services and not just as yet another smart solution for a particular problem.
Only by introducing, testing and validating general tools, we will be able to help keep future healthcare costs within the boundaries of affordability and manageability - at the

very least for individual patients like ms. Brown and ms. Blair.

The SuperAssist project also employs diabetes as a vehicle to learn general lessons about how to introduce trustworthy and cooperative electronic or software assistants into the everyday life of real people. Until recently, human-agent interaction has been studied in settings where a single agent assisted a single user, as in supporting searching, information filtering, and critiquing systems. In the SuperAssist project, research and development are extended to MUMA or multi-user multi-agent communication and cooperative problem solving. This novel research area does not only introduce new types of questions and new types of problems but it is also a starting point for the integration of separate intelligent applications into intelligent overall services.

REFERENCES

Carroll, J.M. 1995. *Scenario-based Design: Envisioning Work and Technology in System Development*. John Wiley, New York.

Erickson, T. 1995. "Notes on Design Practice: Stories and Prototypes as Catalysts for Communication". In *Scenario-Based Design: Envisioning Work and Technology in System development*, 1995, J.M. Carroll (Ed.). John Wiley, New York, 37-58.

FIPA. The Foundation for Intelligent Physical Agents. See: www.fipa.org

Herder, E. 2003. "Utility-Based Evaluation of Adaptive Systems". In *Proceedings of the 2nd Workshop on Empirical Evaluation of Adaptive Systems* (Johnstown, PA), 25-30.

Kemp, J.A.M. and van Gelderen, T. 1996. "Co-discovery Exploration: an informal method for the iterative design of consumer products". In *Usability Evaluation in Industry*, 1996, P.W. Jordan, B. Thomas, B.A. Weerdmeester and I.L. McClelland (Eds.). Taylor and Francis, 139-146.

Lindenberg, J., Nagata, S.F. and Neerincx, M.A. 2003. "Personal Assistant for onLine Services: Addressing human factors". In *Human-Centred Computing: Cognitive, Social and Ergonomic Aspect*, 2003, D. Harris, V. Duffy, M. Smith & C. Stephanides (Eds.). Erlbaum, London, UK, 497-501.

Maes, P. 1998. "Reflections on ... Agents that Reduce Work and Information Overload". In *Readings in Intelligent User Interfaces*, 1998, M.T. Maybury and W. Wahlster (Eds.). Morgan Kaufmann, 525-536.

Neerincx, M.A. and Streefkerk, J.W. 2003. "Interacting in Desktop and Mobile Context: Emotion, Trust and Task Performance". In *Proceedings of the first European Symposium on Ambient Intelligence - EUSAI* (Eindhoven, The Netherlands). Springer-Verlag.

PALS. Personal Assistant for onLine Services. IOP/MMI project website; see http://www.tm.tno.nl/pals

Rizzo, A and Bacigalupo, M. 2004. *Report on testing of full scale scenarios*. Deliverable 3.2 of CARE/CREA Project. Available from: www.dblue.it/pdf/crea/CARE as CREA_UNISI_WP3_D3.2_v1.1.pdf

Stewart, R. 2004 (Ed.). *The Diabetes Manual*. Warwick Diabetes Care, University of Warwick, UK.

Tarassenko, L., Gibson, O. J., Hayton, P. M., Cobern, W. R., Farmer, A. J., Hannaby, K., Dudley, C. and Neil, A. 2004. "Mobile Phone Technology to Support the Self-management of Diabetes". In *Proceedings of the Diabetes UK Annual Professional Conference* (Birmingham, UK, March 2004), 17-19.

Weiser, M. 1991. "The Computer for the 21st Century". *Scientific American* 265, No.3, 94-104.

BIBLIOGRAPHY

GEERT DE HAAN studied Cognitive Psychology at the University of Leiden and has a PHD in Mathematics and Computer Science from the Vrije Universiteit Amsterdam. After a number of projects in which he worked on formal design models, usability evaluation, scenario-based design in Digital Libraries and Ubiquitous Computing, he is currently employed at the man-Machine Interaction Group at the Department of Electronic Engineering, Mathematics and Computer Science at Delft University of Technology, as a Postdoc on the SuperAssist project. His main interest is in design methods for, among others, personalization and ubiquitous computing.

CHARLES VAN DER MAST has a PHD from Mathematics and Computer Science from Delft University of Technology where he is employed at the Man-Machine Interaction group at the Department of Electronic Engineering, Mathematics and Computer Science as a Associate Professor. He teaches Multimodal Interfaces and Virtual Reality, Developing Highly Interactive Systems, Multimedia, and Usability Engineering, and developed the Bachelor/Master curriculum in Media & Knowledge Engineering. His interests include using various media to improve teaching, VR therapy for phobia treatment.

OLIVIER BLANSON HENKEMANS studied Business Communication and Media at Tilburg University and worked on website development and Augmented Multiparty Interaction. He is currently pursuing a PHD at both TNO Human Factors and the Man-Machine Interaction group at the Department of Electronic Engineering, Mathematics and Computer Science at Delft University of Technology.

MARK NEERINCX studied Cognitive Psychology at the University of Leiden and has a PHD in Cognitive Psychology from the University of Groningen. He worked at the TNO Institute of Preventive Health Care, the University of Leiden and the University of Amsterdam, after which he became research scientist at TNO Human Factors and heads the Usability Engineering group. He is also parttime professor in the Man-Machine Interaction group at the Department of Electronic Engineering, Mathematics and Computer Science at Delft University of Technology. His research interests include usability engineering, human-computer interaction, cognitive task analysis, mental load, ambient intelligence, personalization, electronic partners, adaptive interfaces and multi-modality.

THE EFFECT OF LOCOMOTION TECHNIQUE ON PRESENCE, FEAR AND USABILITY IN A VIRTUAL ENVIRONMENT

M.J. Schuemie
B. Abel
C.A.P.G. van der Mast
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands
Email: c.a.p.g.vandermast@ewi.tudelft.nl

M.Krijn
P.M.G. Emmelkamp
University of Amsterdam
Department Clinical Psychology
Roetersstraat 15, 1018 WB Amsterdam
The Netherlands

KEYWORDS

Human-computer interaction, virtual reality, tele-medicine, phobia treatment.

ABSTRACT

In this study the effects of locomotion techniques were investigated in virtual worlds made for phobia treatment. Three locomotion techniques walk-in-place, hand-controlled viewing and gaze-directed steering were systematically varied for several taks in different virtual worlds. A series of variables were measured to show the effects on presence, fear, avoidance, and simulator sickness. In the experiment was found that more natural locomotion techniques contribute to higher levels of presence and fear. This researh provides guidelines to improve the usability of systems for virtual reality exposure therapy.

INTRODUCTION

An important, and perhaps the defining aspect of presence is how people react to virtual stimuli. Do they react to a situation as if it is real, or as if they are just looking at pictures of that situation? One example of reactions that people have been known to show is fear of certain objects or situations in Virtual Environments (VE). When confronted with for instance a virtual height, most people with fear of heights will report anxiety and show behavior that is associated with fear towards a height situation. This fact has led to Virtual Reality (VR) being used effectively in treatment of amongst others fear of height (e.g. Hodges et al. 1994), fear of enclosed spaces (Alcaniz et al. 1998) and fear of flying (Krijn et al., 2004).

The relationship between presence as measured using questionnaires and fear was first investigated by (Regenbrecht et al. 1998). In an experiment with 37 non-phobic subjects they did not find a significant correlation ($r = 0.251$, $p>.10$) between presence and fear. A regression analysis did show that presence was the best predictor of fear. (Schuemie et al. 2000), in an explorative study with ten subjects being treated for fear of heights, did find a significant correlation between fear and presence reported on questionnaires. It is unclear however whether a higher sense of presence led to more fear or whether people with a higher fear of heights, because of their higher fear in the VE, report higher levels of presence. In fact, a small study performed by (Krijn 2002) indicated that people diagnosed as having fear of heights indeed report higher presence levels in VE depicting height situations when compared to neutral VEs, while non-phobics reported equal presence levels in both types of VEs. In the same study the level of presence was manipulated by increasing the level of interactivity of the VE. Higher levels of interactivity did lead to a higher level of presence but not to a higher level of fear. The most probable explanation for this outcome is the fact that the higher interactivity allowed subjects to avoid fearfull situations, which they did.

In phobia treatment a patient is confronted with situations that generate fear in the patient, while allowing the fear to attenuate through habituation (Bijl et al. 1998). It therefore seems likely that, for the treatment to be effective, it is essential that the patient experiences at least some level of fear. For effective Virtual Reality Exposure Therapy (VRET) of phobias, we should therefore design our VR setup to optimize this fear. In this paper we will investigate whether increasing the sense of presence will increase the fear a subject is experiencing.

Interaction and presence

Already some research has shown how aspects of the VE can influence the level of presence (see for instance the survey in Schuemie et al. 2001), and these aspects could thus possible also influence the level of fear a patient experiences. One aspect of the VE that can determine the level of presence is the type and level of interaction possible in the VE. Several studies have investigates the relationship between interaction and presence (Welch et al. 1996) showed that interaction, when compared to no interaction, increases the sense of presence and that longer delay between action and feedback has a negative effect on presence. (Barfield et al. 1998) showed the type of input device to have no significant effect on presence, when comparing a 3 Degrees o Freedom (DoF) joystick with a 3 DoF space mouse. Both (Hendrix & Barfield, 1996) and (Schubert et al. 2000) found a significant positive effect for including headtracking. (Slater et al., 1998) showed that more body movement can lead to a higher sense of presence. (Slater et al. 1993).

We need to determine which locomotion gives the highest sense of presenceand fear and is this best suited for VRET. Given the large number of locomotion techniques available to us it is infeasible to test all of them. Fortunately, especially the previous research by (Bowman 1999) has already investigated the accuracy and speed with which users can move through the virtual world with most of the techniques. However, it is still not very clear if a more natural technique will lead to higher presence and a higher fear. To investigate this, we compared the least natural technique to the most natural one. Although of course chosen subjectively, we would like to argue that hand-controlled viewing is the least natural interaction method, because here users have to move a device to rotate the viewpoint and press a button to move forward, which is not at all similar to the way we move in the real world. One of the most natural locomotion technique today that allows the user to move larger distances in the VE than by simple headtracking alone is walk-in-place (Slater et al. 1998). Here the user can rotate the viewpoint by rotating his or her head and can move forward by making a walking motion. An advantage of this technique is that is doesn't require an extra tracking device and is therefore relatively inexpensive.
To determine the separate effects of the headtracking and the walking motion detection, we can add a third locomotion technique: gaze directed steering, where rotation is controlled by headtracking and the user can move forward by pressing a button.

METHOD

Design

The experiment consisted of three conditions:

- *Walk-in-place*
- *Trackball* (hand controlled viewing using a handheld trackball device)
- *Headtracking (gaze-directed steering)*

The walk-in-place technique was implemented by feeding the coordinates obtained from the tracking device into a multi-layered perceptron (Herz et al. 1993) that was trained to distinguish between walking and non-walking behavior.
Subjects were required to complete a course through a VE. Each subject was exposed to only one condition.

Virtual Environment

To generate the images a PC was used equiped with a 3D-Labs Oxygen G420 graphical card and a Visette Pro Head Mounted Display (HMD). This HMD has a Field Of View (FOV) of 70 degrees diagonally and a resolution of 640*480 color elements. The position of the HMD was tracked using an Ascension Technology Flock of Birds. For a full description of the system, see (Schuemie & van der Mast 2001).
The VE used in this experiment consisted of three parts that were connected, forming a single VE: a small training space, where subjects could familiarize themselves with the interaction technique, a room designed to determine the controllability of the interaction technique and a part containing height situations aimed at determining the effect of the locomotion technique on the fear of the subjects.

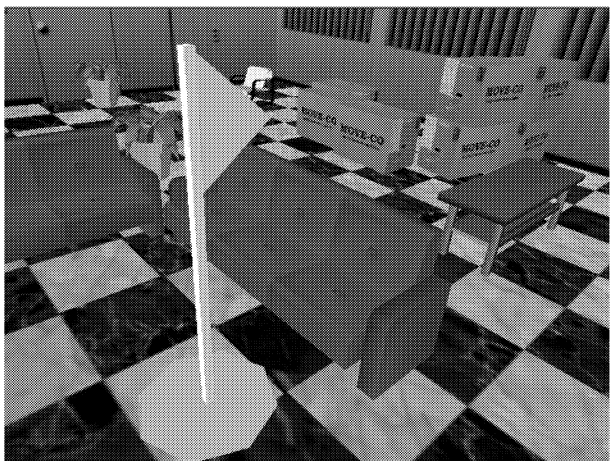


Figure 1: First part of the VE, with a flag as target for the subject to move to and obstacles such as plants and couches.

The room designed for determining the controllability of the technique contained several objects such as couches and plants. A specific location in the room was marked with a flag and a spot on the ground as depicted in *figure 1*. The subject was instructed to move to this marker and press a button to indicate that he or she had reached it. After this the marker would disappear and another would appear somewhere else in the room. All the markers were easy to find, but differed in difficulty to reach. This way we prevented the subjects from losing their orientation in the VE. Subjects were also instructed to try to avoid collisions with the objects in the room. By instructing the subjects to avoid collision instead of telling them to complete the task as quick as possible, the emphasis of the test was put on controllability instead of speed.
When the subjects reached all seven markers, they were instructed to move into an elevator, which led to the final part of the VE. The subjects were told that the accuracy test was over and were now instructed to locate boxes in the VE. These boxes were placed in the last part of the VE to stimulate the subject to look around. The top side of these boxes was open and on the inside a figure was depicted. Subjects were instructed to tell the experiment leader which figures they saw. *Figure 2* shows the firestairs found at the beginning of this part, along with a box.

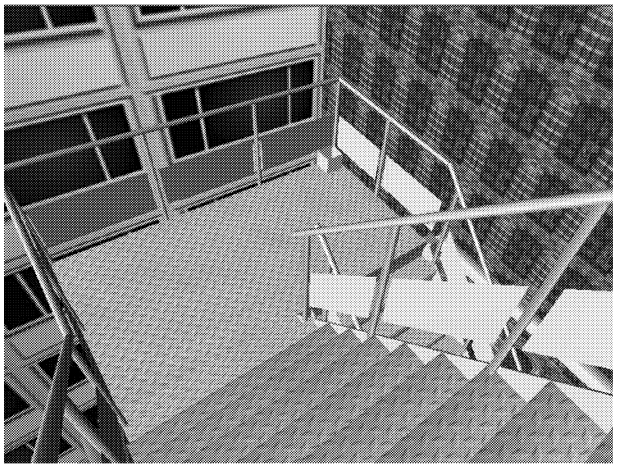


Figure 2: Firestair with a box placed in the far corner, aimed at stimulating the subjects to look in certain directions.

The top section of the VE was specifically designed to measure the avoidance behavior of the subjects. *Figure 3* shows an overview of this section. In the first half of the section (Situation 1), subjects could avoid the great depth by keeping to the right. In the second half (Situation 2), an obstacle prevented them from doing this. Situation 1 and 2 are shown in *figures 4 and 5* respectively.
To avoid any unwanted effects caused by differences in framerate between different parts of the VE and between conditions, the framerate was fixed at 15 frames per second.

Measures

Prior to the experiment itself, the subjects were required to fill in the following questionnaires:

- Acrophobia Questionnaire (AQ): This questionnaire, developed by (Cohen 1977), consists of 40 items and can be divided into a fear and an avoidance subscale.
- Motion Sickness Tendency questionnaire (MST): A short questionnaire to measure the subject's tendency to suffer from motion sickness.
- Computer Experience questionnaire (CE): 5-item questionnaire regarding the subject's experience with computers and 3D programs
- Tellegen Absorption Scale (TAS): Developed by (Tellegen & Atkinson 1974), this questionnaire aims to measure the subject's openness to absorbing and self-altering experiences.
- Simulator Sickness Questionnaire (SSQ): This questionnaire, designed by (Kennedy et al. 1993), is aimed at measuring the simulator sickness experienced by the subject. The questionnaire was filled in just prior to the experiment to acquire a base-rate.

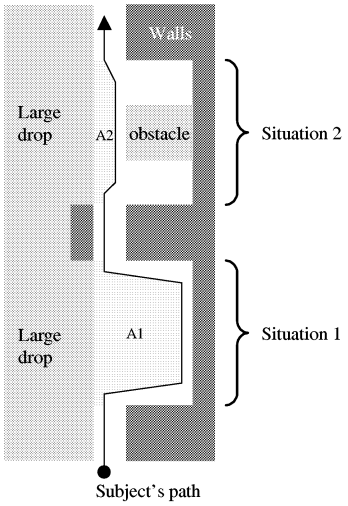


Figure 3: Overview of the top section of the VE, showing a typical subject's path and the extent of the avoidance behaviour in situation 1 (A1) and in situation 2 (A2).



Figure 4: Situation 1, where subjects could avoid the depth by keeping to the right.

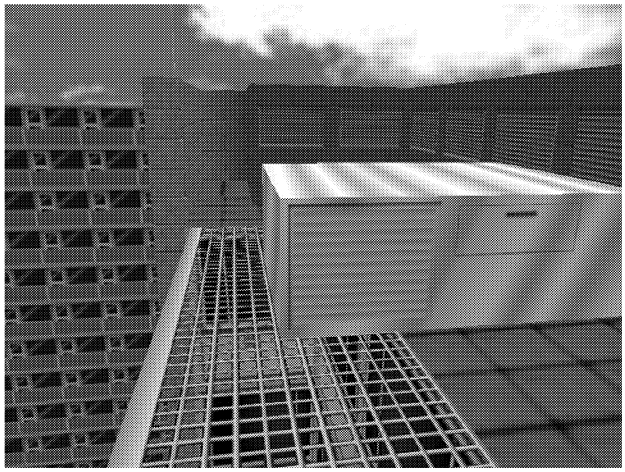


Figure 5: Situation 2, where avoidance was hindered by introducing an obstacle.

During the first part of the experiment, where the controllability of the interaction technique was tested, the following data was automatically recorded:

- *Number of collisions*: The number of times the subject collided with objects in the virtual world, and the magnitude of the collision, where a head-on collision resulted in the highest score.
- *Accuracy*: The accuracy with which subjects positioned themselves near the flags in the first part of the experiment, measured in cm.

In the second part of the experiment aimed at determining the control of fear the interaction technique provided, the following measures were recorded:

- *Heartrate:* With a five second interval, the heartrate of the subjects was recorded with an external device.
- *SUDs:* At certain predefined locations in the VE, the subjects were requested to rate their fear on a scale of 0 to 10. At the beginning of the second part of the experiment, a base-rate measure was taken to compensate for any general anxiety that the subjects might be experiencing.
- *Head-Down Rotation* (HDR)*:* Average number of degrees of downward rotation of the subject's head, used as an indication of whether subjects were looking down into the depth in height situations.
- *Avoidance (AV)*: Extend to which the subject avoided the edge of the large drop in Situation 1 and Situation two, measured as the integral of the distance in cm from the edge.

After the experiment the subjects were instructed to fill in the following questionnaires:

- *Simulator Sickness Questionnaire (SSQ):* Similar to the base-rate measured just prior to the experiment. The final score of the SSQ is calculated by subtracting the base-rate score from this score.
- *Igroup Presence Questionnaire (IPQ):* The questionnaire developed by (Schubert et al. 1999) to measure the subject's sense of presence.
- *Usability Questionnaire (USA):* Small questionnaire aimed at measuring the subject's subjective evaluation of the interaction technique. (see appendix A)

Subjects

To acquire a wide variety of subjects in terms of age, gender and computer experience, representative of the typical user population for VRET, subjects were recruited through two radio programs on Radio West, advertisement in local newspapers and in the news bulletin of the Fobievrienden Nederland foundation, posters on the campus of Delft University of Technology and email newsflashes at the Hogeschool voor de Kunsten Utrecht. Subjects were excluded from the experiment if they were suffering from epilepsy, had eyesight with deficiency greater than ± 2 dioptres (and did not wear contact lenses) , or were wearing a pacemaker.
42 subjects qualified for these criteria, 23 women and 19 men. These subjects were randomly distributed over the three conditions and received €10 upon completion of the experiment.

RESULTS

Subjects

Of the 42 subjects, three were not able to complete the experiment due to extreme fear, resulting in some missing data in the fear section of the experiment. One of these subjects was removed from the experiment altogether because of too extreme values on several scores. *Table 1* shows the some descriptive statistics of the remaining 41 subjects.

	Min	Max.	Mean (SD)	Max. Score*
AQ-fear	15	77	44.85 (16.07)	120
Computer Experience	8	20	13.66 (2.67)	25
Age	18	62	30.41 (12.42)	

* Max. score = highest possible score on the questionnaire.

Table 1: Descriptive statistics of the subjects

To test the distribution across the groups an ANOVA with condition as dependent variable was performed for AQ (F=2.309, p=.113). Age (F=1.351, p=.271), Gender (F=.233, p=.801), CE (F=.188, p=.830), MSQ (F=1.439, p=.250) and TAS (F=1.808, p=.178), indicating that the subjects were evenly distributed over the three conditions.

Measure reliability

A reliability analysis has been performed for the various questionnaires, using Cronbach's alpha as shown in *table 2*. The heartrate measure was extremely hampered by technical problems such as magnetic interference from the tracking device and was unusable.
The SUD scores were found to correlate significantly with the AQ-fear scores (Pearson correlation=.437, p=.005), confirming that people with a higher fear of heights report higher fear in the VE, suggesting that the SUD measures were reliable.

Measure	N of cases	N of items	Alpha
IPQ	41	14	.8308
AQ	64	32	.9203
MSQ	42	5	.8380
TAS	42	6	.6632
USA	41	11	.6408
SSQ(pre)	42	28	.8269
SSQ(post)	42	28	.8572
CE	64	5	.6148

Table 2: Chronbach's alpha as calculated for the various questionnaires

Accuracy of the locomotion technique

Two measures are related to the accuracy with which subjects can move through the virtual world: the accuracy with which subjects positioned themselves on the markers during the accuracy test and the number of collisions and the magnitude of these collisions with objects in the room. We investigated whether age, gender or computer experience had any influence on these two measures. No correlation was found between these variables.

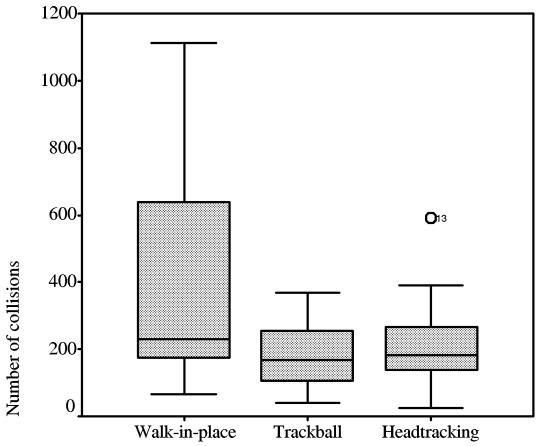


Figure 6: Boxplot of the number of collisions in the different conditions.

An ANOVA showed that there was no significant difference in positioning accuracy between conditions (F=2.087, p=.138), but that there was a significant difference in the number of collisions (F=3.273, p = .049). In *figure 6* the boxplot shows that particularly in the walk-in-place condition some people tended to have much higher numbers of collisions.

Presence

An investigation into what factors could have influenced the presence score revealed that there was no correlation between the IPQ scores and the AQ, gender, computer experience and TAS. This is interesting, since we included the TAS because we expected absorption to be important for presence. Since this did not appear to be the case, the TAS was ignored in the rest of this study.

We did however find a significant positive correlation between age and presence (Pearson correlation =.380, p=.015), indicating that older people tend to score higher on the IPQ. To compensate for this effect in our analysis, age was used as a covariate in an ANCOVA with condition as independent factor and IPQ as dependent factor. Our analysis showed a significant difference in presence between conditions (F=4.722, p =.015). *Figure 7* shows the boxplot of the IPQ scores, showing that presence is highest in the walk-in-place condition.

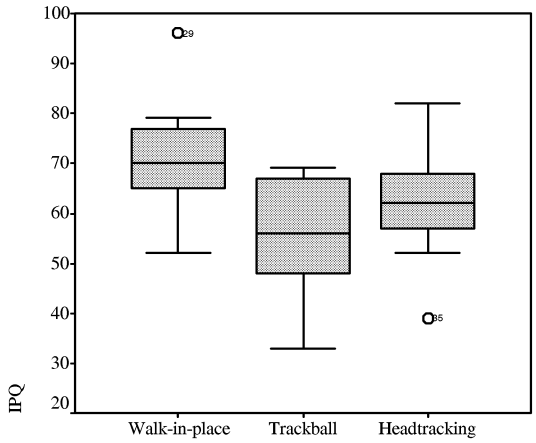


Figure 7: Boxplot of the IPQ scores.

Fear

As stated previously, fear as measured using the SUD scores correlated with the AQ-fear scores. It was therefore used as a covariate in an ANCOVA with condition as independent factor and SUD as dependent variable. For this we used the SUD measure prior to situation 1 and situation 2 (minus the base-rate measure). Other variables were not found to correlate with fear and were therefore not included in this analysis. Results showed a significant difference between conditions (F=4.015, p=.026). The boxplot of the SUD scores in the different conditions in *figure 8* shows that the fear was highest in the walk-in-place condition.

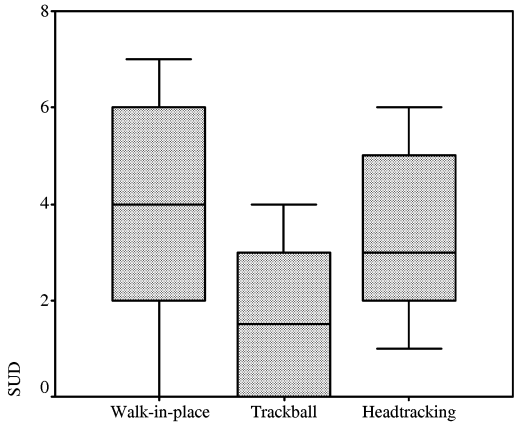


Figure 8: Boxplot of the SUD scores.

Furthermore, fear has been found to be significantly higher in situation 2 when compared to situation 1 (F=46.402, p<0.001), indicating that when avoidance is possible, this will reduce the fear a subject experiences. The reduction in SUD scores, resulting from the avoidance behavior, was not significantly different between conditions (F=.599, p=.555), showing that avoidance behavior was just as effective for all locomotion techniques.

Avoidance

The subjects could avoid the deep drop in situation 1 and 2 by either looking away as measured using the downward head rotation (please note that only the upward or downward head rotation is relevant here. Because all locomotion techniques steered the subject in the way he or she was looking, sideways head rotation resulted in moving away from the edge as measured by the second avoidance measure.), or by moving away from the edge as measured in the general avoidance measure. An ANOVA for repeated measures showed that for both measures there is a significant difference in avoidance between situation 1 and situation 2. Interestingly, people tended to look down more in situation 2 ($F=5.923$, $p=.020$). The general avoidance measure however was completely in line with our expectations, showing that people moved away much more in situation 1 compared to situation 2 ($F=70.000$, $p<0.001$).
An ANOVA showed that there was no significant difference in either avoidance measure between conditions, indicating that for each locomotion technique, subjects avoided equally.

Simulator Sickness

By subtracting the pre-test SSQ score from the post-test score, we obtained a measure of the extent to which the VE experience attributed to the subjects level of feeling sick. Oddly enough, several subjects scored negatively on this cumulative score, indicating that they felt less sick after exposure. Contact with other researchers using the same questionnaire showed that this has indeed occurred in other research as well, for reasons that are yet little understood.

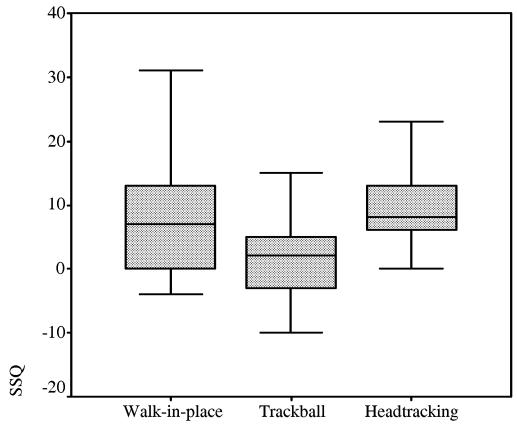


Figure 9: Boxplot of the SSQ scores

An investigation into factors contributing to simulator sickness showed no correlation between SSQ and age, gender, computer experience, AQ and MSQ.
An ANOVA showed SSQ scores to be significantly different between conditions ($F=3.257$, $p=.049$). *Figure 9* shows that SSQ is higher in both conditions using headtracking for viewpoint rotation, compared to the trackball condition.

General usability

The scores on the usability questionnaire did not correlate with the scores on computer experience, AQ, MSQ, age or gender. There also was no difference between conditions ($F=.705$. $p=0.500$), indicating that subjects rated all locomotion techniques equally.

DISCUSSION

The accuracy with which users could move through the VE did not differ greatly. Only the number of collisions was significantly higher in the walk-in-place condition.
As discussed in the previous chapter, the locomotion technique can influence fear in two ways: by changing what the user sees and by changing the user's sense of presence. The results of this experiment show that subjects indeed could use the locomotion technique to lower their fear, resulting in lower fear scores in situation 1 when compared to situation 2. Results also showed that all three locomotion techniques allowed for the same avoidance behavior, since there was no difference in avoidance as measured through the distance from the edge and the downward head rotation between conditions. This provided us with an excellent opportunity to test whether presence does indeed influence fear directly, because what the subjects saw in all three conditions was now the same.
The more natural locomotion techniques resulted both in higher presence scores and in higher fear scores, indicating that presence has a positive effect on fear. The simulator sickness was higher for the more natural locomotion techniques, and the results suggest that this is mainly attributable to the headtracking used to determine the rotation of the viewpoint. These results are in line with previous research, and can be explained by the fact that nobody actually keeps his or her head perfectly still, resulting in continuous movement of the viewpoint, whereas in the trackball condition one simply does not have to touch the trackball to keep still. The continuous movement in the headtracking conditions will confront the user more with the latency of the system and will therefore lead to more sensory conflict and thus to more simulator sickness.
The overall scores of the subjects concerning the usability of the system was not different between the devices, indicating that subjects found all techniques equally easy to use.
This experiment suggests that more natural locomotion techniques can contribute to higher levels of presence and fear in a VE, and that all techniques offer comparable levels of control over what the patient sees and hears in the HMD. The result on simulator sickness suggest that adding headtracking increases simulator sickness, but more natural interaction techniques such as walk-in-place do not add further to simulator sickness. In all, the most natural locomotion technique seems to be the one most suitable for VRET.

Conclusions

This experiment clearly confirms that, when a virtual fearful situation can be avoided, people with fear of heights will indeed show avoidance behavior in a VE, and this avoidance behavior will lead to a significant reduction in that person's fear. This experiment also showed that, when avoidance behavior is equal, presence does correlate with fear and that more natural locomotion techniques not only lead to more presence but also to more fear.

There apparently is very little difference in the general usability of the different locomotion techniques, or even between patient control and therapist control. Also, there does not seem to be much difference in simulator sickness, although headtracking does seem to add to sickness reported by subjects.

It should be noted however that only a limited number of locomotion techniques has been tested in these experiments. Therapist control was only compared to one type of patient control, and our own research has shown there to be differences between different types of patient control (i.e. between different types of locomotion technique). Also, the difference between patient and therapist control was minimized because with either type of control, patients could use headtracking to look around in the VE.

All subjects behaved as if the fearfull situation was real: instead of taking the shortest route, they chose to avoid the height situation and take the long way round.

REFERENCES

Barfield, W.; K.M. Baird; and O.J. Bjorneseth. 1998. "Presence in Virtual Environments as a function of type of input device and display update rate." *Displays*, Vol.19, 91-98.

Bowman, D.A. 1999. *Interaction Techniques for Common Tasks in Immersive Virtual Environments*, PhD thesis, Georgia Institute of Technology, June.

Cohen, D.C. 1997. "Comparison of self-report and behavioral procedures for assesing acrophobia." *Behavior Therapy*, Vol.8, 17-23.

Hertz, J.; A. Kroch; and R.G.Palmer. 1993. *Introduction to the theory of neural computation*, Addison-Wesley Publishing Company, 1993.

Hendrix, C. and W. Barfield. 1996a. "Presence within Virtual Environments as a Function of Visual Display Parameters." *Presence*, Vol.4, No.3, 274-289

Hodges, L.; B.O. Rothbaum; R. Kooper; D. Opdyke; T. Meyer; J.J. de Graaff; and J.S. Williford. 1994. "Presence as the defining factor in a VR application." *Technical report GIT-GVU-94-5*, Georgia Institute of Technology.

Kennedy, R.S.; N.E. Lane; K.S. Berbaum; and M.G. Lilienthal. 1993. "Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness." *The International Journal of Aviation Psychology*, Vol.3, No.3, 203-220.

Kennedy, R.S.; K.M. Stanney; and W.P. Dunlap. 2000. "Duration and Exposure to Virtual Environments: Sickness Curves During and Across Sessions." *Presence*, Vol.9, No.5, 466-475.

Krijn, M. 2002. Personal communication in OPROEP report.

Krijn, M; P. M. G. Emmelkamp; R.P. Olafsson; and R. Biemond. 2004. "Virtual reality exposure therapy of anxiety disorders: A review", Clinical Psychology Review 24 (2004) 259-281.

Krijn, M.; P.M.G. Emmelkamp; R. Biemond; C. de Wilde de Ligny; M.J. Schuemie; and C.A.P.G. van der Mast. 2004. "Treatment of acrophobia in Virtual Reality: the role of immersion and presence." Behavior Research and Therapy, 42(2), 229-239.

Regenbrecht, H.T.; T.W. Schubert; and F. Friedman. 1998. "Measuring the Sense of Presence and its Relation to Fear of Heights in Virtual Environments." *International Journal of Human-Computer Interaction*, 10(3), 233-249.

Schubert, T.W.; F. Friedmann; and H.T. Regenbrecht. 1999a. "Decomposing the Sense of Presence: Factor Analytic Insights." p*resented at the 2nd International Workshop on Presence*, University of Essex, UK, 6-7 April.

Schubert, T.W.; H.T. Regenbrecht; and F. Friedman. 2000. "Real and Illusory Interaction Enhance Presence in Virtual Environments." *presented at the Presence 2000 workshop.*

Schuemie, M.J.; M. Bruynzeel; L. Drost; M. Brinckman; G. de Haan; P.M.G. Emmelkamp; and C.A.P.G. van der Mast. 2000. "Treatment of Acrophobia in Virtual Reality: a Pilot Study." In *Conference Proceedings Euromedia 2000, May 8-10, Antwerp,* Belgium, F. Broeckx and L. Pauwels (Eds.). 271-275.

Schuemie, M.J. and C.A.P.G. van der Mast. 2001. "VR Testbed Configuration for Phobia Treatment Research." In *Proceedings of the Euromedia 2001 Conference, April 18-20, Valencia, Spain*, M.E. Domingo; J.C.G. Cebollada; and C.P. Salvador (Eds.). 200-204.

Schuemie, M.J.; P. van der Straaten; M. Krijn; and C.A.P.G. van der Mast. 2001. "Research on Presence in Virtual Reality: A Survey", *Cyberpsychology and Behavior*, Vol.4, No,2. April, 183-202.

Slater, M., Steed, A. , McCarthy, J., Maringelli, F., (1998b) The influence of body movement on subjective presence in virtual environments, *Human Factors*, Vol.40, No.3, 469-477.

Slater, M.; A. Steed; and M. Usoh. 1993. "The Virtual Treadmill: A Naturalistic Metaphor for Navigation in Immersive Virtual Environments.", in *First Eurographics Workshop on Virtual Reality,* M. Goebel (ed.), 71-86.

Tellegen, A. and G. Atkinson. 1974. "Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility." *Journal of Abnormal Psychology*, Vol. 83, 268-277.

Welch, R.B.; T.T. Blackmon; A. Liu; B. Mellers; and L.W. Stark. 1996. "The Effects of Pictoral Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence", *Presence*, Vol.5, No.3, 263-273.

ALMOST BLUE: THE DESIGN OF A COOPERATIVE GAME THROUGH INTEGRATING ACCESSIBLE INTERACTION

Stefano Cacciaguerra
Silvia Mirri
Paola Salomoni
Department of Computer Science
University of Bologna
Mura Anteo Zamboni, 7 40127 Bologna, Italy
{scacciag, mirri, salomoni}@cs.unibo.it

Manuel Baldassarri
Corso di Laurea in Scienze dell'Informazione
University of Bologna (Cesena)
Via Sacchi 3 47023 Cesena (FC), Italy

KEYWORDS
Game accessibility, Simulation based training, Interactive cooperative tools, Cooperative story telling.

ABSTRACT

Currently on-line entertainment is an important quality of life issue and represents a cornerstone of human culture and communication. With the growingly significant diffusion of mobile and multimodal devices, on-line entertainment is increasingly considered in an anytime, anywhere, anyone dimension. In this context new forms of on-line entertainment are offered to users, combining networked gaming, story telling and high interactivity in mixed reality contexts. This paper presents design issues related to the development of an accessible game where users cooperate to catch a serial killer. The application has been designed taking into account accessibility and offers multimodal access to sustain users with disabilities in being completely involved in the game scenario. The game, that supports edutainment issues, is built on an agent based simulation system.

1. INTRODUCTION

Playing games is a keystone of discovery, learning and, finally, growing. Playing is a way to observe, try and manipulate things or ideas, to make experiments, understand and invent. Moreover, playing with others consistently involves people in mutual communication (Shell 2002). These considerations encouraged designers to make on-line gaming an inclusive technology that supports access for anyone, regardless of disability. At the moment, on-line games are based on inaccessible interfaces and people with disabilities have difficulties or are totally left out of the game. As an example, people who are totally blind or visually impaired cannot play most of the on-line games currently available, which are exclusively based on visual cues. Failing inclusion issues is critical when games are used in edutainment applications to support teaching and learning activities. Otherwise, an accessible design could effectively offer on-line gaming access also to mobile players that use portable devices with low bandwidth, slow CPUs or small displays (IGDA 2004). Different analysts consider mobile playing one of the most interesting challenges for the on-line gaming market (Kim 2004) and recently radical innovations (in both devices and networks) have fueled the development of mobile and/or nomadic forms of entertainment. Games and other types of entertainment, as for example interactive story telling, offer users nearly virtual life experiences. Reality and virtuality are frequently mixed, by combining virtual and real characters in both virtual and real scenarios.

In this context, we present a system that implements multi-player gaming and supports accessible interactions. The game is freely inspired by *Almost Blue* (Lucarelli 1999), a thriller written by Carlo Lucarelli and set in the city of Bologna. The game combines a turn-based rule set with some features inherited from interactive story telling applications (Shell 2002). The aim of the game is to cooperate in catching a serial killer (*Iguana*) who is murdering students (and alongside running away). Players can move using only public transportation. Characters are both real and virtual and play roles that are taken from Lucarelli's story, including *Simone*, a blind guy who has listened to Iguana's voice and may be able to recognize him.

The game scenario is defined by the transportation map of Bologna and in particular is constituted by the graph of connections that are actually available. This feature supports players in learning the city's topology, understanding how to use bus and metro lines, and (being trained on) how to move in Bologna. Game Based Learning (GBL) represents one of the newest trends in e-learning and combines content provision with computer games to obtain a more effective way to learn. The use of GBL in schools implies that games must be designed and implemented to be accessible. An example of accessible GBL project is Schoolfrags (Westin 2003), an open multi-user 3D game and editor focused on 3D Web gaming accessibility. The development of accessible games is fully supported by the International Game Developers Association, through a specific workgroup, the Game Accessibility Special Interest Group (IGDA 2004).

This initiative empathizes the importance of accessibility as a design issue, especially for GBL applications.
In this context, we present a GBL system designed and developed to support the improvement of orienteering and mobility skills. These important abilities are truly essential for young people with disabilities and this consideration drove to include accessibility as a fundamental issue during the design process. The reminder of our paper is organized as follows. In (the next) Section 2 we discuss design issues whilst in Section 3 we briefly introduce the architecture of our system. Section 4 presents preliminary results and Section 5 finally concludes the paper by providing some final remarks.

2. DESIGN ISSUES

The main goals in designing the architecture of our cooperative game were:

- Helping players to learn city's topology, by offering on demand explanations on how to reach a specific destination, where bus stops are, where bus lines go, and other information related to the city context.
- Supporting accessible interaction and offer users with disabilities the opportunity to fairly play the game with all the other users and learn the city topology.
- Supplying access through mobile terminals, supporting end to end continuity and offering multimodal interaction.
- Mixing real and virtual characters: each game can be played by real users and by agents that act as *Almost Blue* characters.
- Mixing real and virtual scenarios: the system is designed to support Mixed Reality Story-telling.

The novel *Almost Blue* (Lucarelli 1999) offers an ideal setting for a game based story telling application designed with the above mentioned issues in mind. This thriller is set in Bologna and the author explores its labyrinthine streets. A young female detective, *Grazia*, who just moved to the city, leads a group of policemen who are trying to catch a serial killer called Iguana, who is murdering University students. A blind guy, Simone, helps the police to find and identify the killer, by scanning the radio waves of the city to eavesdrop on other people's lives. In the meantime Iguana is choosing Simone as his next victim. Freely inspired by this story, the game is based on the following set of criteria:

- *World*: The game world is the map of Bologna. Players can move their characters by bus, by metro or by foot. All buses and metro lines follow defined paths (the same ones really available in Bologna). Indeed, the system could represent also a different city, by specifying the appropriate topology and the related transport service map.
- *Characters*: The main characters are Iguana (who kills and runs away), Simone (who listens to conversations and informs the police) and Grazia (who tries to catch the killer and to protect Simone). The other players can act as policemen (trying to catch Iguana) or as students (trying to save their lives).
- *Main Rules*: When he escapes through the streets of Bologna Iguana is invisible, nobody can see his moves. Simone can perceive the position of Iguana every three turns. Players are initially placed randomly in the world game and Iguana's position is unknown. On each turn players move by bus, by metro or by foot to a position directly linked to the current one. Players can catch public means of transport if their position is corresponding to a bus or a metro stop.

Transport services move following their actual routes, simulating realistic time tables. As a future feature, the actual position of buses and metros could be measured with a GPS device and communicated to the system (Flintham et al. 2003). The game ends after a certain number of turns (in this case Iguana wins) or when a policeman (included Grazia) catches Iguana.
The game and the story themselves offer an appropriate environment to describe the topology of Bologna, teaching players to get their bearing in the city centre and to use public transport services. The presence of a blind character spurs designers to consider accessibility issues and to offer users an accessible interface to the game. Combining these two features, the game supports orientation and mobility capabilities of blind or, more generally, visually impaired players. The game is also designed to be used by mobile users, allowing context aware interactions. With a future availability of real bus and metro positions, users on public transportation could be located within the game space. A context aware interface may inform mobile players on game status evolution, by considering events only in the surrounding area. Moving may cause other users to miss a link and have effects on the game. The presence of mobile players (potentially unreachable) implies that the system has to set up a time out for each turn. A simple strategy to overcome the absence of an unreachable player is the following: if a player does not move within a specific time limit, he passes the turn (i.e. he can not move in this turn). This constrain could be really penalizing in a reverse hide and seek (sardines) game, like the one described above. To overcome this limit, our system proposes to delegate the move of every unreachable player to an agent that plays the assigned role until the player reconnects himself to the game.

3. ARCHITECTURE

Many architectural solutions have been used to implement cooperative games and a quite novel and promising one is represented by Multi-Agent System (MAS). A MAS is a software environment that promotes the capability to cooperate, coordinate and negotiate among different software agents (Wooldridge 2002). Programming the different agents (according to specific system rules) allows developers to build realistic virtual environments. Properly programmed, a

MAS can implement different synthetic virtual environments to explore any natural, social, or other phenomenon by involving different disciplines, such as art, biology, chemistry, physics, computer science, earth science, games, mathematics and social science (Netlogo 2004). In its current form, the MASs serves mostly didactic purposes. In other cases, they result in a quasi-experimental observation-generation environment, where it is possible to conduct realistic tests. An alternative approach (Wilensky et al. 1999; Cacciaguerra et al. 2004) suggest expanding the capability of these systems with participatory simulation activities. Hence, during an agent-based participatory simulation activity, each single user can play some system entities and see how the behavior of the system as a whole is affected by their individual behaviors. These synthetic virtual environments promote the interaction among agents, controlled by a pre-set behavioral model and human-driven agents. MAS offers many features that support the development of networked multi-player games, by providing support to synthetic virtual environments and agent-based participatory simulation. Assuming to implement the game on a MAS platform, we have developed an Accessibility Management Framework (AMF) that provides support to accessible interactions. The architecture of the system is depicted in Figure 1, which shows the integration between MAS and AMF. These two components are described in the following subsections.

3.1 Implementing *Almost Blue* on MAS

The MAS used to implement the game environment is SPADES (System for Parallel Agent Discrete Event Simulation), a middleware system for agent-based distributed simulation (Riley 2003). SPADES is designed to build agent-based simulations where each computational entity works in three phases:

- it receives sensations from a synthetic world (*sensing phase*).
- It spends some amount of computation (*thinking phase*).
- It then returns some actions (*acting phase*).

The synthetic world operates by scheduling (single) simulation events (realized one by one), and the agents simply receive sensations and send actions. SPADES has four main components: a simulation engine, a communication server, several agents and a world model. The simulation engine and the communication server are supplied as part of SPADES, while the agents and the world model must be implemented by the developers according to the set of rules that define the game. The simulation engine adopts a discrete event scheduler to advance the game. This component is in charge of queuing all pending events and coordinating all network communications.

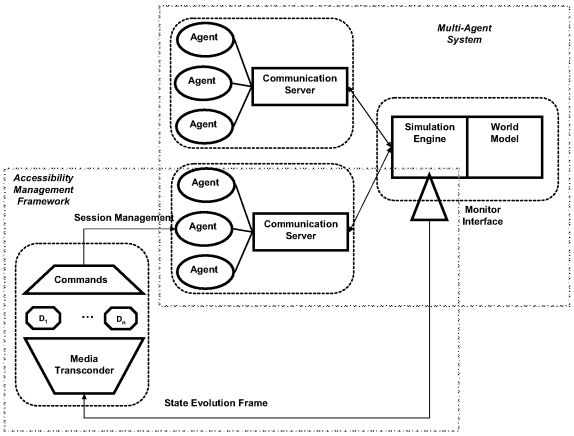


Figure 1: System Architecture

The world model is built to implement the synthetic world and its rules. The world model links up to the simulation engine as a library to constitute a single process. In particular, the world model inherits the C++ classes of SPADES (in order) to interact with the simulation engine.
Finally, the agents interface with the communication server by means of read and write system-calls over pipes, thus the agents may be programmed as the developer wants. Furthermore, it is possible to connect the simulation directly to an external program and get periodic updates about the evolution of the world a monitor interface.
In order to create our *Almost Blue* multi-player game based on SPADES, agents and world are implemented according to the set of rules described in Section 2. As a consequence, policemen, Simone, Iguana, busses and metros are programmable agents, while the map of Bologna is represented by the world model. Each bus and metro is controlled by a pre-set behavioral model. Policemen, Simone and Iguana may be both controlled by a pre-set behavioral model or driven by a human being. All the groups of agents should pursue a specific goal in the game: all policemen try to track Iguana, Iguana tries to reach Simone, and so on. A session mechanism is set up to pass the control of a character to a behavioral model when a user becomes unreachable. Busses and metros follow a pre-set model circulating on their defined paths. In particular, buses halt at their stops following a cyclical route, while the metro travels on the same railway in both directions. The world is in charge of synchronizing these agents by implementing the map of Bologna and its physical rules. Hence, the world verifies if any action is possible and correct (according to the game rules) and updates the game status. For example, every time a character wants to catch a bus or a metro at a stop, the world controls the current position of public transport to eventually change the agent status.

3.2 Integrating the Accessibility Management Framework

The above mentioned system exports the game status to the AMF, which is responsible to control the interaction between the user and the game, offering support to multimodal input and output. Every time a new game state is computed from the MAS, a text-based frame of the newly calculated state will be exported to all final users through the AMF. Using the monitor interface, a text-based snapshot of the game evolution is sent to the client media transcoder (see Figure 1). The media transcoder interprets the textual data contained in the frame and builds the appropriate output format that is finally delivered to the user. Currently the transcoding system offers three different outputs:

- A 2D visual representation of the game world, as shown by the captured screenshot (reproduced) in Figure 2.
- A 3D visual representation of the game world, that is depicted in Figure 3. The 3D interface is implemented by using OpenSceneGraph (OpenSceneGraph 2004).
- A textual description of the game status, implemented by using Ncurses (Ncurses 2004). The description corresponding to the state represented in Figure 2 and 3, is shown in Figure 4.

This last output mechanism allows a player that needs textual access (i.e. a blind user) to perceive the game evolution and use the keyboard to move its agent in real time. In order to test the textual interface we used a screen reader called Jaws for Windows (Freedom Scientific Inc. 2005). As mentioned above, the text document may be used as an output to allow game participation on hardware-limited devices such as PDAs or smart phones. In order to offer more effective support to blind or visually impaired users, the AMF provides also audio alarms every state update and keyboard shortcuts to access the textual description. The system offers users all the support needed to navigate within the game world and better understand the game status, i.e. zoom tools for graphical (2D and 3D) interfaces and dialog interaction for the textual representation.

4. RESULTS

To show the effectiveness of our approach, we conducted a preliminary study based on the use of a MAS-based game integrated with ours. In this scenario, we focused our attention on the results of the output devices of a single player. In line with this, the actions of a policeman are driven by us, while the generation/delivery of events coming from other agents are controlled by pre-set behavioral models. The player participates in the game running on a laptop machine (equipped with an AMD Athlon XP 1400, 256MB of RAM, S3 Twister K 8MB Shared RAM video card, and Sound Blaster audio card). The laptop mounted a Windows XP professional OS.

The laptop was connected to the MAS through a TCP/IP connection on a 100 Mb Ethernet card. A desktop PC (equipped with AMD Athlon XP 2400, 512MB of RAM, 100 Mb Ethernet card, mounting a Linux OS) accommodates the *Almost Blue* game running on the top of SPADES architecture.

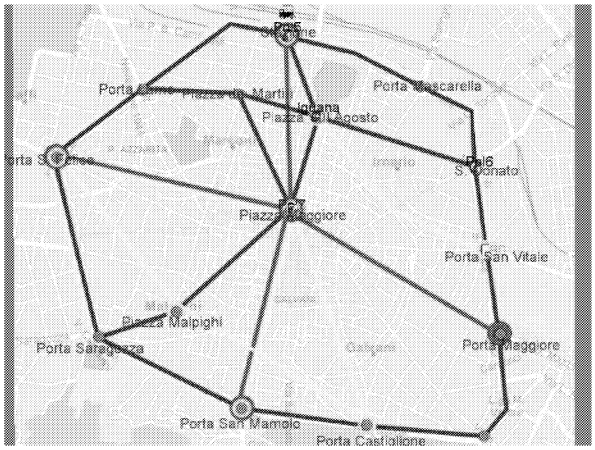


Figure 2: 2D representation of the game world (Bologna).

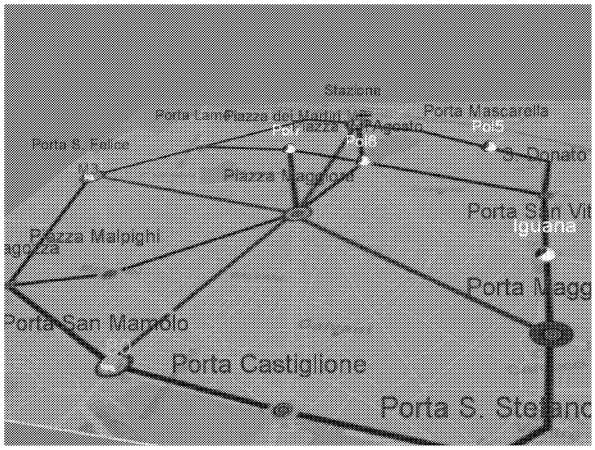


Figure 3: 3D representation of the game world (Bologna).

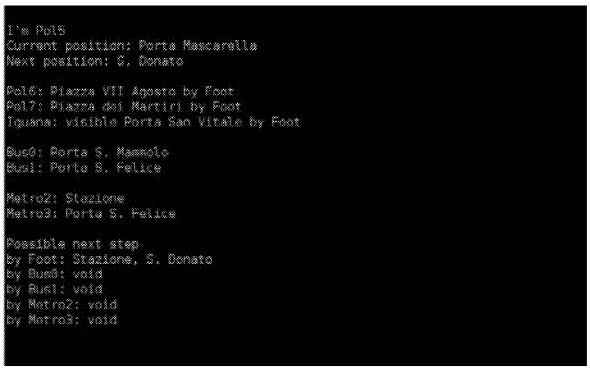


Figure 4: Textual representation of the game world.

Some of the results obtained during a test game are shown in Figure 2, 3 and 4. Several considerations are due here. Firstly, we wish to point out that our 2D/3D and text representations are rendered dynamically, following the game evolution. In the example above, agents are rendered, while they play the moves on the map. Secondly, all the different outputs representing the game world provide the same informative content by using different perceptive channels and different media. Furthermore, the user can spend his entire turn trying to understand the game evolution before make his move. As a consequence of these two features, all users, regardless of disability, can play the game with a fair possibility to win. Finally, during the development of the system, the availability of different output game state representations helped implementers to both verify the correctness of each output (by comparing visual and textual ones) and improve their expressivity.

5. CONCLUSIONS AND FUTURE WORK

We designed and developed a multi-player game architecture supporting accessible interactions that allows users with disabilities to be involved in the game and cooperate or compete in fair conditions. The game encourages players to understand and learn a city topology, describing its actual transport service network. In line with these considerations, our game-base learning application reproduces a real city and its transport services through a sophisticated simulative environment, showing the world evolution on different output devices. The accessibility goal has been reached by developing an Accessibility Management Framework that is currently able to offer 2D/3D video and textual representation of each game scene. Moreover, we can also use our framework to promote the immersion of a player inside the game environment. Two further rendering mechanisms are still under development, which will offer new interfaces based on 3D stereography and 3D positional audio. Thus, the Accessibility Management Framework is evolving in a Multimodal Interaction Framework that supports these new mixed-reality interactions. Furthermore, a new version of the game is now being designed to enforce story telling features and support the development of non-linear narrative flows with dynamic characters. This version will force players to closely interact one with the others, so as to reciprocally understand their aims and cooperate in the game. Finally, on each new game the system will create and represent a brand new version of the novel *Almost Blue*.

REFERENCES

S. Cacciaguerra, M. Roccetti, M. Roffilli and A. Lomi, "A Wireless Software Architecture for Fast 3D Rendering of Agent-Based Multimedia Simulations on Portable Devices". *In proceedings of the 1st IEEE Consumer Communications and Networking Conference (CCNC'04)*, IEEE Communications Society, Las Vegas, Nevada (USA), January 2004.

M. Flintham, R. Anastasi, S. Benford, T. Hemmings, A. Crabtree, C. Greenhalgh, T. Rodden, N. Tandavanitj, M. Adams, J. Row-Farr, "Where On-Line Meets On-The-Streets: Experieces With Mobile Mixed Reality Games". *In proceedings of the conference on Human factors in computing systems*, Ft. Lauderdale, Florida (USA), 2003.

Freedom Scientific Inc., JAWS 6.0, January 2005, http://www.freedomscientific.com/.

International Game Developers Association (IGDA), January 2005, http://www.igda.org/.

A.J. Kim, "The Network is the Game: Social Trends in Mobile Entertainment" *in online proceedings of the Game Developers Conference (GDC '04)*, San Jose, California (USA), March 2004.

C. Lucarelli, "Almost blue". City Lights Books, 1999.

Ncurses, OpenSourse, GNU, January 2005, http://www.gnu.org/software/ncurses/ncurses.html.

NetLogo User Community Models, official site Jan 2005, http://ccl.northwestern.edu/netlogo/models/community/.

OpenSceneGraph, OpenSource, OSG Community, January 2005, http://www.openscenegraph.org/.

P. Riley, "SPADES: System for Parallel Agent Discrete Event Simulation". *AI Magazine, 24(2):41–42, 2003.*

J. Schell, "Understanding entertainment: story and gameplay are one", *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications archive*, Lawrence Erlbaum Associates, pp. 835 – 843, 2002.

T. Westin, "Schoolfrags". I *Utmaningar* nr. 1, vol. 1. Stockholm: Multimedia pedagogik-teknik, Pedagogiska institutionen, Stockholms universitet, 2003

U. Wilensky and W. Stroup. Learning through Participatory Simulations: Network-based Design for Systems Learning in Classrooms. *In proceedings of the Computer Supported Collaborative Learning Conference*, Stanford University, December 1999.

M. Wooldridge, An Introduction to MultiAgent Systems, Wiley Series, 2002.

ETEC

ADAPTIVE RFID TAG IDENTIFICATION FOR ECOMMERCE APPLICATIONS

Selwyn Piramuthu
Decision and Information Sciences
University of Florida
Gainesville, Florida 32611-7169
USA
E-mail: Selwyn@ufl.edu

KEYWORDS
RFID, Supply Chain Management, Machine Learning.

ABSTRACT

Radio Frequency Identification (RFID) is promising, as a technique, to enable tracking of essential information about objects as they pass through supply chains. Information thus tracked can be utilized to efficiently operate the supply chain. An efficient supply chain is of paramount importance for smooth operation of any eCommerce system. One of the problems encountered while reading RFID tags is that of collision, which occurs when multiple tags transmit data to the same receiver slot. Data loss due to collision necessitates re-transmission of lost data. We consider this problem from Framed Slotted Aloha protocol. Using machine learning, we adaptively configure the number of slots per frame to reduce the number of collisions while improving throughput.

INTRODUCTION

RFID (Finkenzeller 2000) refers to technologies that use radio waves to uniquely identify an object. It generally involves an RFID transponder or tag, which comprises a chip and an antenna that are together attached to an object that is to be identified and/or tracked. The antenna allows the chip to transmit stored information about the object of interest to a reader. Although RFID technology is decades old, there has been renewed interest in utilizing its beneficial properties from researchers and practitioners in supply chain management. This is primarily due to (1) ease of exchange of information across supply chains, (2) decrease in cost of RFID tags, and (3) additional information (e.g., history of temperature readings of the object to which it's attached) that could be provided by these tags (unlike bar codes, which are also used for object identification) with minimal additional cost.

The beneficial aspects of RFID tags are numerous including means to track, locate, and identify a specific object of interest in real-time along a supply chain, ensuring that the object has been handled properly (some perishables, for example, need to be maintained within a certain temperature range) throughout the supply chain, keeping track of delays in the supply chain, inventory management at every stage in the supply chain, reduced cost through automation, theft detection/reduced shrinkage, record keeping/R&D audit trail, improved supply chain visibility, identification of spurious/ counterfeit products, and overall improved service levels. The creation of EPC (Electronic Product Code) coupled with lower tag costs has eased incorporation of RFID throughout the supply chain.

The last few years have seen rapid growth in implementation of RFID tags spurred especially by US-DOD as well as large retailers like Wal-Mart, Tesco, among others. Furthermore, the European Parliament has announced legislation which obliges all goods to be traceable throughout the supply chain by 2005. Commercially, there are several offerings that compete for the expanding RFID marketspace, including those from Texas Instruments (e.g., Tag-It RFID platform), Philips Semiconductors (e.g., I-Code HITAG), Hitachi (e.g., μ-Chip), and Checkpoint Systems.

In a typical system, RFID tags are attached to objects that are to be monitored. Each tag has a certain amount of internal memory where information about the object such as serial number, product composition, etc. is stored. When a tag passes through field generated by a reader, they transmit this information to the reader, thereby identifying the object. Tags can either be passive or active. Passive tags use the reader field as source of energy for the chip as well as for communication between the reader and the tag. Active tags, on the other hand, are battery-powered devices that have an active transmitter on-board. The operating range for passive tags are in the order of meters and in the order of several kilometers for active tags.

One of the advantages in using RFID tags versus barcodes is the speed at which these tags can be read where line-of-sight is not required. Several RFID tags can, in principle, be read at the same time. This leads to the problem of collision that occurs when multiple tags respond to a signal sent by the reader by transmitting data to the same receiver slot. This paper address a specific approach that is used to address this problem: Framed Slotted Aloha protocol. Using machine learning, we adaptively configure the number of slots per frame to reduce the number of collisions while improving throughput.

The next section provides a brief discussion of a few of these protocols. The proposed framework is presented in the following section. The proposed framework is illustrated using a small synthetic data set. This is followed by a brief discussion section that concludes the paper with some extensions to this study.

COLLISION AND MULTI-ACCESS PROTOCOLS

Collision occurs when more than one tag transmits at the same time in a broadcast channel. Protocols that solve the resolution problem dynamically are called Multiple Access Protocols. These can be categorized as contention protocols and collision-free protocols. Contention protocols resolve a collision after it occurs by executing a collision resolution protocol after each collision. Collision-free protocols ensure that a collision never occurs. Even when collision occurs, the pessimistic case assumes all data involved in a collision are lost. The optimistic case assumes data with highest signal strength (e.g., because of proximity to the reader when all data are of same strength or because of sheer intensity of signal strength) is read (capture effect) while the others are discarded.

Aloha was developed in the 1970s for a packet radio network by the University of Hawaii. Whenever a station has data, it transmits the data. The sender station finds out whether the transmission was successful or experienced a collision by listening to the broadcast. If there is a collision, the sender retransmits after a random wait period. This protocol involves the simplest of all reader design where the reader just listens. This protocol quickly adapts to varying number of tags. The theoretically proven maximum utilization with this protocol (assuming Poisson arrival) is 18.4%.

The Slotted Aloha protocol was an improvement where time is divided into discrete time intervals (slots) and a packet can only be transmitted at the beginning of a slot. It reduced the collision duration. Here, packets either collide completely or do not collide at all. I.e., no partial-collisions like in the pure Aloha case. This protocol doubles the channel utilization of pure Aloha to 36.8%.

The Framed Slotted Aloha protocol incorporated further discretization of time by grouping accesses into frames, with N slots per frame. Within a frame, tags transmit at most once in a randomly selected slot. This was later extended to incorporate adaptively expanding and contracting the number of slots per frame in the next upcoming round. The number of slots per frame is selected as per the number of tags in the field. There is no deterministic way to handle this problem of adaptively varying the number of slots per frame at any given point in time. We address this issue in the next section.

ILLUSTRATION OF PROPOSED FRAMEWORK

The Proposed Framework

In the Framed Slotted Aloha protocol, the number of slots per frame depends mainly on the number of tags in the reader's field. The proposed framework contains the following modules: sampler, learner, performance element, knowledge-base, and decision maker. A schematic of the proposed framework is given in the figure below.

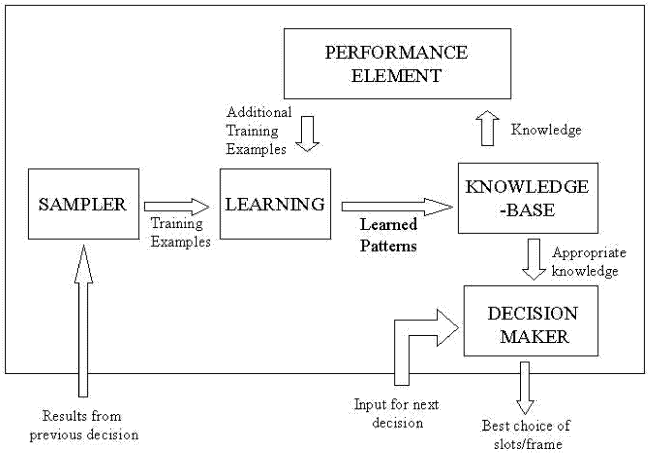


Figure 1: The Proposed Framework

Initially, the system starts with a clean knowledge-base which is later populated with "decision rules" or knowledge on what constitutes a good slots/frame number for collision-free data transmission and receiving. The sampler is simulated several times to obtain "training examples" that are used as instances to train the learning module. The learning module itself can contain any learning algorithm that is selected by the user. Some examples include decision trees, genetic algorithms, neurofuzzy algorithms, neural networks, among others. The learning module learns patterns that are associated with collision free cases as well as those that are associated with cases where there was data collision.

The learned patterns are stored in the knowledge-base. Periodically, the quality of the knowledge stored in the knowledge-base is evaluated and any deficit in knowledge is rectified by appropriately generated additional training examples that are used to incrementally update the knowledge-base.

When a new cycle is about to begin, necessary data from this cycle (e.g., number of tags in reader's field, slot length required) are input to the decision maker module, which uses this along with patterns stored in the knowledgebase to decide on the best choice of slots per frame. We assume all slots to be of the same length. This is not a restricting assumption since the proposed framework applies for varying slot lengths. Once this cycle is over, the results from this cycle (input parameter values and whether this cycle was successful without any collisions, and the number of collisions if unsuccessful) are input to the sampler module. The sampler module in turn generates appropriate training examples and feeds it to the learner module. The learner module then incrementally updates the knowledgebase.

Vogt (2002) uses intervals of low and high values for the number of tags and identifies the best number of slots per frame. We consider the dual of the problem where we use machine learning technique to identify whether a combination of these led to collision or collision-free data transmission.

An Illustration of the Proposed Framework

We illustrate the proposed framework using a simple synthetic data. The framework is applicable for data set with more fields (attributes) as well as data points. Assume the only attributes of interest are the number of tags, number of slots/frame, and a binary collision (yes/no) attribute.

example #	no. of tags (N)	slots/frame	collision ?
1	6	16	n
2	15	16	y
3	18	32	y
4	23	32	n
5	29	32	y
6	31	32	y
7	37	64	y
8	49	64	n
9	55	64	y
10	61	64	y
11	85	128	n
12	91	128	n
13	108	128	y

Table 1: An Example Data Set

We use SEE5.0 (Rulequest.com) for generating a set of decision rules from this data set. The decision rules generated from this data set using SEE5.0 are as follows:

IF (slots/frame <= 64) AND (N > 49)
THEN (COLLISION=Y)

IF (slots/frame <= 64) AND (37 < N <= 49)
THEN (COLLISION=N)

IF (slots/frame <= 64) AND (23 < N <= 37)
THEN (COLLISION=Y)

IF (slots/frame <= 64) AND (N <= 6)
THEN (COLLISION=N)

IF (slots <= 64) AND (6 < N < 23)
THEN (COLLISION=Y)

IF (slots/frame <= 64) AND (18 < N <= 23)
THEN (COLLISION=N)

IF (slots/frame > 64) AND (N <= 91)
THEN (COLLISION=N)

IF (slots/frame > 64) AND (N > 91)
THEN (COLLISION=Y)

Again, there is no significance to these specific rules. However, such rules when generated in a specific real-world context are stored in the knowledge-base module. Using this knowledge-base, a new example can be categorized as either collision or collision-free.

Clearly, it is important to note that such a knowledge-base evolves over time depending on the need to keep the knowledge-base current and updated as per existing patterns in the system. Hence, an incremental-learning machine learning algorithm would better serve such a framework.

DISCUSSION

We presented a framework to automatically choose the number of slots per frame to avoid collision in RFID systems. Although the example used is small and simple, the framework can be just as easily applied to any size of data. We are currently working on extending this by taking a closer look at the Framed Slotted Aloha protocol. Specifically, we are considering varying slot lengths, incorporating capture effect, among others.

REFERENCES

Finkenzeller, K. 2000. *RFID Handbook: Radio-Frequency Identification Fundamentals and Applications*. John Wiley & Sons.

Vogt, H. 2002. “Efficient Object Identification with Passive RFID Tags.” *International Conference on Pervasive Computing*, LNCS, Springer-Verlag.

A KNOWLEDGE-BASED DECISION-SUPPORT-SYSTEM FOR E-COMMERCE

Prof. Dr.-Ing. Stephan Kassel
Prof. Dr.-Ing. habil. Christian-Andreas Schumann
Dipl.-Inf. Kay Grebenstein
Dipl.-Inf. Claudia Tittmann
Westsächsische Hochschule Zwickau (FH) – University of Applied Sciences
P.O. Box 201037
D - 08012 Zwickau
E-mail: Stephan.Kassel@fh-zwickau.de

KEYWORDS
e-Commerce, Public Market Places, Business Intelligence, Market Intelligence, Knowledge Management, Expert Systems.

ABSTRACT

Commercialisation of the Internet has led to an advent of new business models and companies providing Internet services. Particularly, the World Wide Web evolved from a pure information repository, to a platform for interactive and multimedia applications, leading to new marketing channels. The execution of business processes is more and more shifting to the internet, causing the emergence of new markets and new technologies.

Especially small and medium-sized enterprises are confronted with the task of using the new information and communication technologies, particularly the technology of e-commerce, to gain competitive advantages. This challenge can be mastered by using commercial products providing a connection to Internet marketplaces like eBay. Thus, eBay becomes integrated into IT-systems as an additional sales channel. This is especially useful for selling seasonal products or remaining stock.

One vendor of such products is AGETO GmbH, located in Jena, Germany. In a joint project of AGETO and the University of Applied Sciences Zwickau a combination and integration of Knowledge Management, Business Intelligence and Artificial Intelligence methods are used for implementing a Decision Support System, aiming at determining the optimum selling time for auctions on public marketplaces. The necessary decision knowledge is derived from a data warehouse comprising of sales data on the electronic marketplace. The architecture of the overall system is described, and the progress of the project is depicted.

INTRODUCTION

The market research institute "Jupiter Research" forecasts that "the online commerce market in Europe will grow from €29 billion in 2003 to €117 billion in 2009, with 61 percent of European Internet users buying online and spending an average of €843 per buyer. Increased online tenure, growth in the online population, and improved broadband uptake will be the main drivers behind this solid growth". (Salcedo 2004)

New and attractive marketing channels for remnants and returned products without additional marketing costs are developing. Selling on public marketplaces like eBay opens up new customer groups, thus showing a favourable way to acquire new customers. Remnants can be liquidated fast and storage capacities are optimised, because of an increased turnover pace. The professional sales execution provides quick receipts of payment.

The enterprises have to change their business processes to adapt to the processes on the electronic marketplace. Oftenly, this is not achievable with the existing human resources in the companies. Particularly, small and medium-sized enterprises (SME) are facing the problem of opening up the potential of the new e-commerce sales channels. For those companies, AGETO GmbH developed a solution, AGETO eBay Web service, which facilitates a simple and fast access to online marketplaces without any intense knowledge of e-commerce processes. This solution is targeting on the transformation of the enormous reach of online marketplaces into measurable economic success for the enterprises.

The AGETO eBay Web service establishes channels to online marketplaces to enlarge existing e-commerce systems of retailers and producers, or provide completely new business opportuinities for enterprises which are so far not using such systems. This Web service permits a simple integration into existing corporate processes and can be adapted easily to the individual needs.

PROBLEM DEFINITION

Despite the simplicity of the technical solution of AGETO, selling goods via public marketplaces leads to new challenges by providing a unprecedented offer transparency. Increased customer orientation and forecasts of optimal sales dates are indispensable in order to succeed on these markets.

E.g. lower prices will be obtained when flooding the market with homogeneous goods. Furthermore, the achieved prices of many goods depend on external terms of sale like weather or daily news.

Online marketplaces have the potential that changing customer needs are reflecting in market behaviour directly, as these platforms are operating day and night; that some transactions are closed directly and that on auction platforms prices can develop favourably to the seller because of an increased interest of the customers (expressed by more and higher bids for offered goods).

The exact forecast of the buyers demand behaviour is of great importance; but it is difficult to predict this behaviour, because the prices on an auction platform depend on a multitude of inhomogeneous factors like the layout, the starting prize or the duration of the auction. The seller needs the control of information and knowledge about the behaviour of the buyer and any other factors of influence.

Unfortunately, this task is solved quite often inconsistently. Enormous data sets taken from internal and external sources reach the seller. In order to act on the market a seller needs information. But just providing more and more information is insufficient in the contemporary fast moving time, because the analysis of the information consumes more and more time, which is needed to react to the changes. This leads to the conclusion that explicit decision knowledge is necessary.

Techniques and methods of Knowledge Management and Business Intelligence, e.g. Data Mining or Data Warehousing, can be utilised to generate decision knowledge. Sellers make decisions by using the collected knowledge, being supported by artificial intelligence methods.

TOOLS FOR SUCCESSFULLY USING DECISION KNOWLEDGE

Knowledge Management

Decisions in the enterprises acting on the market are made based on corporate internal and external informations. Since the uprise of information society, the challenge of an ever-growing data mass exists and becomes more and more dramatic. The differentiation in manual and automated information management has been fundamentally developed from the 60's to the 80's. Subsequently, information resources management (IRM) was developed (Horton and Marchand 1982). Within the IRM manual and automated methods of data processing and management are integrated resulting in the management of the knowledge transfer processes.

In the 90's "Knowledge Management" came up due to further integration, complete penetration and dependency on information technologies in the organisation. Major tasks of Knowledge Management consist of the extraction of knowledge from different sources, the transformation and storage of generated knowledge and the supply of knowledge according to requirements. (Kuppinger and Woywode 2000)

As the main potential of knowledge is carried by the individual co-workers, Knowledge Management leads to a fundamental change of the corporate and management philosophy. Therefore, technologies are used for saving knowledge and making it available to the organisation. But to be successful, it is even more important to motivate each employee to share his knowledge with his co-workers to integrate it into the work process. Revealing the employees knowledge should be honoured and should strengthen their position in the enterprise. Above all, within global enterprises, cultural differences should be paid attention to, as they are of high importance and thus a great challenge (Grebenstein et al. 2003).

Business Intelligence

The decisions made in enterprises are usually based upon corporate data. But the enormous amount of data already existing in the company and furthermore permanently increasing causes problems. There are efforts made to support technical and high-level officers in solving their tasks by appropriate data processing tools. (Vitt 2002)

First developments were already realised by management information systems (MIS) in the 60's and 70's. These systems predominantly served to build up autonomous reporting systems. However, MIS lacked for support for interactivity and methods for problem solving. These drawbacks were succeeded by the subsequent decision support systems (DSS). The focus of these systems was set on the support of planning and decision-making processes. The next step was the development of executive information systems (EIS) in the middle of the 80's. Unfortunately, there were discrepancies between the flexibility of these systems and the requirements of high-level managers of an enterprise. The systems offered high performance visualisation tools as well as analytical tools. But they could not adequately show all necessary information from the various heterogeneous data sources in the enterprise. (Hannig 2002)

The concepts of Data Warehouses as central data collections were developed as a solution to the problem of having an inconsistent and redundant data set. At the beginning of the 90's, the term Business Intelligence was defined to denominate software tools extracting enterprise-internal and external data on one hand and transforming them into a sound database model to finally produce decision relevant information on the other hand. (Gadatsch 2002)

Most modern enterprises use more and more Business Intelligence solutions in order to handle the vast number of data necessary for a successful Customer Relationship Solution including cross selling analysis. (Vitt 2002)

Artificial intelligence and expert systems

The preparation and provision of the entire knowledge of a company, and therefore the creation of a new competition factor became a new corporate target of many companies. The demand for expert knowledge is steadily increasing. Knowledge should not be limited to only one specialist, but, summarised in a database, it should be available for all employees. (Sol 2002)

Additionally, the transparency of the existing knowledge is very important for all companies, as the quality can be increased and errors avoided through the trace ability of the results. This transparency empowers the management to improve the quality of necessary decisions, and even other experts have an easier access to the collected corporate knowledge. All experiences made by solving everyday tasks are assimilated into the system, automatically leading to the solution of similar problems. (Abramowicz 2003)

Expert systems were developed in order to meet these demands. They are a practical application of Artificial Intelligence in the form of computer programs combining – similar to human experts – domain-specific as well as domain-independent knowledge with procedures to draw conclusions (Puppe 1996).

PROJECT RESOV – AGETO DECISION SUPPORT SYSTEM

Initial Situation

The AGETO eBay Web Service has been built as an ASP solution for selling seasonal products and remaining stock at eBay while using the companies existing IT infrastructure, business processes, and logistics.



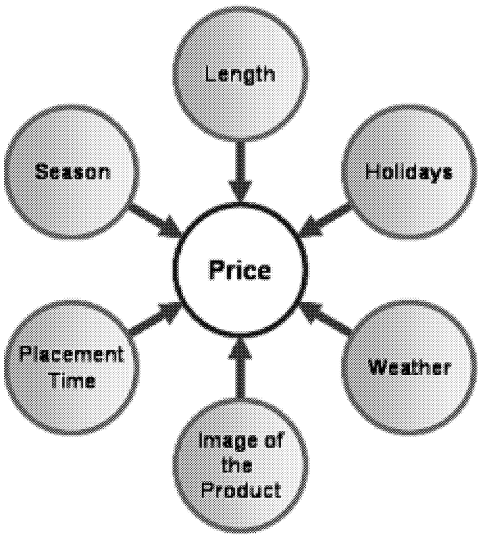
Figures 1: Architecture of AGETO eBay WebService

AGETO eBay Web Service extends existing e-commerce systems of mail order companies and manufacturers. It provides an easy integration into existing business processes like online shops and places the products of the retailers automatically on the public market place.

In the initial solution, auctions are placed automatically at the public market by setting the parameters placement time, total quantity and minimal distance of two auctions. There is no warranty of success for the placed auctions because the placement is not necessarily fitting to the search and buying patterns of the buyers on eBay.

Defining the challenge

There is a wide range of useful information helping to predict the demand pattern of the customers. Basically, the development of prices, the history of the market behaviour and external factors are taken into consideration (see figure 2). Customer tendencies are going to be forecasted by the thorough utilisation of available information; therefore recurring tendencies can be used in order to achieve higher prices for the traded goods.



Figures 2: Influences on the development of prices

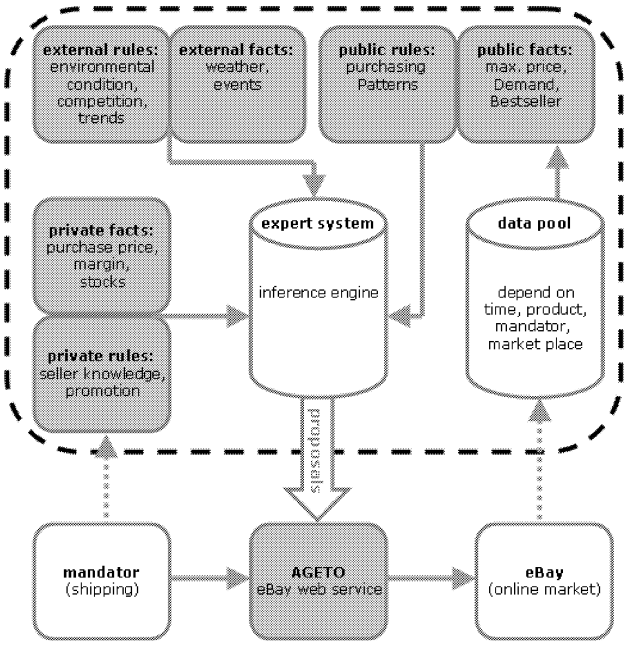
Target System

In future the AGETODSS shall optimise placement time and thereby gain sales advantages. This is achieved by the inclusion of knowledge, like the general sales behaviour of the customers and special influences, like weather and holidays, in a decision support system. Beyond that the sellers can determine whether the auctions are optimising the prices or the sold quantities.

The AGETODSS is based on an expert system. But expert systems work by using the knowledge of experts. As Decision Support Systems they deliver suggestions for decisions about a large problem area, or they work as Cockpit Systems controlling variables of the enterprise and informing the users if these variables deviate from given goal-values.

In both cases the expert systems need the knowledge of the experts and a broad basis of information from the enterprise

and external sources. This information must be gained, consolidated and prepared from appropriate data sources. The AGETO Decision Support System uses the knowledge of power sellers, but needs additional information from sales statistics and relevant data, like holidays or weather forecasts.



Figures 3: Architecture of the AGETODSS

The expert system can only make sound decisions if it comprises a rather complex Knowledge Base, which can only be built by using methods of Knowledge Management and Business Intelligence, which are themselves utilizing the global information base of the public market places. A sketch of this complex knowledge base and the heterogeneous data used in the inference process is shown in Figure 3, where you can identify the different facts and rules for private, public and external knowledge.

Development of the solution

Together with (and funded by) the company AGETO, currently a decision support system is developed in three phases to optimise the placement time of goods at electronic markets and finally to create competitive sales advantages of the AGETO eBay WebService.

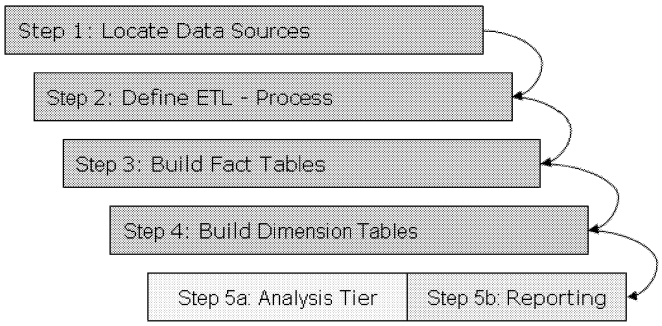
In the first (finished) phase a suitable expert system shell has been selected. It had to be an expert system that can be integrated easily into the existing system of the AGETO eBay Web service. The open source solution Mandarax (Dietrich 2004) was chosen after evaluating available expert system tools on the market.

Mandarax is a java class library for deduction rules. It provides an infrastructure for defining, managing and querying rule bases and is based on backward reasoning. Data can be integrated on the fly at query time, no replication is necessary, e.g. database records can be easily integrated as sets of facts and reflection is used in order to integrate functionality available in the object model. Other data sources like XML and SQL-based databases can be integrated as well.

After the selection of the tool, an initial rule set was developed to demonstrate the effects of using knowledge of the typical order of events on the market into the placement of auctions for several goods. The scenario was further enhanced by integrating external facts like weather forecasts or events into the prototypical system.

The striving knowledge base of the expert system has to include up-to-date information of purchase patterns on the market. On that score in the second phase a database is constructed, taking up information of different possible influence factors on user behaviour on public market places. To provide the database with these information, the technology of Data-Warehousing is used. The data warehouse is supplied with data from the databases of the AGETO eBay Web services containing details of happened transactions as well as additional external sources.

Figure 4 shows the detailed steps for building the data warehouse. After the location of the appropriable data sources, the processes of extracting, transforming and loading the data has been specified. Actually, the fact and dimension tables of the data warehouse are defined and some historical data are imported to initially fill the database. The next step consists of implementing a reporting interface for human usage of the data warehouse.



Figures 4: Progress-Model of Phase II

In the third phase the results of phase I and phase II will be combined. From the stored information of the Data-Warehouse classification numbers are computed describing auctions of specific products on certain market places. The static rules of the expert system (phase I) are changed to dynamic rules through the integration of key dates. Then the expert system can include actual changes of the behaviour of the customers on the market into its decision process. This will make the expert system reactive to new trends on the market and changing preferences of customers.

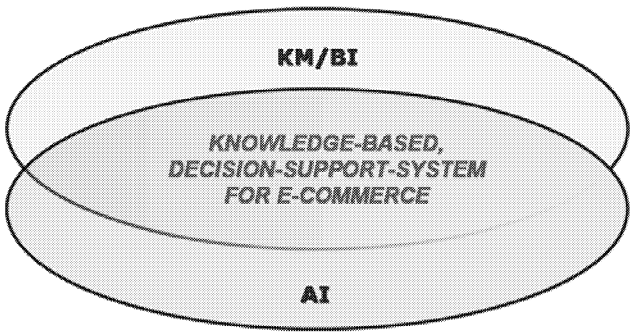
SUMMARY

The software solution AGETO eBay web service enables a simple and fast access to online marketplaces and enhances existing e-Commerce systems of retailers and producers. But nowadays, small and medium-sized enterprises want e-commerce systems reacting fast and flexibly to the constantly changing conditions of the market as well.

The AGETODSS generates decision knowledge by the techniques and methods of Knowledge Management and Business Intelligence, like Data Warehousing. The Data-Warehouse collects information from internal and external data sources and analyses them for obtaining requested and useful data. Experiences in this field were made in the cooperation with AGETO GmbH.

After the first two phases the expert system arrives at decisions using the complex Knowledge Base filled with the knowledge and experience of powersellers achieved by the methods of Knowledge Management and Business Intelligence enhanced with information on the trends and developments of the public market.

After the integration of classification numbers in the knowledge base of the expert system the AGETODSS can react fast and flexibly to changing trends and conditions on public market places like eBay. This allows a corporate usage of automated knowledge akquisition processes.



Figures 5. Integration of the results of the project

The results and experiences of the project are leading to a knowledge-based decision-support-system for e-commerce.

REFERENCES

Abramowicz, W. 2003. "Knowledge-Based Information Retrieval and Filtering from the Web". Kluwer Academic Publishers.

Aebi, R. 2000. „Kundenorientiertes Knowledge Management". Addison Wesley Verlag.

Deutscher Managerverband e. V. 2003. Die Zukunft des Managements. vdf Hochschulverlag AG. Zürich

Dietrich, J. 2004. "A Rule-Based System for eCommerce Applications". In *Proceedings of Knowledge-Based Intelligent Information and Engineering Systems: 8th International Conference* (Wellington, New Zealand, Sep. 20-25). Springer LNCS 3213 / 2004, Heidelberg, 455-463.

Gadatsch, A. 2002. „Management von Geschäftsprozessen". Vieweg Verlag, Braunschweig/Wiesbaden.

Grebenstein, K.; Schumann, C.A.; Tittmann, C.; Tsering, G.; Weber, J.; and Wolle, J. 2003. "Globale IT-Infrastruktur für die interkulturelle Kommunikation". In *Kommunikation in der globalen Wirtschaft 2003,* Bleich, S.; Jia; and W.; Schneider, F. (Eds.) Peter Lang Verlag, Frankfurt a.M., 135-153.

Hannig, U. 2002. "Knowledge Management und Business Intelligence". Springer Verlag. Berlin – Heidelberg – New York.

Horton, F. W. and Marchand, D. A. 1982 "Information Management in Public Administration". Information Resources Press. Arlington.

Kassel, S. and Grebenstein, K. 2004 „AI-based integration of business intelligence and knowledge management in enterprises", Proceedings Conference AIAI 2004. Toulouse

Krallmann, H. 2000. "Wettbewerbsvorteile durch Wissensmanagement". Schäffer-Poeschel Verlag. Stuttgart

Kuppinger, M. and Woywode, M. 2000. "Vom Internet zum Knwoledgemanagement Veränderungen in der Informationsgesellschaft". Carl Hanser Verlag München Wien.

Lehner, W. 2003. „Datenbanktechnologie für Data-Warehouse-Systeme". dpunkt verlag Gmbh. Heidelberg.

Ortmann, G.; Sydow, J. 2001. „Strategie und Strukturation". Gabler Verlag. Wiesbaden.

Puppe, F.; Gappa, U.; Poeck K. 1996. "Wissensbasierte Diagnose- und Informationssysteme". Springer Verlag. Berlin – Heidelberg – New York.

Salcedo L. Market Forecast Report European Commerce, 2003–2009. 2004. JupiterResearch. Jupitermedia Corp.

Sol, H. 2002. "Expert Systems and Artificial Inteligence in Decision Support Systems". Kluwer Academic Publishers.

Vitt, E. 2002. „Business Intelligence : Making Better Decisions Faster". Microsoft Corporation. Redmond

AUTHOR BIOGRAPHY

STEPHAN KASSEL was born in Bad Kreuznach, Germany. He studied computer science at the University of Kaiserslautern. After obtaining his degrees, he worked at several universities in Germany in the areas of Distributed Artificial Intelligence and Information Systems. He earned his Ph. D. at the Technical University of Chemnitz, in 1998. In 1999 he started to work for Intershop, an E-Commerce vendor, in international e-commerce projects. Since 2003, he holds a chair for Information Systems at the University of Applied Sciences in Zwickau, Germany.

EXPERIENCES FROM BUILDING AN INFORMATION SECURITY E-LEARNING ENVIRONMENT FOR INDUSTRY

Jorma Kajava*, Rauno Varonen**

* University of Oulu, Department of Information Processing Science, Linnanmaa, Oulu
P.O.Box 3000, FIN-90014 UNIVERSITY OF OULU, Finland
Jorma.Kajava@oulu.fi

** University of Oulu, Language Centre, Linnanmaa, Oulu
P.O.Box 7200, FIN-90014 UNIVERSITY OF OULU, Finland
rvaronen@cc.oulu.fi

KEYWORDS: e-Learning, information security, information security awareness.

ABSTRACT

Focusing on security education, this paper describes an e-learning environment that has been constructed to increase information security awareness among employees of a Finnish telecommunications company. Generally, such systems are designed and developed for educational environments, but the organization in question here is a commercial provider of public services. The system has already entered the testing phase, and the preliminary results are interesting – and somewhat unexpected: the designers' assumptions about the current knowledge level turned out to be too high. On the other hand, the design principle, based on making the components of the system as simple as possible, seems to have produced a system that delivers both functionality and stability.

INTRODUCTION

This paper discusses an e-learning based information security project carried out by a small organisation in the telecommunication service sector. Interestingly, the organisation has outsourced all software development work. All 600 employees on the payroll, mostly technically-oriented specialists, have the basic information processing skills required by their jobs, but only a few have a broader experience of the information sciences.

In a small organisation, it is relatively easy to construct a learning environment and test its functionality in practice. To be useful, the environment must be built in close collaboration with the intended users, i.e., the environment and its contents must be based on their opinions and experiences and meet their practical needs. In the present case, an extensive survey was conducted to find out the employees' wishes and needs regarding the environment.

Experiences gathered during this project form the basis for its adaptation to the university setting. The arrangement is advantageous, because development work in a small organisation is fairly flexible, and experiences accumulate quickly. Universities, on the other hand, although constituting open learning environments, tend to suffer from a higher degree of bureaucracy, which may affect the design or implementation of the learning environment. Thus, a trial system tested in practice is an essential resource for planners and constructors [3].

The topic area, information security, was selected as it is rapidly becoming a key issue for businesses, institutes of education and society at large. Traditionally, the highlight has been on the technical aspects of information security, but during the past few years, human and organisational aspects have assumed an increasingly prominent role in discussions on security [1, 10, 11].

E-learning based information security education strives to raise the awareness level of all employees. The aim is to equip them with the necessary skills and knowledge to meet the challenges that they may have to face in their everyday work [5].

The ultimate purpose of this paper is explore the use of an electronic learning environment to enhance information security awareness.

METHODS AND TECHNIQUES USED

This study was carried out in the target organisation in 2002. The immediate goal was to investigate the information security knowledge of the employees by means of questionnaires and interviews. The results were

then used to create an educational programme to correct the perceived weaknesses. As teaching material, the programme utilised various organisational guidelines complemented by educational materials compiled at the University of Oulu.

In addition to the actual learning environment and its contents, the organisation has also implemented an automatic online assignment sheet for tracking and monitoring learning. This form was designed to automate the learning process such that anyone who possessed the necessary skills and knowledge, could take the test and – on successful completion – be automatically exempted from having to go through the learning material. The system automatically handles registration and also updates registry files, when students pass the test. In essence, the environment enables company employees to study at their own pace. What it still lacks is a rewarding system for employees who receive a good grade.

Research tends to progress from theory to practice. Having become familiar with information security from various perspectives including the user and end-user perspectives, we decided to go the other way – from practice to theory. Our starting point was that, within information security, relevant knowledge usually resides within the organisation in question. What an outsider, such as a consultant, can contribute is a model or a general framework for exploring, enhancing and utilising this knowledge. On this view, pertinent information that is possessed by company employees is collected and analysed by an outside consultant who introduces a theoretical framework for analysis and may also assist in the utilisation of the results of such analysis.

The practice to theory approach is also supported by the fact that all information security events comprise a variety of aspects, some of which are strongly in relief, while others can best be described as weak signals. Even these can be taken into account thanks to the increased computing power of modern computers, which allows the unique features of each information security incident to be analysed in detail.

This study investigated the information security knowledge of different employee groups using a semi-structured theme interview. At the same time, we were able to establish which areas of information security knowledge needed improvement. This information formed the foundation for the design of the e-learning programme.

Although information security forms a central part of all e-learning, the focus of this paper lies elsewhere. Consequently, instead of a presentation of information security, we shall only refer to international standards [4] and previous research results in this area [7, 8, 12, 13].

SECURITY LEARNING

Understanding information security issues from the technical point of view is an advantage that employees of the case company have [6]. Nevertheless, since they do not have a wider perspective on other aspects of security, such as organisational or end-user related issues, they need information security training. The problem is that, being small, the company does not have the resources to allow its personnel to take time off from work to participate in security training [7]. This paper seeks to answer the question of what technical tools the company could use in this situation.

One solution is to resort to e-learning and construct an online learning environment. Many e-learning environments are realized by long distance networks, via the Internet, but our solution was to build an intranet-based environment, within the company network. Most e-learning solutions consist of very sophisticated and complicated systems, filled with content that is more entertainment than work-oriented, but we proposed a solution that is both simple and practical.

In the long run, the project reported here aims to develop a five-level solution consisting of different guidelines custom-tailored for different groups. At the first stage of this research, the focus is on guidelines that apply to all user groups. First, a questionnaire on currently prevalent practices is sent to every group. Then, having analyzed the results, the most important guidelines are collected for organisational use using the e-learning environment. It is important that these guidelines are easy to understand and follow - and it would not hurt if they were presented in a humorous way [9].

In addition, it is vital that the information security education programme is automated and computer-supported as well as transportable to different environments, including those based on older systems. As a result, the e-learning environment can be transported to other organisations working in the same field. Moreover, as the basic guidelines pose no problems for employees with a deeper knowledge of the subject, they must have the option of going directly to the test part, without having to trudge through the entire programme.

REQUIREMENTS FOR THE LEARNING ENVIRONMENT

The starting point for this project was the fact the organisation under study is a profit-seeking commercial enterprise. Aiming at improving the security level of this

organisation, the project also offers it a competitive edge through the provision of more secure telecommunication services. Technical solutions, although constituting the foundation of security, are insufficient and must be incorporated into a wider approach.

In a drive to promote information security across the organisation, the company initiated the e-learning project reported here. It seeks to find new, cost-effective, ways of offering security education to company employees. A guiding principle in this undertaking is that the education offered must be meaningful and immediately relevant to the employees.

In carrying out their everyday tasks, people tend to place a high value on usability, sometimes at the expense of security. Sadly enough, the significance of information security is often realized only after some mishap occurs.
Another balancing act is frequently observed in the context of e-learning. Striving toward a more exciting and entertaining approach, educators sometimes lose sight of their original purpose, and become entertainers rather than educators. From the organisational viewpoint, the crucial question is which is more beneficial for the functionality of the organisation: a learning environment that is highly enjoyable or a learning environment that is highly usable and functional [2]?

The organisation under study here has posed a number of requirements for its new e-learning programme. For a start, it has to run on all computer platforms used within the company. In terms of structure, it has to be sufficiently simple to be reliable in all environments [7, 8].

PRESENTATION AND LEARNING

Creating a multimedia e-learning environment requires not only technical and content-related expertise, but also a pedagogical advisor. Chief among the tasks of this advisor is to devise ways of presenting learning materials in a manner that enables learners to assimilate new knowledge into their previous knowledge structures - and thereby understand what they have learned. Another function of the pedagogical advisor is to take account of different learning strategies and styles to maximise individual learning results. If no such advisor is available, and the design of learning materials is left to technical experts, for example, the various multimedia elements may have more entertainment than educational value [9].

All learners have their own learning strategies. Part of each individual's learning strategy is their learning style, which is an essential element of the learning process. We all have our own strengths, which we rely on when processing information. Some people are characterised as holistic, while others are best described as analytic learners. The difference lies in the way they tend approach a task; holistic learners immediately strive for the big picture, whereas analytically-oriented learners favour a piecemeal approach.

Also our senses play an important part in the learning process. We receive and process information on the basis of our vision, hearing, tactile or kinetic sense. As a result, we have preferences as to how we want learning materials to be presented to us, how we want to see, hear, feel or experience the materials. Schools depend heavily on vision and hearing, at the expense of pupils who learn better by doing things, for instance. Some learners remember things as images, others as stories. A third group consists of learners who like to try things out through trial and error. Kinetically-oriented persons may memorise a telephone number or a melody, for example, as a sequence of finger movements, and learn certain things best while jogging or biking. Some people prefer to discuss things with other people, others teach themselves by talking aloud. Most people have one or more preferred senses for receiving and processing information [7].

Learning styles based solely on one way of learning are very uncommon, as most people have their own learning strategies based on their strengths, habits and preferences. Consequently, a learning environment catering a large target group must be designed to accommodate a range of approaches and styles.

LEARNING CONTENTS

The e-learning environment described here can be divided into five sections:

- topic-driven learning modules
- instructions and guidelines
- learning tasks and exercises
- glossary of information security
- feedback.

Having set the goal of creating an educational model based on these five sections, the planners had to decide, whether there should be a core learning package for all employees or several packages aimed at the various employee groups. The adopted solution was to create a core learning package consisting of the following learning modules:

- information security essentials
- passwords
- classification of data and information
- social engineering
- malicious programmes
- visitor routines and practices
- work station security
- Internet and e-mail
- guidelines for telephony.

All these modules are in current educational use.

Two of the five sections were specifically designed with the needs of organizational management in mind. In reality, however, this specialization was unnecessary, although it proved a crucial point that also senior management completed the core learning package. This resulted in the construction of three sections which, though not implemented in the system, will nonetheless form an integral part of the organization's self-study programme. These sections are:

- learning modules for management
- learning modules for maintenance personnel
- learning modules for front-end services

SIGNIFICANCE OF AUTOMATED LEARNING ENVIRONMENT TO THE SYSTEM

The purpose of the e-learning project was to construct a learning environment that requires no additional hardware or software. In addition, the environment must be accessible from all workstations within the organisation [8].

In accordance with this principle, no special software or hardware components were installed on these workstations running on various Windows operating systems. It was only required that the PCs contain standard peripherals like monitor, mouse, keyboard and, importantly, sound card for listening to the recorded samples included in the learning package. No additional video programs were necessary, as the package contained a viewing programme.

For best possible functionality, the adopted solution seeks to minimise the effects of features that exist solely to increase the entertainment value of the environment. A major concern was the fact that an increased degree of automation inevitably leads to larger program code which, in turn, poses a threat to the stability and functionality of the entire system.

These two basic requirements, accessibility and full functionality from all workstations, relate to the levels of programming languages. It is a well-established truth that the lower the level of the used language, the faster the code and the smaller the memory requirement. Such code is also more secure. Nevertheless, implementing any system involves a compromise between automation and reliability, but the old adage "small is beautiful" is well worth bearing in mind.

ESSENTIAL RESULTS

When we set out to design the e-learning environment, it was assumed that the basic information security guidelines of the organisation would be well-known by all employees. Therefore, it was a quite a revelation when the first tests in January 2004 indicated that some of the supposedly simple questions proved very hard to answer satisfactorily.

It was also revealed by the theme interviews that a great number of employees had no clear understanding of what information security is. All employee groups tended to describe it in terms of individual or isolated components. Moreover, about half of the interviewees could not explain in what ways information security issues would be relevant to their work. This shows that security education should start by breaking down the definition of information security and analysing how it affects everyday work.

Feedback relating to the use of the e-learning system was mainly concerned with its technical implementation. Typical comments include "it is slow" and "it takes a long time to start". Generally, the learners either wanted the multimedia components to load quicker or they wanted more functionality, including muting or resolution changes on the fly.

The actual content matter of the programme was not commented on. What little feedback was received indicated that the intended practical approach was appreciated and that the learning topics were experienced as having a practical value. This relative lack of feedback may be explained by the fact that the learners did not have any expectations as to the content matter, since they were unfamiliar with the subject. Also, the design of content may have been better than the technical

implementation. In general, the e-learning system was described as an interesting novelty and a number of learners indicated an interest to participate in similar training on other topics as well. With an average of 5 - 6 hours, many learners stated that they had used less time than expected on the tasks[3].

Measuring information security awareness is a difficult undertaking. One way of approaching it is to observe employees while they are working to establish the degree to which they follow the given guidelines. However, this study investigated the topic through an interview conducted among the learners. These interviews started by exploring how the learners understood the concept of information security which, after all, constitutes the foundation of information security awareness. The latter term refers to how well employees and members of society understand various information security threats and the related responsibilities. The results show that a high level of awareness has been achieved when all personnel understand the meaning of information security in its full extent and apply this knowledge in their work[3]. In addition, personnel must also be able to identify and manage a range of information security threats. Finally, they must also know what to do to avert these threats.

This bring us to the critical question: is it possible to achieve this goal using an online learning environment? Answering that question involves separating two different kinds of responsibility. Technical issues and software concerns should be tackled by hardware and software manufacturers, while issues relating to users and organisations are the province of education and are best dealt with by universities. It then follows that universities have the responsibility to define the structure and content of information security studies. In a very broad sense, this involves taking responsibility for ensuring the wellbeing of future society, which is highly dependent or communication and computers.

CONCLUSIONS

This paper discussed an e-learning environment for information security education, designed and constructed by a small Finnish telecommunications company. The experiences gathered so far indicate that the implementation of an extensive learning system of this kind must be based on simple solutions that minimise system load.

It became clear during this study that, to be successful, e-learning requires that the designers and tutors are familiar with the learners' needs and learning styles. Diverse ways of presenting the learning materials makes it easier for individuals with different learning styles to take in the information. What renders the entire task more challenging is that the content matter of information security is often fairly abstract, highlighting the importance of careful design and presentation. The education offered must provide a range of possibilities for interaction, because the chosen medium, online teaching, markedly lessens personal contact among teachers and learners. Other important aspect include the provision of support to the learners and ways of creating an inspiring atmosphere conducive to learning. Feedback provided by the teacher is an integral part of learning, and its role is even more important in online teaching, where studies are usually conducted in (relative) isolation from other learners. Time must also be allocated to electronic communication between the participants. And finally, attention must also be given to developing the proficiency of the teachers and promoting their interaction.

An interesting – and quite unexpected – finding was that the planners' assumptions about the level of security knowledge among company employees was fairly low. In the information age, every citizen should have a basic understanding of information security, otherwise they will be weak links in the chain, vulnerabilities that can be exploited. Increasing the level of security knowledge necessitates the implementation of various kinds of information security awareness programmes, an example of which is provided by the e-learning environment described here.

DISCUSSION

The prevailing situation in the organization discussed here was very confusing: IT functions were outsourced, although staff were experts in communication and information technology. In a sense, the organization concentrated all efforts on its core activities while outsourcing administrative routines. As a result, it is hardly surprising that the organization has gone on to transfer it information security services to a national umbrella organization. Currently, the development of the e-learning programme is in the process of being put out on the market as a new product.

Claiming that information security is not a major concern in the small and medium sized business sector is completely wrong. They are as likely to be affected by security breaches as major organizations. Moreover, networking can be described by process models and implies that each component in a network is like a link in a chain, and neglecting the information security of a small organization may have an adverse effect on the information security of the entire network. Thus, whatever your business is, information security is everyone's business.

Using e-learning only in the academic setting is an outdated concept. As our example shows, this channel can also be used to keep industrial and social organizations aware of important developments affecting them. In the present case, the e-learning environment was designed to promote information security awareness, but such environments may be harnessed to accommodate practically any topic of interest to any type of organization.

REFERENCES

[1] BS7799-1:fi. Standard. – Information Security Management. Part 1: Code of practice for information security management systems. (1999).

[2] Epelboin, Y. : "E-learning: putting documents 0n the web – Do and Don't". Workshop in EUNIS 2002. Porto, Portugal. (2002).

[3] Heikkinen, I., Ramet, T., "e-Learning as a part of information security education development from organisational point of view". Oulu University. Oulu. May (2004) (in Finnish).

[4] ISO/IEC JTC1/SC27, Guidelines for the Management of IT Security (GMITS). (1995).

[5] Kajava, J., Varonen, R., "Incorporating Information Security into University Infrastructure". EUNIS 2002. Porto, Portugal. (2002).

[6] Kajava, J., "Information Security Challenges for Users, End-Users and Organizations in the Beginning of the new Millennium" (Abstract in English). *Administrative Studies Journal*. Volume 19, Number 2. Tampere, Finland. (2000).

[7] Kajava, J., Varonen, R., Tuormaa, E. Nykanen, M., "Information Security Training through eLearning - Small Scale Perspective". In Eveline Riedling (ed.): *VIEWDET 2003.* Vienna International Conference on eLearning, eMedicine, eSupport. Vienna University of Technology. Nov. 26. - 28. Vienna, Austria. (2003).

[8] Kajava, J., Varonen, R., "e-Learning as a Tool: Framework for Building an Information Security Awareness Programme for a Local Teleoperator". Euromedia'2004. Hasselt, Belgium. EUROSIS. Ghent, Belgium. (2004).

[9] Neal, L., Perez, R., Miller, D., " eLearning and Fun". CHI'04 SIG. ACM. Vienna, Austria. April 26 - 29. (2004).

[10] Parker, Donn B., "Computer Security Management". Prentice Hall, Reston, USA. (1981).

[11] Royal Canadian Mounted Police, "Security in the EDP Environment". Security Information Publication, Second Edition. Gendarmere Royale du Canada. Canada. (1981).

[12] Thomson, M.E., von Solms, R., "An Effective Information Security Awareness Program for industry". *Information Security - from Small Systems to Management of Secure Infrastructures.* IFIP TC-11 Sec'97: WG 11.2 and WG 11.1 Copenhagen, Denmark. (1997).

[13] Walsh, T., "Measuring the Effectiveness of Computer Security Training". 23th Annual Security Conference and Exhibition. CSI. November 11 - 13. Chicago, Il. (1996).

DISTRIBUTED MULTIMEDIA DATABASES AND MULTIMEDIA ADAPTATION

MULTIMEDIA DATABASE SYSTEMS: WHERE ARE WE NOW?

Harald Kosch
Institute of Information Technology, University Klagenfurt
Universitätsstr. 65/67, A -9020 Klagenfurt, Austria
harald.kosch@itec.uni-klu.ac.at

KEY WORDS
Multimedia Database Systems, MPEG-7 and MPEG-21

ABSTRACT

This paper shows the evolution of Multimedia Database Systems and derives a list of open issues to be handled in the future

INTRODUCTION

A *Multimedia Database Management System (MMDBMS)* must support multimedia data types in addition to providing facilities for traditional DBMS functions like database creation, data modeling, data retrieval, data access and organization, and data independence. The area and applications have experienced tremendous growth. Especially with the rapid development of network technology, multimedia database system gets more tremendous development and multimedia information exchange becomes very important. This paper reviews the history and the current state-of-the-art in MMDBMS.

THE FIRST WAVE

The first MMDBMS rely mainly on the operating system for storing and querying files. These were ad-hoc systems that served mostly as repositories.
The mid 90s saw a *first wave* of commercial, implemented from-the-scratch, and full-fledged MMDBMS. Some of them were MediaDB, now MediaWay [1], JASMINE [2], and ITASCA that is the commercial successor of ORION [3]. They were all able to handle diverse kinds of data and provided mechanisms for querying, retrieving, inserting, and updating data. Most of these products disappeared from the market after some years of existence, and only some of them continued and adapted themselves successfully to the hardware and software advances as well as to application changes. For instance, *MediaWay* provided early very specific support for a wide variety of different media types. Specifically different media file formats varying from images, and video to PowerPoint documents can be managed segmented, linked and searched. Information on the current solutions can be found at http://www.mediaway.com.

THE SECOND WAVE

In a *second wave*, commercial systems were proposed which handle multimedia content by providing complex object types for various kinds of media. The object-oriented style provides the facility to define new data types and operators appropriate for the new kinds of media, such as video, image and audio. Therefore, broadly used commercial MMDBMSs are extensible Object-Relational DBMS (ORDBMSs). They are successfully shipped since 1996-1998, started by Informix. The current releases are significantly improved in performance and integration into the core systems. Future works include the extension of search services (mainly similarity) to video and audio and advanced presentation and browsing facilities.

The most advanced solutions are marketed by Oracle 10g, IBM DB2 and IBM Informix. They propose a similar approach for extending the basic system. As sample, we consider the IBM DB2 Universal Database Extenders. The

The *IBM DB2 Universal Database Extenders* extends the ORDBMS management to images, video, audio, and spatial objects. All these data types are modeled, accessed, and manipulated in a common framework. Features of the multimedia extenders include importing and exporting multimedia objects and their attributes into and out of a database, controlling access to non-traditional types of data with the same level of protection as traditional data, and browsing or playing objects retrieved from the database.

For instance, the *DB2 Image Extender* defines the distinct data type *DB2IMAGE* with associated user-defined functions for storing and manipulating image files (http://www-306.ibm.com/software/data/db2/extenders/). The actual contents of the image file that DB2Image describes can be stored as binary large objects (BLOB) or outside of the database in a file system. The following SQL-insert statement shows how an image is stored into a column named *image*, in a table named *example*. In this example, the content of the image comes from a server file and stored as a BLOB in the database:

```
INSERT INTO example (image) VALUES(
  DB2IMAGE (
    CURRENT SERVER,
    'pisa.jpg',                 /* source_file  */
    'JPG',                      /* source_format  */
    1,                          /* 1=BLOB, 2=file pointer */
    'my Image File'             /* comment  */
  )
)
```

The DB2 Image Extender provides similarity search functionality based on the *QBIC* technology (http://wwwqbic.almaden.ibm.com/) for images stored in the DB2IMAGE type. The QBIC technology provides the ability to query, or search, for images based on their content. Using this query mechanism, one may specify image content features, such as color values and another image file as input to a query. These features are matched against the contents of the images stored in the database, and a score is assigned to each image. A score is a double-precision floating-point value between 0 and 1 that indicates how closely an image's features match those specified in the QBIC query. The lower the score, the closer is the match. The image features that can be used in QBIC queries are average color, histogram color distribution, positional color values and the texture of an image. For instance, the following SQL statement shows an example of executing a QBIC query that ranks each image in the image column based on how closely its average color matches red:

```
SELECT CONTENTS(image),
 QBScoreFROMStr(`averageColor=<255,0,0>', image) AS SCORE
 FROM signs ORDER BY SCORE
```

Besides the commercial products, several research projects have implemented full-fledged multimedia database systems. Recently successfully terminated projects include the following:

MIRROR [4], an acronym for Multimedia Information Retrieval Reducing information OveRload and developed at the University of Twente, is a research MMDBMS that is developed to better understand the kind of data management that is required in the context of multimedia digital libraries. Its main features are an integrated approach to both content and traditional structured data management. MIRROR provides probabilistic inference mechanism, during the interaction with the user [5], which has been adopted from cognitive theories. MIRROR is implemented on top of the ORDBMS Monet database system. More information on MIRROR may be obtained at: http://wwwhome.cs.utwente.nl/~arjen/mmdb.html.
On top of MIRROR runs the *ACOI* system (http://monetdb.cwi.nl/acoi/) that is a platform for indexing and retrieval of video and image data. The system provides a plug-in architecture to subsequently index multimedia objects using various feature extraction algorithms [6]. ACOI relies on the COBRA (COntent-Based RetrievAl) video data model (only low-level descriptors). COBRA introduces a feature grammar to describe the low-level persistent meta-data and the dependencies between the extraction mechanisms [7]. Figure 1 shows the system architecture of the MMDBMS product.

DISIMA [8,9], an acronym for Distributed Multimedia DBMS, developed at the University of Alberta, is an image database system which enables content-based querying. Additional information at: http://db.uwaterloo.ca/~ddbms/projects/multimedia/
Figure 2 shows the system architecture of DISIMA.

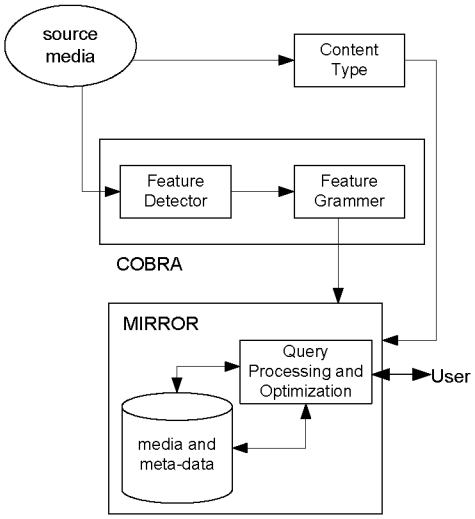


Figure 1: ACOI/MIRROR System.

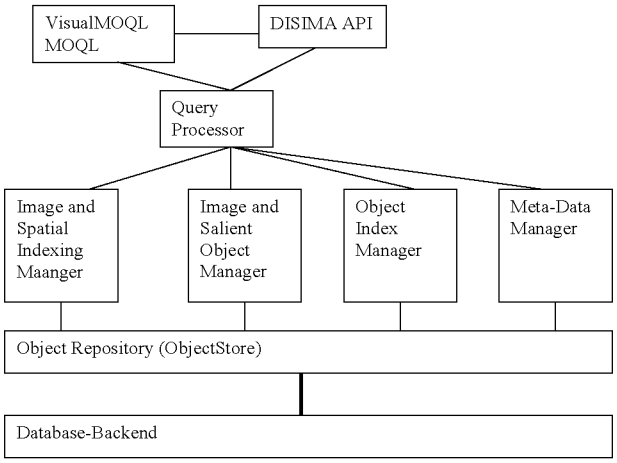


Figure 2: DISIMA System

The prototype was implemented on top of the DBMS ObjectStore. Queries are specified in the query language MOQL (or Visual MOQL for images only), which relies on a new type concept model for both image and spatial applications. The associated query languages (MOQL and Visual MOQL), extending OQL, allow spatio-temporal querying as well as the definition of a presentation specification. For instance, the following simple query allows to find all images in which a person appears (assuming that an Image table m and Person table p is defined beforehand).

```
SELECT m
   FROM Images m, Persons p
   WHERE m contains p
```

THE THIRD WAVE

The *third wave* includes currently running projects (recently finished or still running). They mainly address the needs of applications for richer semantic content. Most of them rely, thereby, on the new MPEG-standards MPEG-7 and MPEG-21. Representative running project is the *MARS* project carried out at the

University of Illinois at Urbana-Champaign and our MPEG-7 Multimedia Data Cartridge.

Before jumping into details of the projects, let us give a quick introduction to MPEG-7 and MPEG-21.
MPEG-7 [10] is the ISO/IEC 15938 standard for multimedia descriptions and issued in 2002. It is an XML-based multimedia meta-data standard, which proposes description elements for the multimedia processing cycle from the capture (e.g., logging descriptors), analysis/filtering (e.g., descriptors of the MDS, Multimedia Description Schemes), to the delivery (e.g., media variation descriptors), and interaction (e.g., user preference descriptors).

MPEG-21 [10] is the ISO/IEC 21000 standard defining an open multimedia framework. Most of the parts are in work. The driving force for MPEG-21 was the current situation that many elements exist to build an infrastructure for the delivery and consumption of multimedia content but that there exits no "big picture" to describe how these elements relate to each others. The vision for *MPEG-21* is to define an *open multimedia framework* that will enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. The intent is that the framework will cover the entire multimedia content delivery chain encompassing content creation, production, delivery, personalization, consumption, presentation and trade.

Both standards had a major influence on the current products, both in design, as implementation.

MARS, an acronym for Multimedia Analysis and Retrieval System [11,12], realizes an integrated multimedia information retrieval and database management system, that supports multimedia information as first-class objects suited for storage and retrieval based on their semantic content. MARS proposes a set of tools for an MMDBMS Back-End, as shown in Figure 3.

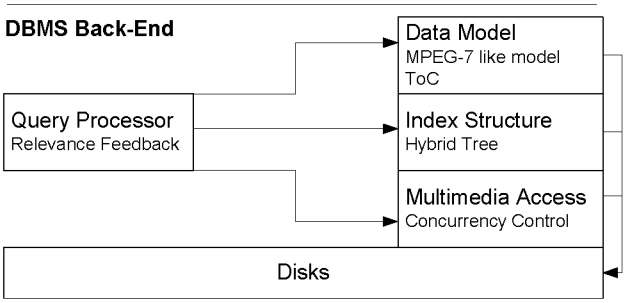


Figure 3: MARS Project

The MARS project includes the conception of a multimedia data model, for content indexing and retrieval, and for database management. The presented multimedia data model influenced the development of the MPEG-7 standard. *MARS* is a from-the-scratch management system, based on a query refinement processing. Furthermore, a table of content (ToC) extraction mechanism for videos has been developed [13]. A hybrid tree data structure to support high-dimension feature indexing in multimedia databases [14] has also been realized. For the multimedia information retrieval, an adapted relevance feedback approach that is able to learn user's information need in image databases is proposed [11, 12]. More information on MARS may be obtained from http://www-mars.ics.uci.edu.

The *MPEG-7 Multimedia Data Cartridge (MDC)* is a system extension of the Oracle 9i DBMS providing a multimedia query language, access to media, processing and optimization of queries, and indexing capacities relying on a multimedia database schema derived from MPEG-7 [15,16]

The *MDC* builds on three main concepts (see Figure 4). At first, the *Multimedia Data Model* is the database schema which is derived from MPEG-7 descriptions. It is realized with the help of the extensible type system of the cartridge environment, i.e., descriptors in the MPEG-7 schema are mapped to object types and tables. At second, the *Multimedia Indexing Framework* (MIF) which provides an extensible indexing environment for multimedia retrieval. The indexing framework is integrated into the query language and enables efficient multimedia retrieval. Finally a set of *internal and external libraries* allow the access to the media and communicate with MDC (query, insert, update, etc.).

The *Multimedia Schema* of MDC relies on the one hand on the structural and semantic parts of the MPEG-7 standard (high-level descriptions). On the other hand, object types for the MPEG-7 low-level descriptors, like color, shape, texture are provided and linked to the high-level descriptions. This enables one to retrieve multimedia data not only by low-level features, but also by semantics in combination with low-level characteristics.

The *Multimedia Indexing Framework (MIF)* offers advanced indexing services to the MMDBMS. It is generic in a way that new index types may be added without changing the interface definitions. The MIF is divided into three modules (as seen if Figure 4). Each module, especially the GistService and the Oracle Enhancement may be used on its own and may be distributed over the network.

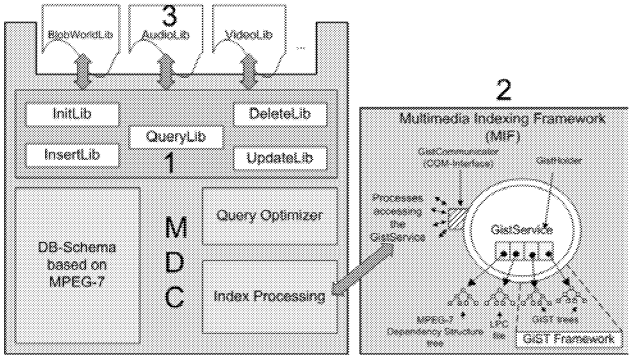


Figure 4: Multimedia Data Cartridge

The *GistService* is the main part and realized in the external address space. It offers services for index management and relies on the Generalized Search Trees

(GIST) framework (http://gist.cs.berkeley.edu/). The original GIST framework was extended to cope with several index trees at time and to be linked to the Oracle DBMS. Several index trees can be employed in our system, in the catgory balanced trees e.g., X-tree and SR-tree, and other access methods not relying on balanced trees (VA-files). This service is split into two components: *GistCommunicator* and *GistHolder*. The GistCommunicator is a COM-object (Component Object Model) used for inter-process communication between the database (by passing through the GistWrapper shared library) and the implemented access methods. Thus, the GistCommunicator supplies the necessary functionality (e.g., creating, inserting, deleting) for accessing the index structures. The GistHolder manages all currently running index trees and the accesses to them. Each index tree is identified through a global and unique ID, which is forwarded to the accessing process. The integration of MIF into the MDC is done via the index extension mechanisms of Oracle 9i. For each new index, a new Oracle *indextype* has to be defined, but the interface to the Gist remains unchanged.

OPEN ISSUES

Since the first multimedia database system ORION was developed in 1987, the area and applications of Multimedia Databases have experienced tremendous growth. From the multimedia database viewpoint two main issues have to be considered:

- **Issues on Multimedia Data Modeling:** A Multimedia Data model must deal with the issue of representing the multimedia objects, that is, designing the high and low-level abstractions of the media data to facilitate various operations. These operations may include media objects selection, insertion, editing, indexing, browsing, querying, retrieval and communication. A good data model should fulfill the following issues:

1. It is necessary to model and store media components in the database. Its storage mechanism will play an important factor affecting the performance of multimedia system.
2. A representation should be provided for the logical media structure. It is essential to represent this structure explicitly both for querying and for representation.
3. Semantic meaning should be modeled and linked to low-level characteristics and the structure of the media.
4. The meta-data necessary for the operation of the system components needs to be determined and stored in the database.
5. It is necessary to rely on international standards, e.g., MPEG-7, in order to guarantee interoperability for data sharing and data exchange.

- **Issues on Multimedia Indexing, Querying and Presentation**: The key functionality in a multimedia database is how to access and how to exchange multimedia information effectively. No matter what data model is used, data store mechanisms are employed, the most important thing is how to retrieve and communicate continuous and non-continuous media with a short real-time constraint. The key to retrieval process is similarity between two objects. The content of the object is analyzed and used to evaluate specified selection predicates, the process called *content-based retrieval*. In order to have an accurate representation of the multimedia objects in the database and the query object, different features have to be combined (e.g., texture, shape and color). The results are high-dimensional feature vectors (usually more than 1024 values). The efficiency of the similarity search in such spaces must be supported by the use of special multi-dimensional indexing structures and by dimension reduction methods, which however take care to the different nature of the feature components. Finally, similarity metrics have to re-visit to take the user perception into account.

In order to retrieve multimedia data from a database system, a query language must be provided. A multimedia query language must have abilities to handle complex spatial and temporal relationships. A powerful query language should have to deal with keywords, index on keywords and semantic contents of multimedia objects. The key to efficient communication is to rely on standards for communicating meta-data and associated media data. An important contribution is the development of the system parts of MPEG-7 and the file format of MPEG-21.

Finally, the issues considered above on data model, storage retrieval and query should not be limited to atomic multimedia data like image, audio or video. It is also interesting to retrieve composite objects, such as when we read news text, and at the same time, we open a window to see video documentaries. Hence we need the ability to deal with several atomic multimedia data, media data presentation, at the same time. MPEG-21 goes in this direction with the provision of the notion of a Digital Item, which surely will influence the multimedia database world in the future. In addition and related again to MPEG-21, it is important for a multimedia database system to use multiple representations of data for different Users and profiles, for Intellectual Property Management and for adaptation purposes.

References:

[1] B. Phillips, Mediaway presses access to multimedia database. *PC Week, 13*(7), 1996, 39-40.

[2] D. Woelk, W. Kim, & W. Luther, An object-oriented approach to multimedia databases. *ACM SIGMOD Record, 15*(2), 1986, 311-325

[3] Setrag Khoshafian, Surapol Dasananda, & Norayr Minaasian, *The Jasmine Object Database: Multimedia*

Applications for the Web (Morgan Kaufmann Publishers, 1998).

[4] Arjen P. de Vries, Mark G. L. M. van Doorn, Henk M. Blanken, & Peter M. G. Apers, The MIRROR MMDBMS architecture. *Proc. of the International Conference on Very Large Databases*, Edinburgh, Scotland, 1999, 758-761

[5] Mark G. L. M. van Doorn & Arjen P. de Vries, The psychology of multimedia databases. *Proc. of the Fifth ACM Conference on Digital Libraries*, San Antonio, TX, USA, 2000, 1-9.

[6] A.P. de Vries, M.A. Windhouwer, P.M.G. Apers, & M.L. Kersten, Information access in multimedia databases based on feature model. *New Generation Computing, 18*(4), 2000, 323-339.

[7] H. E. Blok, M. A. Windhouwer, R. van Zwol, M. Petkovic, P. M. G. Apers, M. L. Kersten, & W. Jonker, Flexible and scalable digital library search. *Proc. of the International Conference on Very Large Data Bases (VLDB)*, Roma, Italy, 2001, 705-706.

[8] V. Oria, M.T. Özsu, & P. Iglinski. VisualMOQL, The DISIMA visual query language. *Multimedia Tools and Applications, 23*(3), 2004, 185-201.

[9] V. Oria, M.T. Özsu, P. Iglinski, S. Lin, & B. Ya. DISIMA: A distributed and interoperable image database system, *Proc. of the ACM SIGMOD International Conference of Management of Data*, Dallas, Texas, USA, 2000, 600.

[10] Harald Kosch, *Distributed Multimedia Database Technologies supported by MPEG-7 and MPEG-21* (CRC Press. 280 pages. November 2003. ISBN: 0-849-31854-8).

[11] Saushik Chakrabarti, Kriengkrai Porkaew, & Sharad Mehrotra, Efficient query refinement in multimedia databases. *Proc. of the IEEE International Conference on Data Engineering (ICDE)*, San Diego, California, USA, 2000, 196.

[12] Y. Rui, T. Huang, & S. Mehrotra. Relevance feedback: a powerful tool for content-based image retrieval, *IEEE Transaction on Circuits and System for video technology, 8*(5), 1998, 25-36.

[13] Yong Rui, Thomas S. Huang, & Sharad Mehrotra. Constructing Table-of-Content for videos, *ACM Multimedia Systems, 7*(5), 1999, 359-368.

[14] Kaushik Chakrabarti & Sharad Mehrotra, Efficient concurrency control in multidimensional access methods. *Proc. of the International ACM SIGMOD Conference on Management of Data*, Philadelphia, Pennsylvania, USA, 1999, 25-36.

[15] Mario Döller & Harald Kosch, An MPEG-7 Multimedia Data Cartridge, *Proc. of the SPIE conference on Multimedia Computing and Networking 2003* (MMCN 2003), Santa Clara, California, USA, 2003, 126-137

[16] Harald Kosch, Laszlo Böszörmenyi, Mario Döller, Mulugeta Libsie, Andrea Kofler & Peter Schojer The life-cycle of Multimedia Metadata *IEEE MultiMedia. 12*(1), January/March 2005.

Selected Links:

Multimedia Database HomePage:
http://www-itec.uni-klu.ac.at/~harald/codac/

Overview of Content-Based Retrieval Tools:
http://give-lab.cs.uu.nl/cbirsurvey/

MPEG-Homepage:
http://www.chiariglione.org/mpeg/index.htm

MPEG-7 Tools:
http://m7itb.nist.gov/M7Validation.html

MPEG-7 and MPEG-21 Applications:
http://www.m4if.com/resources.php#section40

Flexible Multimedia Query Rewriting

Georges CHALHOUB, Samir SAAD, Richard CHBEIR, Kokou YETONGNON
Computer Science Department
LE2I – Bourgogne University
BP 47870 21078 Dijon CEDEX - FRANCE
Phone: (+33) 3 80 39 36 55, Fax: +333 80 39 68 69
GChalhoub@bdl.gov.lb, samir.saad@upa.edu.lb, {rchbeir, kokou}@u-bourgogne.fr

1 Introduction

A user query cannot always be satisfied by a system especially when it is based on terms and/or constraints that are not compatible with the database content. Several attempts in information retrieval aim to rewrite a query using relaxing and enriching techniques in order to relate its content to the database (or corpus) content. In spite of the existence of many proposals for textual and metadata reformulation during retrieval process, a few of them treat multimedia queries where several criteria (low-level, spatial, and semantic) and more complex constraints are considered. This paper addresses the adaptation of multicriteria queries in multimedia DBMS
To explain the motivation of this framework, we give an example of a multimedia database containing six objects shots of different traveling means in different locations (as shown in Figure 1). These objects are related within the application domain taxonomy (Figure 2): each object is represented by a node.

- If O1 is a child node of O2 then O1 *belongs to O2*
- If O1 is a father node of O2 then O1 *contains to O2.*

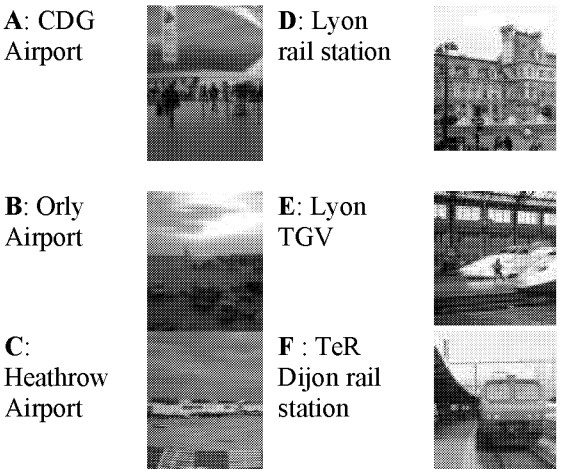


Figure 1: Multimedia Database

We believe that, in Multimedia DBMS (M²DBMS), there is a real need to rewrite the initial query every time the user is not sure of what he needs. For example, if a user wants to *find all image objects similar to the query object of* Figure 1*A*, what *the result should contain?*

- All images globally similar to the query object? The query result in this case must be Figure 1A only.
- All images *Belonging to* Paris Airport? The query result here gives Figure 1A and Figure 1B and will not give Figure 1C
- All images *showing an* Airport? The query result here gives Figure 1A, Figure 1B, and Figure 1C because Airport *integrates* London Airport which *Contains* Heathrow one in his turn.
- All images *Belonging to* Traveling? The query result here gives all the table of Figure 1 because traveling *Contains* Airport and rail station concepts.

Furthermore, we believe that a user must have the possibility to exclude objects from the result. For example, if the user wants to *find all image objects similar to the query object of* Figure 1A **except** *SNCFs'*. The query result here will not give Figure 1D, Figure 1E, and Figure 1F.

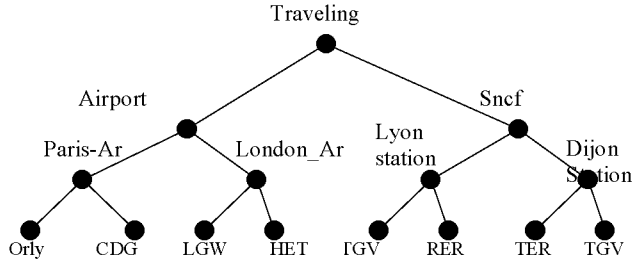


Figure 2: Belong and composition relations

In a previous work [18], we addressed the adaptation problem of multicriteria queries in multimedia DBMS. We proposed an original method to rewrite a query by relaxing and enriching terms, terms features, and several types of relations: semantic, topological, directional, temporal. This approach is based on two functions: relaxing function and constraint function. The first function consists of relaxing term and allows providing additional multimedia objects using a *neighborhood*

graph according to the desired result. We considered an undirected neighborhood graph for each term in the query. The results of the relaxing function, applied to a term, are objects that are located within a specific distance of the initial object (or term) and directly linked to it. After relaxation, a constraint function is applied to the result in order to include and/or exclude terms and to respect a display order according to the user requests.
In this paper, we improve our approach by providing more cooperative answering possibilities to the user needs. First, we extend the approach by generalizing the relaxing function in order to be able to consider in the result even the nodes that are not directly linked to the initial term. This function is recursively applied using a neighborhood directional graph where each direction represents a relation. Second, we generalize the constraint function so that the user would be able to exclude/include a term or a set of terms (sub graph) from the result.

2 Related work

Query relaxation is widely studied in *Information Retrieval* (IR) and recently in the semantic web [9, 13, 14] where problems of text search are observed, including synonymy, polysemy (a word with more than one meaning), language, etc. Term relaxation and suggestion techniques are commonly used. Relaxation techniques have been studied for decades [8, 11, 12, 10, 15]. They consist of enlarging or shrinking the search range or relaxing an answer scope to include additional information. Relaxation is considered as a platform for cooperative query answering in [17] where the authors define the relaxation as a method for expanding deductive database queries and the answers must be related in some way to the original query. In [19], the authors defined a query similarity measure to realize a collaborative search. Term suggestion techniques attempt to suggest relevant terms for queries in order to help users to formulate more effective queries and reduce unnecessary search steps. Usually, term suggestion and relaxation techniques are supported by the use of a set of logic rules [5, 7] and a thesaurus or taxonomy for representing vocabulary knowledge [8]. The problem of suggestion and relaxation becomes harder when the user uses a different ontology (or structure data source) with the system's. To integrate data from disparate ontologies, the *semantic correspondences* between their elements should be known. Manually finding such mappings is tedious, error-prone, and clearly not possible. For that reason, several automatic mapping techniques [1] are provided in the literature in order to link two taxonomies.
A few works have addressed the textual and visual similarity in multimedia DBMS [3, 16]. In [16], the multimedia query is represented in XML. To rewrite the query, the authors define two types of rules: domain dependent the authors define two types of rules: domain dependent rules and feature-dependent rules. Using these rules, an image query is then relaxed according to its components and features. However, the authors did not address spatial relationships and did not provide relaxation threshold to reduce rewriting process latency. Another interesting approach is given in [3] where the authors provide a cooperative SQL by integrating several relaxation operators to consider spatial data. This approach lacks considering three important aspects: 1-the authors only address the TAH taxonomy having generalization/specification relations but they did not provide a relaxation algorithm able to deal with these relations.2-they did not provide other semantic relations taxonomy (synonym, polysemy, related to, etc 3- the use of Euclidian distance to relax spatial relations is not appropriate for all types of relations such as topological ones.

3 Proposal

In our previous proposal, a query is represented as:

$$Query = \Sigma\{t_i[r\ t_j]\}_{Ontology_u} \rightarrow Query' = \Sigma\{t_i \cup T_i\ [r \cup R\ \ t_j \cup T_j]\}_{Ontology_{s/u}} \rightarrow Result_{Ontology_{u/s}}$$

where Query is the initial user-ontology based query, Query' is the final reformulated system-ontology query, Result is expressed using the user ontology, t represents a term, r represents a relation, T represents the additional set of terms after relaxation, and R is the additional set of relations after relaxation. To rewrite a system-ontology query Q to a new adaptive system-ontology Q^A, we can write:

$$Rewriting(Q, \{R, T\}, \{C\}) \bullet \ Q^A$$

Where R is a neighborhood graph, T is a threshold, and C is a constraint set.
In this paper we extend our rewriting process to consider a set of relations represented in the same graph and a set of threshold. Formally we write:

$$Rewriting\ (Q, \{\{R\},\{T\}\}, \{C\}) \bullet \ Q^A$$

Below, we present our approach by describing: 1- what neighborhood graph means, 2-how to weight neighborhood graph nodes, 3- how to enrich and relax initial query, 4- and how user and system constraints are applied.

3.1 Neighborhood Graphs

In [18], we defined the neighborhood graph (NG) as a uni-typed graph representing the closeness between concepts or terms. We used several NG to represent an ontology (commonly used to express the knowledge of an application domain), and each NG describes only one type of relations such as Synonym, Hierarchy, Co-occurrence, Antonomy, Related to, Compatible with,

Composition, etc. Figure 2 shows an example of neighborhood graph for the *Composition* relation. Each node has a weight that reflects its position in the graph and its importance to the user. The weighted neighborhood graph of **R** can be observed as a N-tree where the nodes are the terms, and the edge between nodes A and B indicates that the relation R relates B to A. The node weight is calculated in function of both its closeness to the root and the graph depth. However two major criteria may modify the initial weight of each node: the term frequency in the database, and the node or term popularity in queries in function of the user profile. The distance between two related nodes A_i and A_j in the graph is expressed by the absolute value of the difference of their corresponding weights $|A_i - A_j|$.

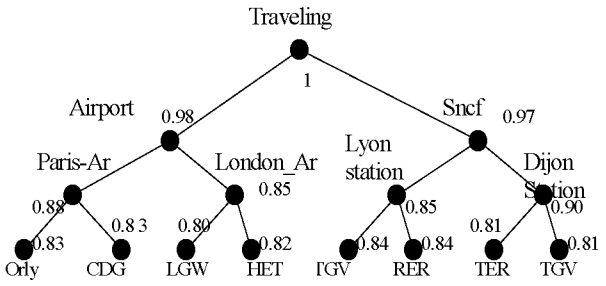


Figure 3: A weighted neighborhood graph

In this paper, we extend the above NG definition to represent several relations on the same graph. We consider a multi-directional NG where each is associated to a relation represented in the graph. For example, the NG of Figure 3 can be used to represent the two relations *Composition* and *Belong* where the descending (from root to leaves) direction is associated to the first relation and the ascending direction to the second one.
For example in the same graph (Figure 3), we represent the 2 relations: *Paris_Ar Belongs to Airport* and *Paris_Ar contain Orly*

3.2 Rewriting process

As mentioned before, our rewriting process gathers two principal functions: the Relax function F_R and the Constraint function F_C. F_R allows returning a set of relaxed terms or relations. F_C controls the returned result of F_R.

For the rewriting of only one term, we can formally write:

$$Rewrite\ (Term) = Rewrite(\delta) = F_C(F_R(\delta)) = F_C(\delta') = \delta'_C$$

We detail these two functions below.

The Relax function F_R

An element δ is a triplet (**A**, **R**, **T**) where:

- **A**: is an attribute, an attribute value or a relationship term that the user wants to relax
- **R**: is the relation for the relaxation of **A**
- **T**: is the relaxation threshold of **A** in R. **T** $\in$ [0, 1] and represents the maximum distance of a node (linked to **A**) to consider in the relaxation.

We define $\mathbf{F_R}$ as the relaxation function to be applied on an element δ (or to set of elements δ). It returns a sorted set δ' consisting of pairs (value, weight) related to **A** in descending order. Each pair is a node (value and weight) selected from the neighbourhood graph of R. The distance (weight difference) between **A** and a selected value is less than or equal to **T**. $\mathbf{F_R}$ is formally formulated as follows:

$$F_R:\ \delta = (A, R, T) \bullet F_R(\delta) = \delta' = \{(Value, W), \leq_W\}$$

For instance, to relax the term O using its components (Figure 4) within a distance of 0.1, we can write:

$$F_R("paris\text{-}Ar","Composition",0.05) = \delta' = \{(Orly, 0.83), (CDG\ 0.83)\}$$

However, users need more relaxation result. In our example users may need not only the node *B* but also its child node *I* and even the nodes *J* and *K*. In the other direction in the graph, they may need the *Root* node in the result of relaxation. We present below the extended form of F_R.

An element $\delta^{(M)}$ is a triplet
($\{\mathbf{A}\}, \{\mathbf{R_I}\}_{I=1..P}, (\mathbf{T})_{IJ\ I=1..P\ J=1..M}$) where:

- **M**: is the number of steps that the function will be applied recursively starting with the term **A**
- **{A}**: is the attribute, attribute values (or set of) or the relationship term that the user wants to relax
- **{R_I}**: is a set of P relations for the relaxation of **{A}**
- **$(T)_{IJ}$**: is the matrix of the relaxation threshold where each $\mathbf{T_{IJ}}$ is the relaxation threshold using $\mathbf{R_I}$ in the j^{th} step of application of the relax function. $\mathbf{T_{IJ}} \in [0, 1]$ and represents the maximum distance of a node (linked to current the relaxed node) to consider in the relaxation.

We extend the definition of $\mathbf{F_R}$ **to become applicable** to an element $\delta^{(M)}$ (or to a set of elements $\delta^{(M)}$). $\mathbf{F_R}$ is newly formulated as follows:

$\delta'^{(M)} = \mathbf{F_R}(\delta^{(M)}) =$

$$\bigcup_{i=1}^{M-1} \delta'^{(i)} \cup F_R(\delta'^{(M-1)}, R_{j\ j=1..p}, T_{jM\ j=1..p})$$

For instance, to relax the term *Paris_Ar* (using its components) using the following:

- Two relations *Composition* and *Belong* (P = 2) represented by the graph of Figure 3
- M=3: The user wants to apply the relax function recursively in 3 steps
- The 2 rows of threshold matrix are T_1={0.05, 0.1, 0.03} and T_2= {0.1, 0, 0} In the second and the third step the thresholds of *Belong* are null

The first step of relax application is applied to the Triplet δ^1 where:

$\delta^{(1)}$=({ *Paris_Ar*}, {*Composition, Belong*},{T_{11}, T_{21}})

$\delta'^{(1)} = F_R(\delta^1)$ = {(Airport, 0.98), (Orly, 0.83), (CDG, 0.83)}

The second step of relax application is applied to each term of $\delta'^{(1)}$ using the thresholds ***{T_{12}, T_{22}}***

$\delta'^{(2)} = \delta'^{(1)} \cup F_R(\delta'^1, \{composition, Belong\}, \{T_{21}, T_{22}\})$

$\delta'^{(2)}$={(Airport, 0.98), (*Paris_ar, 0.88)*, (*London_Ar, 0.85)*, (Orly, 0.83), (CDG, 0.83)}.

The third step of relax application is applied to each term of $\delta'^{(2)}$ using the thresholds ***{T_{13}, T_{23}}***

$\delta'^{(3)} = \delta'^{(2)} \cup F_R(\delta'^2, \{composition, Belong\}, \{T_{13}, T_{23}\})$

$\delta'^{(3)}$={(Airport, 0.98), (*Paris_ar, 0.88)*, (*London_Ar, 0.85)*, (Orly, 0.83), (CDG, 0.83), (HET, 0.82)}.

By applying the extended form, we obtained more terms in the relaxation result.

The Constraint function Fc

Usually, the user desires to exclude (or include) terms from the result set δ'. To accomplish this, we previously defined a constraint set **C** and a constraint function **Fc** as follows:

- **C** is a set of attributes, attribute values or relationship terms. It is represented by a set of couples (Val, W) where Val contains the value to include or to exclude from the relaxation result according to W value.
- $\mathbf{F_C}$ is a function that applies C constraints to the results δ' as follows:

$$F_C: (\delta', C) \bullet F_C(\delta', C) = \delta_C' = \delta' - \{(val, W') \text{ where } (val, W) \in C \text{ and } 0 \le W < 1\} \cup \{(val, W) \text{ where } (val, W) \in C \text{ and } 0 < W \le 1\}$$

In other words:

- If W = 1 then $\mathbf{F_C}$ adds the couple (Val, 1) to the relaxation result δ'
- If W = 0 then $\mathbf{F_C}$ excludes the couple containing Val from relaxation result δ'
- If 0 < W < 1 then $\mathbf{F_C}$ modifies the weight of Val in the relaxation result (if Val exists in the result)

For instance, if the user wants to:

- Include LGW to the result of relaxation even if the distance from *London_Ar* to LGW is greater than 0.03 (0.85 – 0.80 = 0.05 as in Figure 3)
- Exclude *Orly* from the result
- Modify the weight of *CDG* in the result of relaxation

C will contain then: {(LGW, 1), (Orly, 0), (CDG, 0.92)}

However, in several cases, users may need to exclude (or include) a node and its child nodes. For example, a user wants to include not only the node *Lyon Station* but also its all child nodes *TGV* and *RER*. Also they may need to exclude (or include) the node *Dijon_Station* and its ancestor nodes.

We extend C and F_C to allow the users to exclude (or include) not only a related node term but also its successor (or ancestor) nodes in the tree. C and F_C are newly formulated as follows:

$$C = \{(\delta_p'^{(M)}, W)\}$$

$$\mathbf{F_C}: (\delta'^{(M)}, C) \bullet \mathbf{F_C}(\delta'^{(M)}, C) = \delta_C'^{(M)}$$

$\delta_p'^{(M)}$ is a subset of $\delta'^{(M)}$ containing terms to include/ exclude from the relaxation result. $\delta_p'^{(M)}$ Can be reduced to a simple term if we associate null matrix to the threshold of $\delta_p^{(M)}$

For instance:

- If the users want to exclude the sub- graph nodes *(Sncf, Lyon_station, Dijon_Station, TER, TGV, RER)* from the relaxation result of our example, C will contain then (δ'^2,0) where:
  $\delta'^{(2)} = F_R(\{Sncf\}, \{, Composition\}, \{0.7, 0.9,)$
- If the users want to exclude only the term *Sncf*, **C** will contain then ($\delta'^{(1)}$,0) where:
  $\delta'^{(1)} = F_R(Sncf, null, 0)$

N-elements query

In the above explication, we described the simplest case where we rewrite only one element δ= (A, R, T) in the relaxation expression. However, usual queries contain more than one element. Similarly, we formulate the rewriting of n-elements $\delta_{i\ i=1..n}$ query as follows:

$F_C(F_R(\{\delta_{i\ i=1..n}\}), C) =$
$\mathbf{F_C}^{(n-1)}(\mathbf{F_C}(F_R(\delta_1), C_1) \times (\mathbf{F_C}(F_R(\delta_2), C_1)), C_2) \times \ldots \times (F_C(F_R(\delta_n), C_1)), C_n) =$
$\mathbf{F_C}^{(n-1)}(\mathbf{F_C}(\delta'_1, C_1) \times (\mathbf{F_C}(\delta'_2, C_1)), C_2) \times \ldots \times (\mathbf{F_C}(\delta'_n, C_1)), C_n)$
where $F_C{}^{(n-1)} = F_C$ o[1] F_C ...o F_C (n-1 times) and X is the operator cartesian product between relaxation results

In addition, in our approach, we can handle several independent expressions with relaxation order given by the user. This issue is out of the scope of this paper.

4 Conclusion

In this paper, we addressed the rewriting query problem in multimedia DBMS. We showed the limits of current rewriting methods when addressing multimedia data. We presented our method of rewriting multicriteria multimedia queries. It is built upon two major functions: the relaxation function and the constraint function
We extended the above functions and applied them recursively to obtain more cooperative results and to allow the user to include or exclude more terms from the results. Multimedia query rewriting process can be used also in the measuring of similarity between multimedia query and multimedia objects. This multimedia similarity covers several types. The future directions will be focused on testing our approach and on providing such types of similarity.

5 References

[1] A. Doan, J. Madhavan, P. Domingos, A. Halevy, "Learning to Map between Ontologies on the Semantic Web", http://www.www2002.org/CDROM/refereed/232/ (Visited at 13/09/2004).

[2] Chalhoub G., Saad S., Chbeir R., Yetongnon K., " A multimedia Meta-Database Model For Distributed Multimedia DBMS", WSEAS Transactions on Systems Issue 4, Vol 3, pp.1514-1520 June 2004.

[3] Chu W.W., Yang H., Chow G., "A Cooperative Database System (CoBase) for Query Relaxation", Proceedings of the third International Conference on Artificial Intelligence Planning Systems, Edinburgh, May 1996.

[4] Chu, Zhenyu Liu, Wenlei Mao. "Techniques for Textual Document Indexing and Retrieval via Knowledge Sources and Data Mining", Information Retrieval and Clustering, Kluwer Academic Publishers 2002.

[5] Cuppens, F. and Demolombe, R. 1989 "How to recognize interesting topics to provide cooperative answering", Information Systems, 14(2), pp. 163-173.

[6] E.N. Efthimiadis, "Query expansion", Annual Review of Information Science and Technology, 1996, 31:121-187.

[7] Hemerly, A., Casanova, M., Furtado, A., 1994, "Exploiting user models to avoid misconstruals", In Demolombe, R., and Imielinski, T., eds, Nonstandard, Queries and Nonstandard Answers, Oxford Science Publications, pp. 73-98.

[8] Y. Jing and W.B. Croft, "An association thesaurus for information retrieval", Proc. RIAO'94, 1994, pp. 146-160.

[9] Frank Meng and Wesley W. Chu, "Database query formation from natural language using semantic modeling and statistical keyword meaning disambiguation", Technical Report CSD-TR 990003, University of California, Los Angeles, 1999.

[10] Y. Qiu and H.P. Frei. Concept-based query expansion. In Proc. 16th ACM-SIGIR, 1993, pp. 160-169.

[11] K. Sparck Jones, "Automatic keyword classification for information retrieval", Butterworth, London, 1971

[12] K. Sparck Jones, "Collecting properties influencing automatic term classification", Information Storage and Retrieval, 1973, 9:499-513.

[13] Thought Treasure, http://www.signiform.com/tt/htm/tt.htm (Visited at 13/09/2004)

[14] Wen-Syan Li, Junho Shim, "Facilitating Complex Web Queries Through Visual User Interfaces and Query Relaxation", Computer Networks April 1998, Vol. 30, N 1-7, 1 pp. 149-159.

[15] J. Xu and W.B. Croft, "Query expansion using local and global document analysis", Proc. 19th ACM-SIGIR, pp. 4-11, 1996.

[16] J. P. Yoon, A. Hafez, and V. Raghavan, "Query Rewriting for Multimedia XML Data", http://www.cacs.louisiana.edu/~jyoon/publications/pub/MIS2000-

[17] T. Gaasterland, P. Godfrey, and J. Minker. *Relaxation as a platform for cooperative answering*. Journal of Intelligent Information Systems, 1:293--321, 1992.

[18] Chalhoub G., Saad S., Chbeir R., Yetongnon K., "Adaptive data retrieval in multimedia DBMS", CSITeA_04 Cairo 2004

[19] L.Fu and Dion Hoe-Lian Goh, "The effect of similarity measures on the quality of query clusters", Journal of information Science issue 30 vol 5 pp. 396-407, 2004

[1] **The symbol o is the composition operator i.e. f o g(x)= f(g(x))**

EXPERIENCING AUDIO INDEXATION ON BROADCASTED NEWS

Messaoud Bengherabi
Centre de Développement des Technologies Avancées
Cité du 20 août 1956, BP : 17-16303 Baba Hassen-Alger
E-mail: bengherabi@yahoo.com

Abdenour Sehad
Centre de Développement des Technologies Avancées
Cité du 20 août 1956, BP : 17-16303 Baba Hassen-Alger
E-mail: sehad@cdta.dz

Ahmed Mostefaoui
Laboratoire d'Informatique de Franche-Comté (LIFC)
Pôle universitaire du pays de Montbéliard
BP 71427, 25211 Montbéliard
E-mail: amostefa@pu-pm.univ-comte.fr

KEYWORDS:
Speaker Indexing, Gaussian Mixture Models, BIC Criteria

ABSTRACT

In this paper, we describe an automatic audio segmentation and classification system for broadcasted news indexation purposes. We specifically focus on speaker identification and audio scene detection. Speaker identification (SI) is based on Gaussian mixture models whereas scene detection process uses the Bayesian Information Criteria (BIC). Preliminary experience results on a small subset of Arabic broadcasted news in MPEG format show very promising results.

1- INTRODUCTION

In the recent years, video indexing and retrieval has become an active field of research. Many applications, especially in entertainment, education, medicine, etc. have shown real needs for content-based retrieval of video information. Nevertheless, this task is still a very challenging one; In fact, in order to retrieve the information from a digital collection, we cannot search natively the raw data as we do for alphanumerical information but only some kind of descriptions summarizing their contents. This is due mainly to the complex nature of video data.

Another source for collecting video indexation information, which we consider in this paper, is the audio source attached to the video data. In fact, information about the speaker identity for instance is very helpful, especially in broadcasted news indexation. It denotes in general a scene change in the video stream. The motivation for our work comes from two facts:

- Time processing needed to perform video analysis is far larger than the one needed to process audio data.
- It is more "easier" to recognize a person by analysing his/her voice than by tracking his/her face in a video stream.

For indexation purposes, we focus, in this paper, on two major processes: (a) speaker Identification process and (b) speech segmentation process. By the combination of these two processes, we show the effectiveness of our approach on a subset of Arabic broadcasted news. The rest of the paper is organized as follow: section 2 presentes the related work and the major techniques that we retain in our work. In section 3, a detailed description of the system is presented. Section 4 presents implementation details and preliminary results. Finally section 5 concludes the paper and highlights the future works.

2- RELATED WORK

Speaker indexing is useful for many multimedia applications such as meeting or teleconference. It sequentially detects the points where a speaker identity changes in a multi-speaker audio stream, and classifies each speaker segment. Speech segmentation is the location of boundaries between speakers in the audio stream. A widely adopted algorithm for the audio segmentation is based on the Bayesian Information Criterion (BIC) (Chen and Gopalakrishnan 1998, Tritschler and Gopinath 1999). Bayesian Information Criterion can divide audio streams into output streams that are long enough to provide useful information. Other advantages are no need to do model training and no prior knowledge concerning acoustic conditions are required.

Speaker identification (SI) is the process of finding the identity of an unknown speaker by comparing his/her voice with voices of registered speakers in the database. It's a one-to-many comparison. The basic structure of SI system (SIS) is shown in Figure 1.

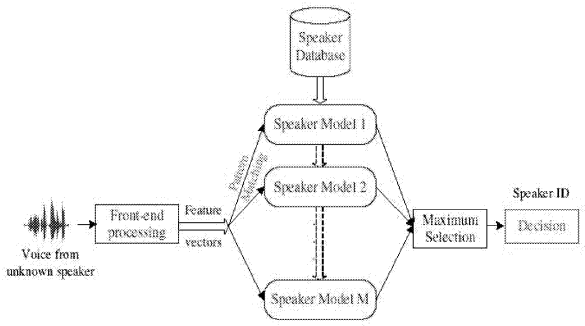


Figure 1: Basic Structure of Speaker Identification

Different classification paradigms using different modelling techniques for text independent speaker identification could

be found (Campbell 1997), such as Vector **Quantization** Modelling, Neural Network and Gaussian Mixture Model. Vector Quantization Modelling (Rosenberg and Soong 1986) is a template modelling where the temporal information of the features is averaged out. A codebook is used to represent the features of the speech. However, this method suffers from the fact that the averaging process discards much speaker-dependent information and it requires long speech utterance (more than 20 seconds) to derive stable long-term speech statistics. Neural Network (Rudasi and Zahorian 1991) does not train the individual model separately. It is trained to model the decision function which best discriminates speakers within a known set. The merits of Neural Network is that it requires a small number of parameters and have better performance in recognition as compare to Vector Quantization. However, the drawback is that the complete network must be retrained when a new speaker or language is added to the data set. This is very time consuming and inconvenient. Gaussian Mixture Model is a statistical model that models the underlying sound of a person's voice.

Gaussian Mixture Model (Reynolds and Rose 1995) represents the broad acoustic classes which reflect general speaker-dependent vocal tract configurations. The vocal tract configuration of a person is unique, so it can be used to model the speaker identity. The Gaussian mixture density provides a smooth approximation to the sample distribution of observations obtained from utterances by a given speaker. It is computationally efficient and only short utterances are needed. Furthermore, it is text and language independent. For these reasons, we retained the Gaussian Mixture Model as the basis for our identification engine.

3- A DETAILED DESCRIPTION OF THE SYSTEM

The block diagram of the speaker indexing system is shown in Figure 2.

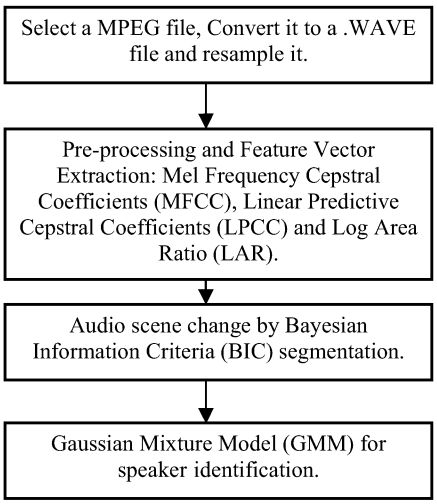


Figure 2: Block Diagram of the Speaker Indexing System

3.1 The Parameter Extraction Module

Before identifying any voices or training a person to be identified by the system, the voice signal must be processed to extract important characteristics of speech. By using only the important speech characteristics, the amount of data used for comparisons is greatly reduced and thus, less computation and less time is needed for comparisons. The steps used before feature extraction are: silence detection, pre-emphasis, frame blocking, and Hamming windowing.

The performance of a speaker identification system is highly dependent on the quality of the selected speech features. Most of the current proposed speaker identification systems use Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC) as feature vectors (Pravinkumar and Wasfy 2001). Although MFCC and LPCC were proved to be two very good features in speech recognition, they are not necessarily being as good in speaker identification. In a recent work (Chow and Abdulla 2004), Log Area Ratio LAR features extracted from LPC (Linear Predictive Coefficients) processing of speech signal (reflection coefficients) are shown to outperform the MFCC parameters with less computational complexity especially for clean speech because it has a linear spectral sensitivity and is more robust to quantization noise.

3.2 The Speaker Identification Engine

Gaussian Mixture Model is a type of density model that use to represent the speaker model. It strictly follows the probabilistic rules. Advantage of Gaussian mixture model is text independent, robust, computationally efficient and easy to implement.

GMM (Reynolds and Rose 1995) is commonly used for language identification, gender identification and speaker identification. It is more accurate for identifying speaker than gender. A Gaussian mixture model is modelled by many different Gaussian distributions. Each of the Gaussian distribution has its mean; variance and weighting in the Gaussian mixture model it models. This is shown in Figure 3. Figure 4 illustrates how a one dimensional GMM with 3 Gaussians is used to model a multimodal histogram which can not be modelled by a single Gaussian.

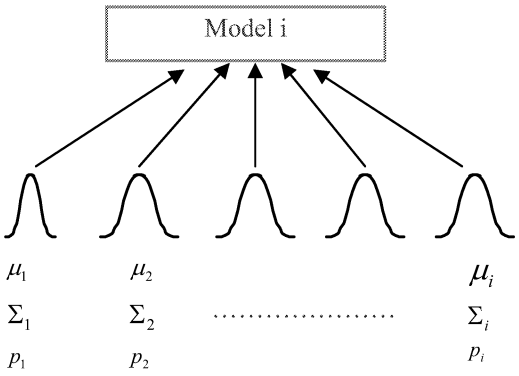


Figure 3: Construction of GMM

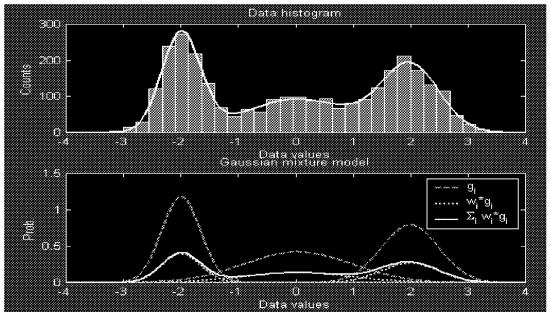


Figure 4: 1 -D GMM with Three Gaussians

assume M is the number of small Gaussian distribution used to model the Gaussian mixture model. The following equation is used to calculate the Gaussian Mixture Density for D-dimensional random vector:

$$p(\vec{x}) = \sum_{i=1}^{M} p_i b_i(\vec{x})$$

where $\vec{x}$ is a D-dimensional random vector
i = 1…, M

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)'\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)\right\}$$

$\vec{x}$ is the input D-dimensional random vector

b_i is the score for Gaussian distribution i

μ_i is the mean for Gaussian distribution i

Σ_i is the covariance matrices for Gaussian distribution i

p_i is the weighting of the score for the Gaussian distribution i where $\sum_{i=1}^{M} p_i = 1$,

3.2.1 Training Procedure for Gaussian Mixture Model

Before you can identify a target speaker, an important step must be done. That is training of the models. For every speaker, we need to collect clips that are long enough for training. After training, a set of parameters are obtained. They are the mean, variance and the weighting for the small Gaussian distribution. They are represented by $\lambda = \{p_i, \vec{\mu}_i, \Sigma_i\}$ in which p_i represents the Gaussian mixture weights. $\vec{\mu}_i$ Represents the mean and Σ_i represents the variance. A possible approach to do training is the Maximum Likelihood parameter estimation. The aim of the Maximum Likelihood parameter estimation is to find the model parameters which maximize the likelihood of the training data. Maximum Likelihood parameters can be estimated using a specialized version of expectation-maximization (EM) algorithm (Dempster et al. 1977). The basic idea is that we first begin with an initial model, and then we estimate the new model such that the new model better represents the data. After training, we will obtain the mean, variance, and weighting of each Gaussian distribution (λ).

Steps for Training:

1. First collect a set of sound clips that is long enough for each speaker you want to identified
2. Begin with an initial model λ Then we calculate the $\overline{\lambda}$ (new mean, variance, weighting) for the model by using the following formula.

3. Gaussian mixture:
   Weights:
   $$\overline{p}_i = \frac{1}{T}\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)$$
   Means:
   $$\overline{\vec{\mu}}_i = \frac{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)\vec{x}_t}{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)}$$
   Variances:
   $$\overline{\sigma}_i^2 = \frac{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)x_t^2}{\sum_{t=1}^{T} p(i \mid \vec{x}_t, \lambda)} - \overline{\mu}_i^2$$

4. Check if the newly calculated parameter is more suitable to model the speaker by using the following formula.
   $$p(i \mid \vec{x}_t, \lambda) = \frac{p_i b_i(\vec{x}_t)}{\sum_{k=1}^{M} p_k b_k(\vec{x}_t)}$$
   We use part of the training data to test the parameter $\overline{\lambda}$. If the score $p(X \mid \overline{\lambda})$ is larger than the score $p(X \mid \lambda)$, then we will use the newly calculated parameter $\overline{\lambda}$ to do the training again. We use the new parameter $\overline{\lambda}$ when the following equation is satisfied.
   $$p(X \mid \overline{\lambda}) \geq p(X \mid \lambda)$$

5. Continue to do the training by repeating step (3) and step (4). It has been proven that this method is a convergent approach (Bilmes 1997). In order words, when we repeat to train the parameter $\overline{\lambda}$, we will get the parameter which is more close to the actual parameter for modelling the speaker (the ratio between the difference in the log-likelihood of the current and previous iteration and the log likelihood of the previous iteration is in the order of 10^{-4}. The error between the actual parameter for the model and λ become smaller and smaller through training.

A critical factor in doing the training is that we need to do the initialization of the model parameters before applying the EM algorithm. Actually the EM algorithm can only guaranteed to find a local maximum likelihood model. However, the likelihood equation for the GMM may have several local maxima. As a result, different initialization model may lead to different local maxima. In our case, we have used the well known LBG vector quantization algorithm (Linde et al. 1980) to initialize the EM. We have noticed that 15 iterations are sufficient for the convergence.

3.3 Audio Scene Change Detection

Speech segmentation is the location of boundaries between speakers in the audio stream. A widely adopted algorithm for the audio segmentation is based on the Bayesian Information Criterion (BIC), applied within a sliding variable-size analysis window (Sivakumaran et al. 2001). In (Cettolo and Vescovi 2003), the BIC algorithm is analyzed in detail, and the Pseudo-code of the multiple change detection algorithms was cited.

4-PRELIMINARY RESULTS

For performance evaluation of the identification engine and after preprocessing the speech signal, a 20 mel-cepstrums (the initial coefficient is discarded), 20 LPCC and 20 LAR were calculated from a 20 ms hamming windowed frames progressing at an overlap of 50%. A GMM model of 32 Gaussian components is used. The training and testing speech was collected from 50 speakers in the TIMIT (NIST 1990)) database which is a noise free speech database (53dB SNR) recorded using a high quality microphone sampled at 16 kHz. In TIMIT, each speaker produces 10 sentences, the first 8 sentences were used for training and the last 2 sentences were used for testing. The average length of each sentence is 3 seconds. The obtained correct identification rates are approximately the same the result obtained in (Wildermoth and Paliwal 2003). The complete system shown in Figure 5 was also tested on small subset of approximately 10 minutes mpeg file spoken interleaving by 3 male speakers obtained from some Arabic TV news.

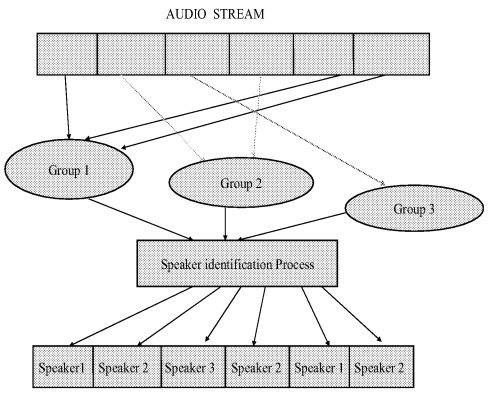


Figure 5: The Segmentation and Classification System

The mpeg files are first transferred to wave files and down-sampled to 22050 Hz, coded by 16 bits. About 15 minutes long speech of each speaker are used to train each speaker GMM model of order 64. In this example the accuracy of indexation is more than 90 %. This promising result encourages us to largely test the system on standard database under different conditions. We are still doing this task.

5-CONCLUSION

In this paper, we presented a system dedicated to audio indexation by identifying and classifying speakers and scene changes. It has been experienced on broadcasted Arabic news with a very promising result (the accuracy was about 90%). We plan in future work to extend the system for other languages by incorporating an engine for language identification (the training and testing must be done in the same language) and to do an extensive comparative study between different features (MFCC, LPCC, LAR) under clean and noisy conditions.

6- REFERENCES

Bilmes. J. 1997. "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models" *Report, University of Berkeley,* ICSI-TR-97-021.

Campbell .J 1997 "Speaker Recognition: A Tutorial", *Proc. Of IEEE,* VOL.85, no.9, Sept.

Cettolo. M. and M. Vescovi. 2003. "Efficient Audio Segmentation Algorithms based on the BIC", *Proc. of ICASSP*, pp. 537-540, Hong Kong, China.

Chen.S. S, and P. S. Gopalakrishnan.1998. "Speaker environment and channel change detection and clustering via the Bayesian Information Criterion," in *Proc. of the DARPA Broadcast News Transcription & Understanding .Workshop*, Lansdowne, VA.

Chow. D, and W. H. Abdullah. 2004. "Speaker Identification Based on Log Area Ratio and Gaussian Mixture Models in Narrow-Band Speech," *Lecture Notes Artifical Intelligence LNAI 3157*, Zhang, C. Guesgen Yeap, W. (Ed.), pp.901-908, Springer. .

Dempster A.P. D, N.M. Laird, and D.B. Rubin .1977."Maximum-likelihood from incomplete data via the em algorithm". *J. Royal Statist. Soc. Ser. B.*, 39.

Linde Y., Buzo A, and Gray, R.1980. "An Algorithm for Vector Quantizer Design." *IEEE Transactions on Communications.* Vol. 28(1), 84-95. (1980) p72-83.

National Institute of Standards and Technology (1990), the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus.

Pravinkumar.P, and B. M. Wasfy .2001. "Speaker Verification/Recognition and the Importance of Selective Feature Extraction: Review", *Proc. IEEE.*

Reynolds, D.A., and R.C. Rose.1995. «Robust text-independent speaker identification using Gaussian mixture speaker models", IEEE Trans on speech and audio processing, Vol 3, No 1, (Jan).

Rosenberg A. E, and F. K. Soong.1986. "Evaluation of a vector quantization talker recognition system in text-independent and text-dependent modes," *in Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing*, pp. 873-880.

Rudasi. L, and S. A. Zahorian.1991. "Text- independent talker identification with neural netowkrs," *Proc. IEEE ICASSP*, pp. 389-392, (May).

Sivakumaran P, J. Fortuna, and A. M. Ariyaeeinia .2001. "On the use of the Bayesian Information Criterion in multiple speaker detection," in *Proc. EUROSPEECH*, Aalborg, Denmark, vol. 2, pp. 795–798.

Tritschler A, and R. Gopinath.1999. "Improved speaker segmentation and segments clustering using the Bayesian Information Criterion," in *Proc. EUROSPEECH*, Budapest, Hungary, , vol. 2, pp. 679–682.

Wildermoth B, and K.K. Paliwal.2003. ``GMM based speaker recognition on readily available databases", *Proc. Microelectronic Engineering Research Conference*, Brisbane, Australia, (Nov).

Using a trace-based simulation tool to test the efficiency of error codes applied to wireless multimedia streaming

Julien Bourgeois, Adrien Henriet
Laboratoire d'Informatique de l'Université de Franche-Comté (LIFC)
Centre de Développement Multimédia
25201 Montbéliard Cedex, France
{Julien.Bourgeois, Adrien.Henriet}@univ-fcomte.fr

Jerome Lacan, Tanguy Pérennou
ENSICA-LAAS,
1, place Emile Blouin, 31056 Toulouse cedex 5
{jerome.lacan, tanguy.perennou}@ensica.fr

Abstract

This article presents an original method to validate the efficiency of the short length maximum separable distance forward error codes when using broadcasting mode in 802.11b/g multimedia streaming. Our method is based on real traces on which the forward error code will be mapped. Then, the quality of the streaming is evaluated by calculating the Peak Signal-to-Noise Ratio (PSNR).

KEYWORDS
Multimedia, FEC, error codes, simulation, wireless, streaming

Introduction

The development of wireless networks bandwidth is changing the way to access to multimedia contents. WiFi, also known as 802.11b/g, is the most popular local wireless network. WiFi can communicate according to two modes: point to point and broadcast. Point to point communications offer reliability by ensuring retransmission of lost packets. The lost packets retransmission is managed by the MAC layer of 802.11b/g. When using broadcasting mode, the MAC layer does not manage retransmission of lost packets. Therefore, in a video streaming session, errors will occur and video quality will decrease. In this article, we propose to evaluate the benefit of using adaptive Forward Error Code (FEC) by calculating the Peak Signal-to-Noise Ratio (PSNR) of the transmitted videos. This experiment based on real traces ensures a result very close to the reality.
In the following section , we present conventional error control mechanisms. Section , discusses the efficiency of short-length MDS FEC. This efficiency is calculated using real experiments that generate traces. Simulations use these traces to obtain results. Section concludes the article and describes future work to be done.

Conventional Error Control Mechanisms

In order to ensure packets transmission reliability, two essential types of mechanisms can be distinguished [3] :

- reactive mechanisms (ARQ: Automatic Repeat request): when a packet is lost, the transmitter retransmits it. The packet loss indication can be carried out by the emission of positive or negative acknowledgments.
- proactive mechanisms (FEC: Forward Error Correction): the transmitter sends redundancy packets *a priori,* making it possible for the receiver to recover certain lost packets thanks to the redundancy packets. These packets are usually calculated by using block error correcting codes. These codes generate $n - k$ redundancy packets starting from a set of k information packets (to be protected) so that the receiver can recover the k information packets as soon as it has received a sufficient number of packets among all the sent packets.

It should be pointed out that these two mechanisms are often combined by using acknowledgments to adjust the quantity of redundancy of FEC codes. Such a combination is often called Hybrid ARQ of type I (packets already received by the receiver are not combined with redundancy packets received after the retransmission request) or of type II (all packets that relate to the same block of k packets are used by the decoder). Being relatively flexible and easy to implement, reactive mechanisms are the most used in packet transmissions (in TCP for example). Proactive mechanisms are particularly used in multicast transmissions (where a redundancy packet allows to compensate for the loss of a different packet for each of the receivers). In addition, real-time transmissions which do not allow delays due to retransmissions, often use proactive mechanisms.
In the context considered here, i.e. real-time multicast video transmissions, the use of FEC is clearly the best solution.
Two main families of FEC can be distinguished: the Maximum Separable Distance (MDS) codes and the Low-Density Parity-Check (LDPC) ones. The MDS codes, often based on Reed-Solomon codes, have the property that the information packets can be recovered as soon as any k packets among the n sent ones are received. However, for long lengths ($n > 500$), these codes, which have quadratic encoding and decoding complexities, become too costly in CPU. On the other hand, LDPC codes have linear complexity and then can be used with long lengths [4, 7]. However, they need more received packets (between 5 and 20% [6]) than MDS ones to recover the information packets, especially for short codes.
A strong constraint on FEC used for real-time transmissions is that they must have a short length. Indeed, the decoder

must necessarily wait the reception of the last packets to begin the decoding. Then it appears that the FEC codes which must be considered here are short-length MDS codes. Next Sections present an evaluation of these codes.

Evaluating the short-length MDS FEC efficiency

To evaluate the efficiency of FEC codes, and extremely important point is to consider a realistic representation of the channel in terms of packet loss rate and losses distribution. Even if some modeling of the 802.11 channel losses were already proposed (see e.g. [8]), we preferred consider real traces to evaluate the codes.

Experiments

A platform of wireless tests was installed in order to collect IP broadcast packets loss measurements. It relies on an 802.11b infrastructure mode network, which comprises one access point, 10 nodes with an 802.11b interface, and one node with an Ethernet interface, directly connected to the access point. Various types of 802.11b interfaces (PCI, USB and PCMCIA) and material architectures (PC, laptop PC, Pocket PC) were used.

Methodology
First, a serie of 5 experiments was carried out during which all the nodes remained motionless. The nodes were placed at various places, with a distance of 0 to 4 walls with the access point. In a second series of 6 experiments, 4 of the nodes were mobile: each one was carried by an operator walking without a specified trajectory while always remaining in the range of the access point.
In each experiment, the fixed node `pacman` transmits a multicast MPEG1/RTP video-only stream, which consisted of 300 to 1400 bytes packets with a mean bandwidth of 500kbps. Each experiment is approximately 90 seconds long. The video stream is produced by the Java Media Studio application and is broadcasted in multicast on the wireless network. The transmitter is fixed in all of the experiments and remains close to the access point, with a minimal attenuation of the signal on the corresponding link. The resulting MAC packets are automatically re-emitted by the access point towards 9 nodes over the wireless network, and 1 node over the Ethernet.
Thus, each experiment made it possible to record the sequence numbers of the RTP packets actually received by each receiver. In 802.11b, there is no acknowledgement nor retransmission in the MAC layer for broadcasted packets [1]. Hence, a received UDP packet was transmitted only once from the transmitter towards the access point and then towards the receivers. By making the assumption that no packet is lost on the Ethernet, the Ethernet receiver `serveur-mitv` allows the measurement of MAC losses over the uplink between the transmitter and the access point, the wireless receivers allowing to measure the additional MAC losses over the downlink.

Results

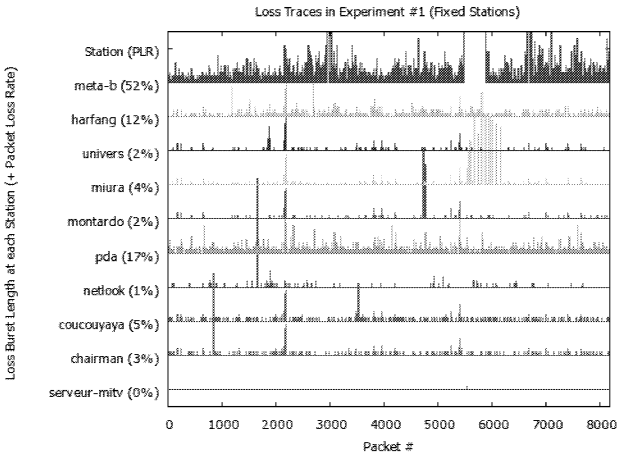


Figure 1. Traces of packets loss with fixed stations

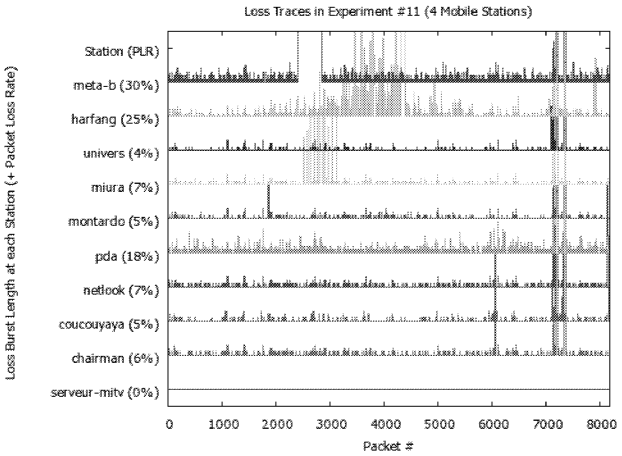


Figure 2. Traces of packets loss with mobiles stations

The 5 experiments, where all the nodes were fixed, produced similar traces, just as the 6 experiments with mobile nodes. We can see here that the mobility isn't a significant factor who influate in packet lost. Figures 1 and 2 gives the traces of one of the experiments with fixed stations and one with mobile stations (`pda`, `miura`, `montardo` and `harfang`). The x-axis corresponds to the sequence of transmitted packets and the y-axis illustrates the burstiness of packet losses for each station. The receiving stations are listed on the left. The percentage given after the station name is the packet loss rate observed for this station in this experiment. Each line indicates if packet #x was received (lowest position) or lost (higher positions) by the corresponding station. Loss

bursts are represented by their length along the y-axis. To get an idea of the length of the bursts, consider that station loss lines are 10 losses apart. In experiment #1 for example, `serveur-mitv` does not lose any packet, which results in a straight line. In contrast `netlook` experiences a big burst of 32 packets starting at packet #1645 and a few other less important bursts: this results in an almost straight line with a high peak for the 32-packet long burst and a few smaller peaks corresponding to bursts of 1, 2, 3 or 5 packets. Finally, `pda` experiences a lot more loss bursts (255 bursts have a length greater than 1, generally 2, 3, 4, or 5 packets long), which results in a very irregular line. The ten receiving stations are classified on the y-axis from the nearest to the most distant with respect to the access point, except for `meta-b` which is close to the access point but receives few packets due to hardware limitations.
The experiments show that practically no data is lost on the uplink, and that many losses are observed on the downlink. They are related to the used material, the position of the fixed receiver or to mobility. Some simultaneous bursts attest the occurrence of important disturbances on the downlink, which could not be diagnosed. The peaks of very important amplitude recorded on `meta-b` reveal a disconnection phase due to a hardware configuration problem.
More generally, a considerable quantity of packets are lost in bursts, which can be temporally correlated for a given receiver and/or spatially correlated for several receivers. The peaks observed for each individual receiver of Figure 2 denote temporal correlation. Spatial correlation can be observed too when peaks occur for the same packets, as in e.g. experiment #1 for packets #2168 to #2088, which are simultaneously lost by all stations except `serveur-mitv` and `netlook`. An analysis of these correlations can be found in [5].

Simulation driven by traces

To evaluate the efficiency of the FEC codes applied to a video sequence, one solution is to count the ratio between well-received and losts packets. If this solution can show the theoritical efficiency of the FEC codes, it does not show the real efficiency of the FEC codes applied to a video sequence. The idea is to calculate the PSNR of the received video. Indeed of the methodology, and the type of trace, we can't know the size of each packet, and so, can't have a more realistic simulation based not only on packet, but on data. This isn't a limitation, because of the structure is the same, and be relatively near a real situation.

Methodology
As described in figure 3, our simulation tool takes two inputs: a movie and the experiment trace file described in section . The movie is the Foreman, a reference sequence for testing multimedia platform. The aim of our tool is to compute the PSNR of the movie when packets are lost according to the trace file. This PSNR represents the available quality for the client for each decoded frame of the stream. Indeed, during multimedia streaming the lost packets will not have the same effect on the video quality. Some packets will seriously affect the visual quality of the video whereas others will not. This difference depends on various parameters like the kind of lost packet, the moment in the group of pictures (GOP) the packet is lost, etc. In fact, if the last P-frame of a GOP is lost, the quality will not be affected because, just after this image, an intra image will be decoded. As the last P-frame of a GOP and the first I-frame of the following GOP do not have any temporal dependances, only one frame will be damaged. An another example is when a packet is lost on an P-frame just before a camera movement. The resulting image will be damaged but the camera movement will delete this error. That is why it is necessary to calculate the PSNR and not just to count the lost packets to evaluate the resulting video quality.
Our simulation tool can compute the PSNR with or without the short-length MDS FEC. In fact, it can apply any kind of processing on the packets. Our simulation tool is composed of two main boxes: the Simulation Box and the PSNR Box.
The Simulation Box is composed of mplayer an open-source movie player and the Damager which is able to alter the packets transmitted by mplayer. The Damager does not randomly alter the packets, it takes as input the trace file generated in section . When a packet is generated in mplayer, the Damager alters it or not according to the trace file. It can also take an error code information into account. Here, we use our short-length MDS FEC as a plugin for the Damager. This plugin modifies the trace file by correcting the lost packets when it is possible.
The PSNR Box receives only images, not packets. MPlayer receives packets and generates Portable Network Graphic (PNG) images. These images are then gathered by the PSNR Box which converts them into a raw format and then computes the PSNR. The output of our simulation tool is a text file which contains the PSNR for each generated image.

Results
The three figures 4, 5 and 6 show the PSNR of the movie Foreman visualized respectively by the client "coucouyaya", "meta-b" and "harfang" in the mobile environment. Strong or light FECs represent the minimum size of the redundancy information in the code. The solid curve represents the PSNR of video sequence without any errors. The dotted curve shows the PSNR calculated when strong FECs are applied to the packets whereas the dashed curve presents the PSNR calculated when light FECs are applied to the packets. The dotted-dashed curve shows the PSNR of the transmission without error codes.
"Coucouyaya" receives the packets with few errors, that is why the three curves, PSNR with strong FECs, light FECs and without FECs are very close. The interesting thing to see is that the use of strong codes or light codes when there are no long bursts of errors is nearly equivalent.
"Meta-b" and "harfang" have nearly the same error profile: lots of errors, and long bursts of errors. For these station, it is clear that the FECs enhance the visual quality of the video.

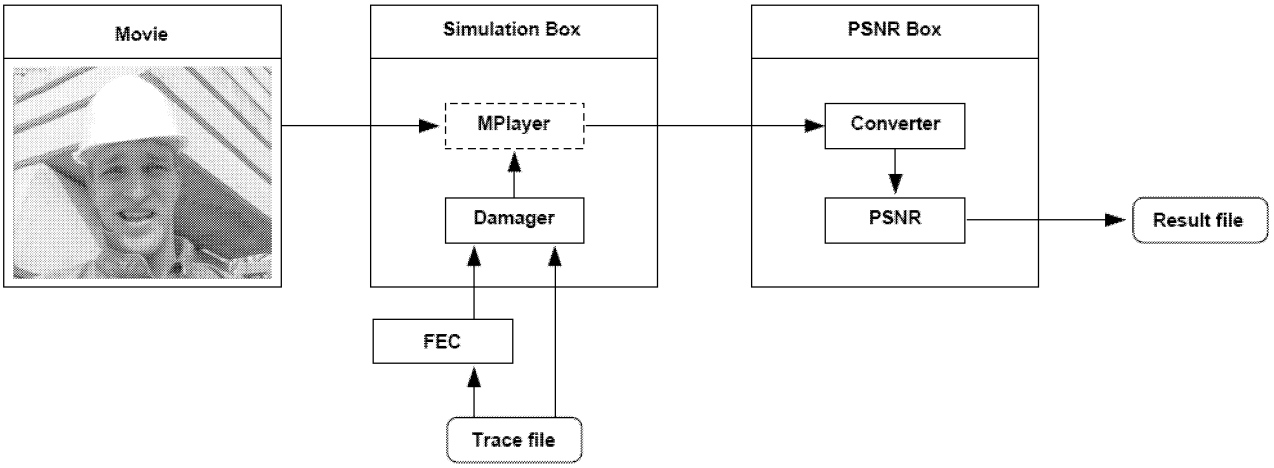


Figure 3. The simulation tool

But, when the bursts are too long, even the strong FECs can not correct the errors and retreiving the original content of the packets. This, is the second interesting fact taught by these experiments.
Thanks to these results, we can say that even strong FECs can not cope with long bursts as the stream will be interrupted. Strong FECs delay the interruption but do not avoid it. The second fact is that when there are no long bursts, light FECs are sufficient. As the strong codes need a lot more bandwidth than the light ones, the idea is to preferably use them. But, the problem of long bursts remains. Solutions will be investigated later.

Conlusion and perpectives

We have presented our method to test the efficiency of the short length maximum separable distance forward error codes. Our simulation tool is able to calculate the quality of any streamed video by using trace files of previous transmission. The results show that the FECs can not cope with long bursts of error. They can delay the error but at last, the video quality will dramatically decrease.
Future works will have to deal with this problem. The idea we will have to experiment is the following. Light FECs are kept to deal with short bursts of error and an RTP mixer [2] capable of regenerating an I-frame is adding between the server and the clients to overcome the long bursts of error. When a interruption occurs, the RTP mixer will transform a P-Frame into a I-Frame, and then the video can continue thanks to this new GOP. With this mixed solution (reactive and proactive), the bigger part of the errors will be treated without adding too much traffic on the wireless channel.

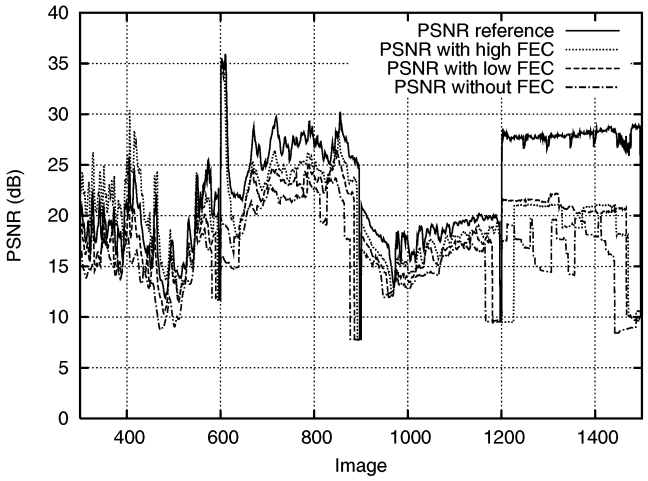


Figure 4. PSNR computation for coucouyaya

REFERENCES

[1] ANSI/IEEE Std 802.11. Wireless LAN MAC and PHY Specifications, 1999.

[2] J. Bourgeois, E. Mory, and F. Spies. Video transmission adaptation on mobile devices. *Journal of Systems Architecture*, 49(10-11):475–484, November 2003.

[3] S. Lin and D. Costello, Jr. *Error Control Coding: Fundamentals and Applications*. Prentice Hall, Englewood Cliffs, NJ., 1983.

[4] M. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman. Efficient erasure correcting codes. *IEEE Transactions on Information Theory*, 47(2):569–584, 2001.

[5] T. Perennou, J. Lacan, and H. Elnatour. évaluation de mécanismes de contrôle d'erreur pour des transmissions multipoints sur des réseaux de mobiles. *TSI*, To appear in 2005.

[6] J. S. Plank and M. G. Thomason. A practical analysis of low-density parity-check erasure codes for wide-area storage applications. In *DSN-2004: The International Conference on Dependable Systems and Networks*. IEEE, June 2004.

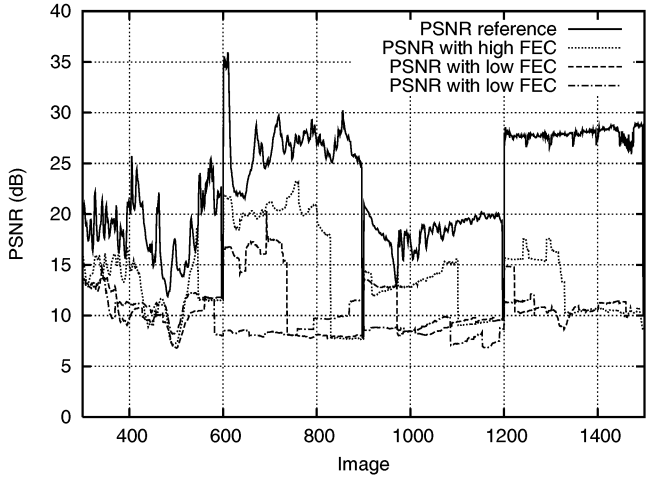


Figure 5. PSNR computation for metab

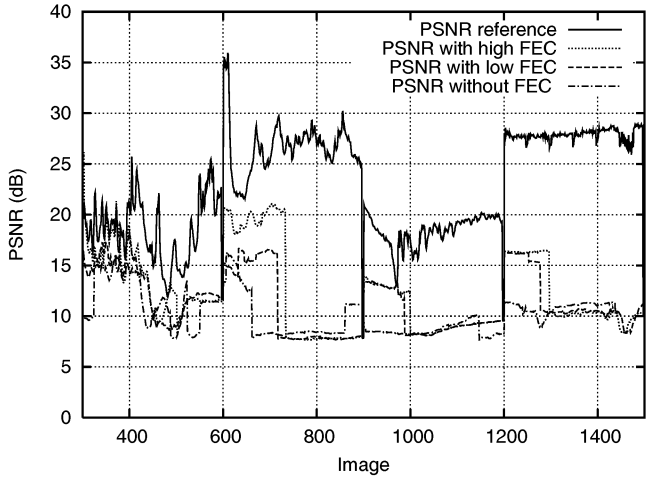


Figure 6. PSNR computation for harfang

[7] V. Roca, Z. Khallouf, and J. Laboure. Design and evaluation of a low density generator matrix (ldgm) large block fec codec. In *Fifth International Workshop on Networked Group Communication (NGC'03), Munich, Germany*, September 2003.

[8] C. Tang and P. McKinley. Modeling Multicast Packet Losses in Wireless LANs. In *Proceedings of MSWiM'03*, 2003.

[9] Mplayer: open source movie player for Linux http://www.mplayerhq.hu.

DISTRIBUTED OPENED CROSS-MEDIA INDEXING PLATFORM

Bassem HAIDAR, Philippe JOLY, Jean-Paul BAHSOUN

IRIT, Institut de Recherche en Informatique de Toulouse
118, route de Narbonne - 31062 Toulouse Cedex 04
France
{haidar, joly, bahsoun}@irit.fr

ABSTRACT

Over the last decade, many research projects have focused on algorithms for segmentation and indexing of multimedia contents so that dedicated system architectures allowing the cooperation of multiple multimedia indexing tools are now a necessity. Pidot (Distributed Opened Cross-media Indexing Platform) is a CNRS project that assembles several competences in the domain of multimedia indexing. The goal of this project is to share multimedia indexing resources between various laboratories, while handing the easy integration of new tools.
For the Pidot architecture, we propose a central server, and multiple distributed multimedia indexing services. The central server allows the control and collaboration between analysis tools which are accessed as web services. The communication is based on SOAP-RPC [1] calls.

INTRODUCTION

Several research teams are nowadays working on multimedia segmentation and indexing tools. Each of them is generally specialized to address one single media. These tools are quite always developed independently, and this introduces a lack of interoperability which is a strong limitation for cross-media analysis. The project Pidot aims at proposing a framework to share multimedia indexing resources between several laboratories and to address new research topics in the domain. It can be, for example, a support to work on the fusion or the aggregation of multiple tools in order to produce improved results. The project framework can also be used to implements methods for benchmarks in order to evaluate multimedia indexing tools.
The international scientific community encourages evolutive projects that allow a diversity of evaluation mechanisms in automatic segmentation, indexing, and content-based retrieval of digital video contents, like OPEN VIDEO [2] and TREC VIDEO [3]. Among projects where the system architecture received a really specific attention, we can cite:
- The project KLIMT (Knowledge InterMediation Technologies) [4] proposed an open framework for the integration of multimedia indexing tools. The framework implemented communication and data transfer mechanisms which used to integrate a whole chain of multimedia indexing tools. The chaining process is customized for a predefined set of scenarios.
- The project AGIR [5] aimed at developing technologies and software tools allowing efficient multimedia document search systems based on the automated analysis of media content and on a normalized content description language MPEG7 [6].
- The Mediaworks project [7] used cooperation between information from text and image in order to develop a system that assists documentalists to index TV archives.
- The project FERIA (Framework for Indusdrial Realization and Experimentation of multimedia Applications) [8] aims at developing a general and open framework allowing the easy development of dedicated industrial applications to edit interactive video documents.
- The large-scale multimedia indexing project, presented in [9], proposed a distributed architecture for media acquisition, content processing, and document indexing. The main goal of the project is to exploit the parallel execution in order to distribute the work load.
The Pidot project is a CNRS project that assembles various skills. Project partners are specialized in a specific domain of multimedia indexing and information retrieval so that those skills are distributed between five laboratories:

- The MRIM (Modeling and Retrieval of Multimedia Information) team [10] of CLIPS-IMAG.
- The Multimedia communication department [11] of EURECOM.
- The SAMOVA (Audio and Video Analysis, Structuring and Modelling) team [12] of IRIT.
- The Image and Sound, Video analysis and indexing team [13] of LABRI.
- The APA (Machine Learning and Knowledge Acquisition) group [14] of LIP6.

The goal of the project is to build a distributed indexing platform allowing to:

- Share indexing resources (corpus, metadata, indexing tools) on a same platform between several laboratories,

- Provide an experimental framework to develop and evaluate new tools of fusion and aggregation of audiovisual indexes, and to
- Provide more flexibility in the collaboration between multimedia indexing tools.

GLOBAL ARCHITECTURE

The platform is composed of distributed services and a central server that implements two principal services: the repository service and the access service (see Figure 1).

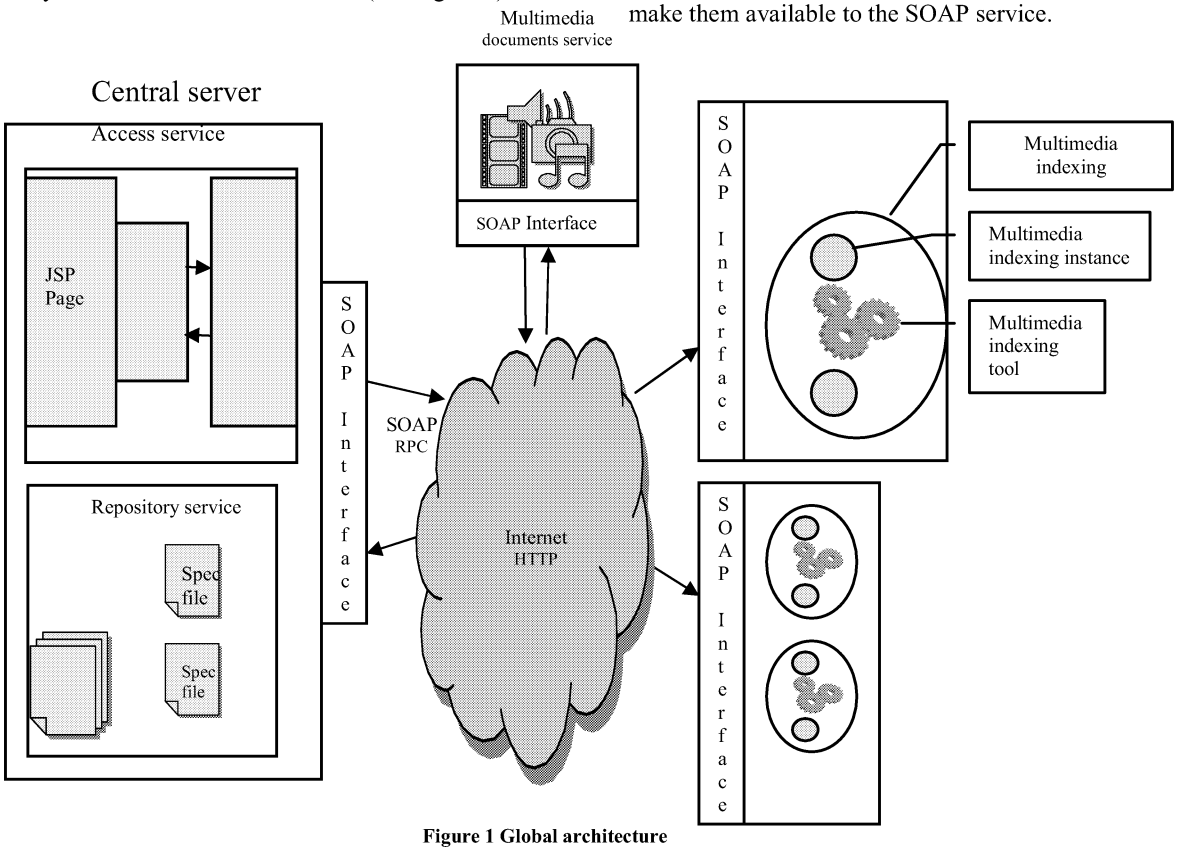


Figure 1 Global architecture

Communication details

We use the SOAP [1] protocol for the RPC (Remote Procedure Call), SOAP (Simple Object Access Protocol) has been developed by Microsoft, IBM, DevelopMentor and UserLand Software. SOAP is a protocol for remote procedure calling and messaging with XML-encoded application data. It presents the advantage of posting data over the HTTP protocol. A SOAP message is written using the XML formalism and is sent over HTTP to a remote server. On the Web service side, the SOAP request is processed and the result is sent back to the client using a SOAP response.

We use Apache Axis [15] for development. Axis was developed by IBM alphaWorks. It provides SOAP support for Apache Tomcat application servers. Apache Axis use SAX for XML parsing.

We use a DataHandler [16] object to transfer the multimedia content. With this DataHandler, the Java Activation Framework provides a serialization and deserialization functionality and attaches it automatically to the sent SOAP message. The data file is not directly integrated in the SOAP envelope, but as a reference to an attachment. The entire SOAP request is sent as a multipart MIME-encoded request.

The first part is the SOAP envelope, and the second part is the attachment. This is basically the same format that allows sending attachments in e-mail messages. When the server receives the multipart request, it automatically understands how to find the referenced attachments and make them available to the SOAP service.

With the previous mechanisms, multiple files can be sent simultaneously by using a table of DataHandler objects, each of them handling the required file.

Transferring data via SOAP can not be achieved for large amounts of data; in a further step of development, we will integrate FTP mechanisms to allow the transfer of large documents such as full-length videos for example.

Opened platform, tool wrapper

A multimedia indexing tool is in general an executable program characterized by a specification file written by the owner or the developer of the tool. The specification file contains information needed to execute this tool like the number and the type of input data, the number and the type of the output data, options needed for the execution, etc.

We have developed a wrapper, which is a package of code able to read the specification file, and then to launch the tool with the required command line which looks like:

```
Prog_name option1 imput1 input2 output1 output2
```

The following example is a specification of an audio segmentation tool:

```
<Component ComponentRole="Index">
<Description>To localize segments of Speech, Music,
Noises and Silences in an audio file</Description>
<Authors>J.Pinquier</Authors>
<Version>v123.2b</Version>
<Input>
  <InputData>
   <input_type>type1</input_type>
   <FileFormat>wav</FileFormat>
   <StorageInputPath>/Input_Path</StorageInputPath>
  </InputData>
</Input>
<Output>
 <OutputData>
  <output_type> AudioSegmentContent</output_type>
  <FileFormat>xml</FileFormat>
  <StorageOutputPath>/output_Path</StorageOutputPath>
 </OutputData>
</Output>
<Options>
<Option>option1</Option>
</Options>
</Component>
```

Figure 2 specification file of an audio segmentation tool

This generic wrapper is independent from the type of the tool, and only the specification file is required.

Tool invocation

The generic wrapper allows the invocation of multimedia indexing tool. Using its specification file, the wrapper builds the command line required to launch the tool. A RPC call *launches (DataHandler[] data)* the transfer of input files thanks to a DataHandler table. An example of this table is explained in the following figure:

Row number	Value
1	Table structure
2	serialised file DataHandler[2]
3	serialised file DataHandler[3]
4	serialised file DataHandler[4]

Figure 3 DataHandler table

The first element in the table specifies the structure of the table. For example, according to the specification given in Figure 4, the second element in the table in the Figure 3 represent the first input parameter (param_in1), the third element represent the second input parameter etc…

Structure	
Parameter Id	Row number
Param_in1	2
Param_in2	3
Param_in3	4
Option_1	-Option1
Option_2	-Option2
….	…..

Figure 4 Table structure

Using the information from the specification file (number of input files, input data format, number of options,…), the wrapper read the table and save input files in order to launch the required command line able to execute the tool. Output data files are also returned using a DataHandler table.

CENTRALISED SERVICES

Repository service

The repository service allows the registration of new services on the platform. A new tool will be integrated on the platform by sending towards the repository service its specification file. The information contained in this file will be available for the other services connected to the platform. The repository service allows a tool, defined by a set of attributes, to be retrieved by a searching mechanism. This mechanism browses the information located in all the specification files.

Semantic types are inspired from the ones defined by the MPEG-7[6] standard when it is possible. Other semantic types have to be added to provide pieces of description which are not available in MPEG-7[6]. We implement some XML research functions in order to browse the XML specifications.

Access service

The access service allows a centralised access to the platform; its interface provides all the functionality of the platform:

- We can search for a given service using different parameters like its output or its input data type, or its name.
- We can search for available multimedia documents.
- We can launch a service by selecting this service and its required input data.

The access service is in fact a web service; its interface implements the previous functionalities. JSP (Java Server Page) is used to instantiate a client for the access web service. This client allows to dynamically load indexing and documents services available on the platform (Figure 5).

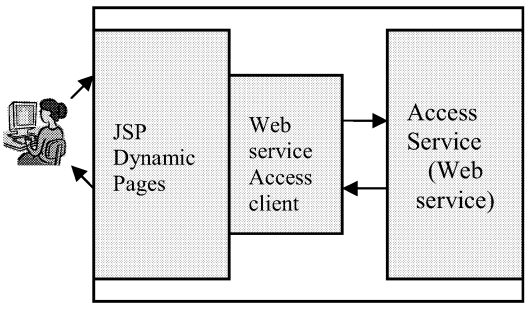


Figure 5 Access Service

The following figure explains the scenario of launching a service via the Access Service.

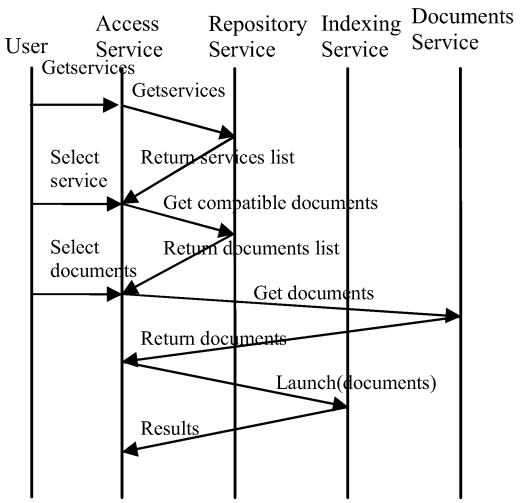


Figure 6 Scenario of launching a service via the Access service

The scenario of launching a service via the access service is illsutrated by Figure 5:
The user asks the Access service for available services on the platform, the access service answers by displaying this list of services provided by the repository service. The user selects the desired service, and then the access service answers by displaying the available compatible documents registered in the repository service. The user selects relevant documents, and then the access service downloads the documents from the multimedia document service. Finally the access service sends the documents to the selected indexing service. This last one processes the documents and returns the results to the access service.

DISTRIBUTED SERVICES

Multimedia indexing services

Indexing service

The goal of the platform is to allow the access of distributed multimedia indexing tools, an indexing service can interface one or several indexing tools, developed by a single research team or specialized in a single media analysis like audio segmentation, or speech transcription. The indexing service must allow a user or another indexing service to use one of the handled indexing tools. Each indexing service is identified by a service identifier by the server side, and contains a table of "indexing tool representation" objects. Services have a SOAP-RPC interface that implements the service functionality.

Indexing tool representation

An indexing tool representation is the representation of one indexing tool of the indexing service; it is created by the central server. More precisely, the scenario of adding a new tool on the platform is the following one:
The owner or the developer of the tool asks the repository service to register his tool, by sending towards the repository service the specification, and the address of the indexing service that will host the tool. Then the central server creates an identifier for this tool, contacts the specified service, and creates an indexing tool representation. The indexing tool representation will be stored in a table by the indexing service side.

Indexing tool instance

An indexing tool may be used by different users simultaneously. So, for each user, we create a specific indexing tool instance that allows several executions of a same indexing tool by different users. An indexing tool instance has an identifier which is associated with a session number. The indexing tool instance implements the functionality of the generic wrapper. It allows to launch the executable tools and to get the output results.

Multimedia documents service

A user can add a multimedia document access service to the platform. As a multimedia indexing service, the multimedia documents service allows to access to multimedia contents, and has a specification file that describes the type of contents. A SOAP interface is defined to transfer data from and to this service. We use DataHandler object to transfer multimedia documents.

AUTOMATIC INDEXING

The Pidot platform can be used for automatic indexing consisting in automatically searching, chaining and executing tools able to generate a given index. A recursive algorithm is proposed in [17]. Based on the input output compatibility, the chaining algorithm builds a workflow able to generate the expected index. We also proposed a distributed execution scenario, that allows to reduce the network charges by establishing the direct communication between distributed services, avoiding so useless data transfer to the central server.

CONCLUSION

In this paper, we have presented the main mechanisms implemented on the Pidot platform to handle distributed indexing services.
The pidot project is intended to be an experimental scientific instrument allowing to share expensive resources (such as ground truth data sets), to compare results of competitive analysis tools, and to develop and evaluate new algorithms for example on the fusion and the aggregation of audiovisual meta-data. An example of metadata fusion is the multimodal person identification in audiovisual documents, using audio and video cues. It will be used also to work on story segmentation using the fusion of multiple results produced by different segmentation algorithms.
This platform development is on progress and a set of different potential evolutions are already identified such as the automatic choice of the most reliable chain of services to produce a given index for example.

REFERENCES

[1] SOAP : Simple Object Access Protocol 1.1 HTTP://WWW.W3.ORG/TR/SOAP/

[2] The Open Video project : http://www.open-video.org/

[3] TREC Video : http://www-nlpir.nist.gov/projects/trecvid/

[4] V. Conan, I. Ferrané, P. Joly, C. Vasserot. KLIMT: Intermediations Technologies and Multimedia Indexing. In Third International Workshop on Content-Based Multimedia Indexing (CBMI'03), Rennes, France, septembre 2003. INRIA, 11-18.

[5] AGIR : Architecture Globale pour l'Indexation et la Recherche par le contenu de données multimédia. http://www.ina.fr/recherche/projets/finis/agir.fr.htm

[6] MPEG-7: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[7] Mediaworks Project: http://www.inria.fr/rapportsactivite/RA2003/texmex/module15.html

[8] FERIA : Framework for Indusdrial Realization and Experimentation of multimedia Applications. http://www.ina.fr/recherche/projets/encours/feria/

[9] H. Mandviwala, S. Blackwell, and JM. Van Thong. A Scalable Multimedia Content Analysis and Indexing System Architecture. Third International Workshop on Content-Based Multimedia Indexing, Sept. 22-24, 2003, Rennes, France. 349-355.

[10] http://www-mrim.imag.fr/

[11] http://www.eurecom.fr/Multimedia/

[12] http://www.irit.fr/recherches/SAMOVA/

[13] http://www.labri.fr/Recherche/ImageSon/index-en.htm

[14] http://www-apa.lip6.fr/English/indexAng.htm

[15] Axis: http://xml.apache.org/axis

[16] http://java.sun.com/j2ee/1.4/docs/api/javax/activation/DataHandler.html

[17] B. Haidar, P. Joly, S. Haidar, A graph based approach to automatically chain distributed multimedia indexing services. To appear in the 9th IASTED International Conference on INTERNET AND MULTIMEDIA SYSTEMS AND APPLICATIONS, EuroIMSA 2005.

Adaptive multimodal communication for multimedia

Cezar Pleşca, Romulus Grigoraş, Philippe Queinnec, Gérard Padiou
IRIT UMR CNRS 5505, ENSEEIHT-Informatique
2 rue Camichel, Toulouse, France
e-mail: {plesca, grig, padiou, queinnec}@enseeiht.fr

KEYWORDS
multimodal communication, adaptive multimedia, distributed systems, streams, causality, partial-order

Abstract

Various distributed multimedia applications need to offer the user a consistent view of the interactions represented by the streams exchanged between dispersed groups of users. Ajay Kshemkalyani has studied temporal interactions of streams by means of intervals and we have proposed an application to distributed multimedia systems. Using strong ordering such as CAUSAL or TOTAL order to capture and deliver streams interactions may be too expensive (in terms of delay) due to the variability of the network conditions. This paper builds upon previous works on expressing streams causality and propose a multimodal communication service in order to integrate different delivery modes (e.g. FIFO, CAUSAL, TOTAL) into a single multimodal channel (with respect to each of these protocols). Moreover, the proposed multimodal channel can handle the mix of any partial or total order protocols. We build an experimental platform for measuring the cost of combining several protocols on various network conditions and show how our communication service can adjust the degree of coherency according to network variations.

INTRODUCTION

The main topic in distributed multimedia systems is how to transfer multimedia data with the demanded quality. One important requirement of distributed multimedia applications is the need for sophisticated quality of service (QoS) management.

QoS parameters can be partitioned into two subsets, namely application-dependent parameters and application-independent parameters. They can also be put into three layers: application, system and network layers as shown in table 1.

QoS Layer	QoS Parameters
Application	Frame Rate, Frame Size Response Time
System	Buffer Size, Process Priority Scheduling Policy
Network	Bandwidth, End-to-End Delay Delay Jitter

Table 1. Quality of Service Layers

Application parameters describe requirements for application services and are specified in terms of media quality and media relations. Media quality includes source/destination characteristics such as frame rate and transmission characteristics such as response time. Media relations specify relationships among media, such intra-stream synchronization (i.e. maintenance of real-time constraints across a single continuous media connection) and inter-stream synchronization (between different media types).

An important issue in the study of media relations is event ordering. Research on this topic deals with the message delivery process with respect to the order of their emissions. Several delivery policies (i.e. Best-Effort, FIFO, CAUSAL, TOTAL) have been defined to express various application requirements.

As multimedia applications are loss tolerant, we propose a FIFOPR (FIFO Partially Reliable) protocol able to deliver messages with respect to the FIFO policy, with a bounded average loss over a sliding window. Critical messages can be sent with 100% reliability. Offering an intermediate QoS between Best-Effort and FIFO, FIFOPR could reduce latency (by diminishing the number of request retransmissions) across a transmission of a multimedia stream, with no (or little) visible impact over the stream quality experienced by users.

A further requirement is to support multiparty communications. Many distributed multimedia applications are concerned with interactions between dispersed groups of users, for example, a remote conference or a

CSCW (Computer Supported Cooperative Work) [4, 5] application. Multiparty communication also adds complexity to synchronization in general. It is important to be able to support a variety of policies for ordering data delivery, for example, FIFO, CAUSAL or other partial orders and even TOTAL ordering. These application level protocols use multicast/unicast transport protocols (UDP/TCP).
The variability of the network conditions affect in a negative way the performance of strong protocols such as CAUSAL or TOTAL. In this paper, we propose an adaptive mechanism that is able to finely adjust the level of coherence (and thus the responsiveness of the communication) to the network conditions. We build upon previous works on expressing streams causality and propose an experimental platform for measuring the cost of combining several protocols on various network conditions. We will also show how our communication service can adjust the degree of coherency according to network conditions.

A COHERENCY PROBLEM

Stream communication raises consistency problems when ordering the stream delivery to a receiving site. Cooperative multimedia applications need to offer the user a consistent view of the interactions represented by the streams.

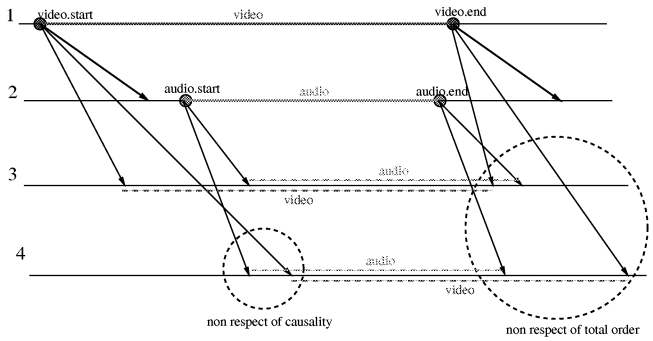


Figure 1. Consistency of Streams Perception

For example, during a videoconference, a video stream (e.g. representing a question) is sent to four participants. One of the participants starts to comment the video stream, sending an audio stream (an answer) that is multicasted to the others. According to [3], these two streams are in a causal relation : the cause of the audio stream is the visualization of the video stream. Thus, all the participants should see the video stream (the question) before they hear the audio stream (the answer). The figure 1 shows the consistency problem on the sites 3 and 4:

- user 4 hears the audio stream before the video stream (that is, user 4 hears the answer before the question). This is a causal inconsistency since the delivery of the video stream starts **after** the delivery of the audio stream, whereas at emission, on site 2, the video stream is received **before** the starting of the audio stream (i.e. the order of streams beginnings is causally reversed on site 4).
- the audio stream ends before the video stream for user 3, and after the video stream for user 4. This is a total order inconsistency: even if the ends of video and audio streams are independent, their deliveries are inversed on sites 3 and 4.

The application should reorder the delivery of the streams according to the causality relations. Simple protocols widely used nowadays in multimedia applications (e.g. RTP, providing FIFO based of timestamps) only provide point-to-point ordering and are unable to capture interactions specific to multiparty cooperative applications (e.g. causality). Multimedia research has been concentrated on synchronisation of complex multimedia presentations composed of various interdependent streams. However, little effort has been made on synchronisation (coherency) according to relations that are not known beforehand, but occuring only while distributed users interact by exchanging streams.

PREVIOUS WORK

Modelling Causality for Distributed Streams Temporal interactions of streams seen as intervals have been studied in [6]. An application to distributed multimedia systems has been proposed in [3]. The authors introduced a model for distributed computations using stream communications and formalized various types of causality between streams by using the concept of stream run. These causal relations can be verified by means of classic timestamping techniques (e.g. Fidge-Mattern clock vectors [8, 9]) or by using a multimodal communication approach.

Merging partial orders Abstract multimodal channels were introduced in [1]. Multimedia distributed applications make use of a mix of various protocols. A FIFO order protocol ensures that the messages are delivered in the same order as they are sent between any two participants. A CAUSAL order protocol ensures

the coordination of group discussion by preventing inconsistencies that may occur when responses are received before questions. A TOTAL order protocol ensures that all participants in a group collaboration receive all the messages in the same order. Cooperative applications need to easily combine these protocols and the article proposes an integration of different delivery modes (e.g. Best-Effort, FIFO, CAUSAL, TOTAL) into a single (abstract) channel, with respect to each of these protocols.

ABSTRACT MULTIMODAL CHANNEL

A multimodal abstract channel provides a simple interface for exchanging messages through various elementary channels.
Each elementary channel (implementation of a protocol) maintains a local control information and provides **send** and **receive** primitives. Upon a **send** call by the application, the protocol builds the control information to be sent with the message, effectively sends the message and updates its local control information. Upon reception of a message, the protocol tests the control information attached to the message with respect to its local control information. If the required properties for the delivery are satisfied (e.g. Best-Effort, FIFO, CAUSAL, TOTAL), the message is delivered to the application and the protocol updates its local control information. Otherwise, the message is buffered and its delivery to the application is postponed to the delivery of another message which will enable it by the induced modification of the local control information.
A multimodal channel is composed of a set of several channels $C_1, C_2, \ldots, C_n$, working as follows:

- Sending a message is done by specifying at the application level a subset of channels the message should be sent to. After sending the message, each channel updates its local control information as described above.
- At the reception of a message over a channel, the delivery condition, for each channel the message belongs to, is evaluated. If they are all evaluated to true, the message is delivered and the local control information updated for all these channels. Otherwise the message is delayed until all the conditions are satisfied.

One can notice that for any messages m_1 and m_2 sent over two sets of channels S_1 and S_2 respectively, the delivery policies for the channels in $S_1 \cap S_2$ must be compatible. The non respect of such a guard can lead to a deadlock situation in the delivery process.
Let's consider a group of sites $G = \{s_i\}_{0 \leq i < n}$ where $e_p(m)$ represents the emission of the message m on the site p, $d_r(m)$ is the delivery event of the message m on the site r, while the relation "$\prec$" expresses the "happens-before" relationship between two events [10]. The three elementary orders (i.e. FIFO, CAUSAL, TOTAL) can be specified as follows :

- FIFO: $\forall p \in G : \forall m, m' \in Messages, r \in G$
  $e_p(m) \prec e_p(m') \Rightarrow d_r(m) \prec d_r(m')$
- CAUSAL: $\forall p, p' \in G : \forall m, m' \in Messages, r \in G$
  $e_p(m) \prec e_{p'}(m') \Rightarrow d_r(m) \prec d_r(m')$
- TOTAL: $\forall p, p' \in G : \forall m, m' \in Messages$,
  $(\forall r \in G : d_r(m) \prec d_r(m')) \vee (\forall r \in G : d_r(m') \prec d_r(m))$

Some consistency rules exist or must exist among protocols; for instance, a causally ordered procotol guarantees the fifo property but a totally ordered protocol must preserve the causal order in case it has to be combined with a causal protocol.
In order to solve the causal coherency problem previously stated in figure 1, the application could use an abstract channel composed of a FIFO and a CAUSAL elementary channels. The start and end messages of each stream (video and audio) are sent on both elementary channels, whereas the other messages of each stream need only to be transmitted on the FIFO channel.
We propose an implementation of a multimodal channel that can properly handle the mix of any partial order protocol. Our implementation is based on a toolkit called JGroups.

MULTIMODAL CHANNEL IMPLEMENTATION

JGroups

JGroups (formerly JavaGroups) is an open source toolkit for reliable group communication written in Java [2]. It is based on IP multicast, but extends it with reliability and group membership.
A client application uses a high level abstraction, *JChannel*, that is composed of a set of $\mu protocols$ (software components taking care of different communication properties) layered above each other. A JChannel can be considered as a group communication socket that

the application uses to send/receive messages to/from the group. This abstraction makes the μprotocol stack transparent to applications. Therefore, the same application code can be reused for different network configurations just by changing the JGroups protocol stack. Figure 2 gives an example of a JGroups protocol stack for an application that needs reliable fifo delivery, flow control, group membership and node failure detection.

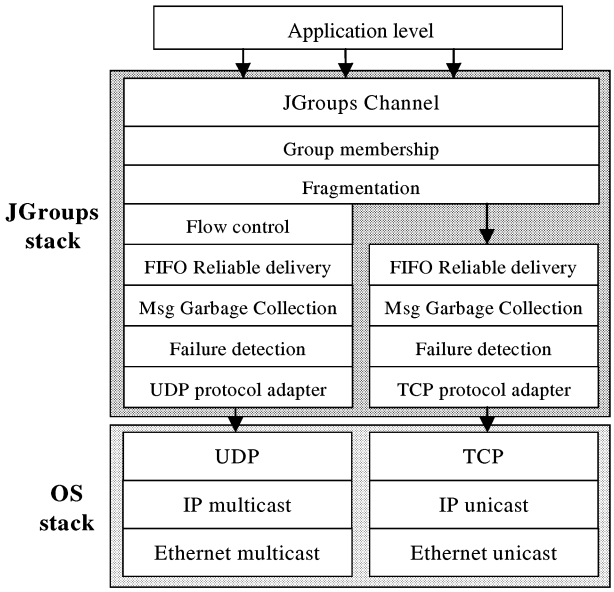


Figure 2. Example of a JChannel protocol stack

Message reliability is implemented using negative acknowledgements. Missing packets are resent using unicast messages. The JGroups protocol stack relies on a transport protocol such as UDP, TCP or any user defined protocol using a protocol adapter. By mixing and matching μprotocols, various application requirements can be satisfied. JGroups comes with a number of μprotocols, for example : fragmentation of large messages, reliable unicast and multicast, flow control, failure detection (crashed members are excluded from the membership), ordering protocols (FIFO, CAUSAL, TOTAL), membership, encryption.
Moreover, a set of μprotocols is offered to simulate various network conditions: DELAY (network delay), SHUFFLE (desequencing), DISCARD (controlled loss). Its pluggable architecture enables the easy development of new μprotocols.

MultiModalChannel

We propose a multimodal channel (which we call MModChannel in the rest of this paper) as an extension of JChannel. Made up of a configurable number of elementary channels, its simple interface is similar to that of an elementary channel.
An application that needs to exchange messages using different channels can create a MModChannel, by specifying the stack protocols for each of the elementary channels. When sending a message, the application chooses to transmit the same message to a set of elementary channels (coded as a mask). The message payload is transmitted only once, but the delivery process ensures that the message is delivered to the receiving application in the correct (partial) order of every channel it has been sent through.
Dealing with two or many partial order protocols (e.g. FIFO, CAUSAL) raises an important difficulty at the delivery process. Let's consider two messages m_1 and m_2 sent on each of two distinct partial order protocols P and Q. Moreover, let's say the two messages are not comparable with respect to P and Q. As P and Q have no order imposed to deliver the two messages, let's suppose that the messages are delivered by the elementary channels to the MultiModalChannel in two different orders. That is, the delivery process for each channel creates a total order which is an extension of its original partial order. Here we experience an artificial deadlock because the two extension orders P' and Q' may be incompatible. Yet, the two messages are independent and could be delivered in any way. The artificial order is nothing but a side effect of the delivery process serialization of any elementary protocol.
This problem lead us to reconsider the partial order at the MModChannel level. Therefore the elementary channels need to expose the partial order to MModChannel. The control information necessary to reconstruct the partial order graph at the multimodal level is transmitted by the elementary level along with the message. In order to keep genericity and express any partial order, this information is modeled by means of Fidge-Mattern vectors. At the multiModal level, the control information (Fidge-Mattern's vector) is extracted and the message is inserted in the partial order graph according to the vector's value.
The delivery algorithm (3) works in the following way : upon reception of a *root* message (message that is independent according to the receiving channel), its delivery condition (its presence among the roots of all channels it has been sent through) is evaluated. If it can be delivered, it is removed from these graphs, the graphs are updated and the message is delivered to the application. Then, the postponed messages depending on it (according to each channel order) that have become roots are

checked. The delivery process continues until there are no more deliverable messages and it is restarted every time a new message arrives at MModChannel level.

```
on receive(Message msg, Channel C) {
    if (msg ∉ roots(C)) return
    candidates ← msg
    while (candidates ≠ ∅){
        choose m ∈ candidates
        mask ← mask(m)
        if (⋀_{c∈channels(mask)} m ∈ roots(c)) {
            for c ∈ channels(mask) {
              remove m from channel c
              update roots(c)
              candidates ← candidates ∪ roots(c)
            }
            deliver m
        }
    }
}
```

Figure 3. MModChannel delivery algorithm

Experimental results

We conducted experiments using a videoconferencing scenario where four users exchange streams between them by means of multicast protocols. In order to repeatedly use the same interactive scenario (figure 4), a simple grammar has been defined to describe the behaviour of all participants and the interactions between them.

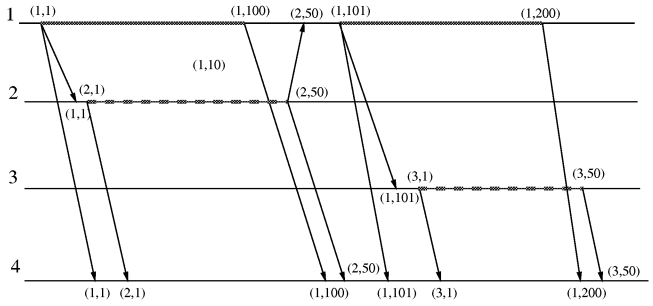


Figure 4. Experimental scenario

User 1 starts sending the first part of a video stream (V_1); after having seen the begining of (V_1), user 2 starts commenting on it by sending an audio stream (A_1). After hearing the last comment from user 2, user 1 continues to send the second part of the video stream (V_2). Once having seen the begining of the second part, user 3 starts commenting on it (A_2). This little scenario of 8 seconds is repeated 40 times, so the session duration is about 320 sec.

Table 2 contains the characteristics of the streams. Note that a video message (streams V_1 and V_2) represents a frame while the messages corresponding to streams A_1 and A_2 could be seen as audio samples.

Stream	Site	Messages	Size(bytes)	Rate(Kbs)
V1	1	100	1500	293
A1	2	50	600	39
V2	1	100	1500	293
A2	3	50	600	39

Table 2. Characteristics of scenarios streams

Three multicast communication patterns were studied:

- FIFO (fifo consistency between any two users)
- CAUSAL (causal coherency among all messages)
- MultiModChannel (FIFO and CAUSAL channels)

We run experiments on four Linux machines (1 GHz PIII CPU, 256 MB RAM) on a 100Mbps LAN. First, we evaluated the cost of the multimodal level only, by observing its overhead. Secondly, we simulated network conditions (loss, delay, desequencing) specific to a WAN, by using our simulation network μprotocols: MDISCARD, MDELAY, and MSHUFFLE.

The end-to-end latency for each message was decomposed into three elementary latencies L_1, L_2, L_3 as depicted in figure 5. These latencies are computed by timestamping the message at four levels: application down, network down, network up, application up.

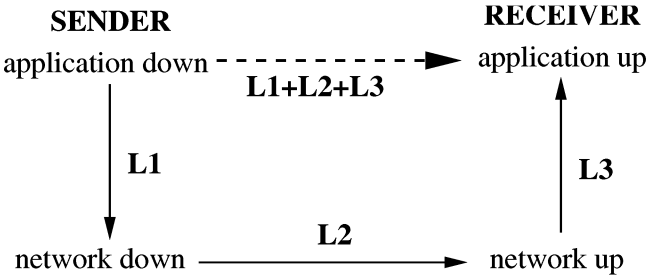


Figure 5. End-to-end latency decomposition

The MModChannel is composed of two elementary channels: FIFO and CAUSAL. Every single message is sent on the FIFO channel and some of them on the CAUSAL channel. The main idea is to ensure at every user the FIFO order and, for some messages, the CAUSAL order.

First experiments were made in ideal conditions (high speed lossless network). In table 3, one can observe the growth of end-to-end latency according to the percentage of messages sent on the two channels. It sould be noted that the overhead introduced at the MultiModal level remains reasonably low (a few percents).

Channel	MULTIMODAL (FIFO + CAUSAL)				
On CAUSAL	0%	25%	50%	75%	100%
Latency	5.01	5.42	8.88	10.33	11.61
Overhead	0.21	0.25	0.38	0.53	0.66
Percentage	4%	4%	4%	5%	5%

Table 3. MultiModal overhead in a lossless network

More realistic (Internet like) network conditions were simulated by introducing a network delay of 100 ms and a loss rate of 5%. The figure 6 shows how multimodal bottom-up latency (L_3) grows (almost linearily) as a function of the percentage of messages sent on the CAUSAL channel.

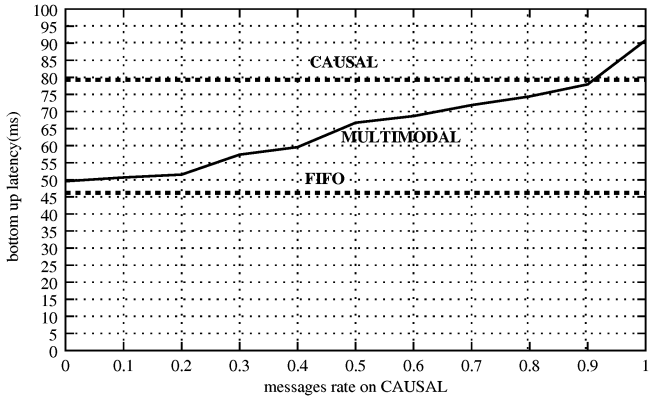


Figure 6. MModChannel latencies in a Internet-like network

ADAPTATIVE COHERENCY SERVICES TO NETWORK CONDITIONS

Figure 6 shows that a multimedia application may find too expensive (in terms of delay) the use of a causal channel for every message of the streams that need causal delivery. The use of the FIFO channel is cheap, but does not provide any causal coherency.

As end-to-end latencies could seriously affect the quality of a videoconference, let's consider an acceptable bound L_{max} for them. We propose the following QoS requirement for a communication path: the percent of messages having latencies above L_{max} is limited to a given p_{max}.

Therefore, we raise the following question: for a given values L_{max} and p_{max}, how many messages can be sent on CAUSAL or, more generally, on the most expensive channel (strongest partial order) in order to respect the QoS requirement ?

Various mechanisms can be used to compute the maximal percent of messages sent on CAUSAL:

- simulate from real network traces;
- simulate from average network estimations.

Using the second approach, we consider several network types with an average delay d of 100ms and a given loss rate $\rho \in [0 : 0.1]$. As average end-to-end latency grows with the percent prc of messages on CAUSAL, one can find a good approximation for the maximal prc_{max} value by means of simulations for some prc values (e.g. $\{0.0, 0.1, 0.2 \ldots 1.0\}$).

For every network type ($d = 100ms$, $\rho \in [0 : 0.1]$) we played the scenario (4); we fixed the percent p_{max} at 20% for an end-to-end latency bound (L_{max}) of 300ms. Figure 7 shows the possible degradation of coherency (in terms of messages sent on CAUSAL) with the raise of network loss ratio (ρ).

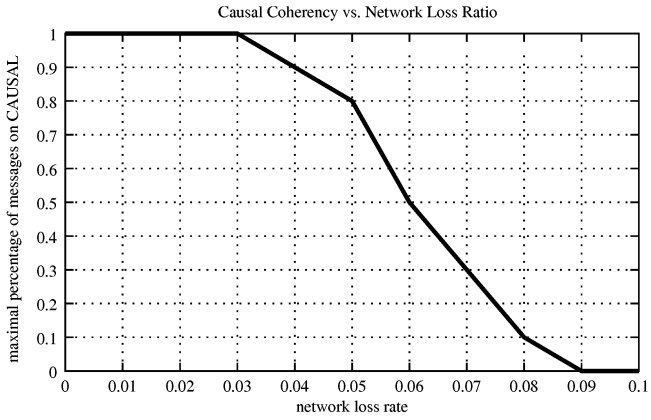


Figure 7. Maximal Percentage on CAUSAL vs. Loss Ratio

By observing the network conditions (e.g. RTT and loss ratio), the MultiModalChannel could use simulations results to determine what percent of messages could be sent on CAUSAL with respect to the maximal percent of latencies over L_{max}.

Moreover, during the multiparty session, the coherency level (the prc percent) could be adapted, by comparing the current percent p of latencies above L_{max}, to p_{max}. If $p < p_{max}$, the MultiModalChannel could choose a higher prc (**more coherency**), otherwise the prc must be fixed to a lower value (**less coherency**).

In conclusion, the MultiModalChannel offers the application a degree of causal coherency adapted to the network condition: the coherency level is choosed such

that the percent of messages arriving beyond a fixed latency remains bounded. An alternative solution to relax causality in distributed multimedia systems is described in [7] where the authors propose a modified CAUSAL protocol (i.e. Δ-causal) that ignores some causalities in order to bound latency.

CONCLUSION AND PERSPECTIVES

In this paper, we proposed an implementation of a generic adaptive multimodal communication service and show how it can be used in a multiparty videoconferincing application. Our proposals are:

- a JGroups-based implementation of a generic multimodal channel;
- a delivery algorithm that is able to correctly deliver messages sent on various partial-order channels;
- a FIFO Partially Reliable μprotocol with bounded average loss (FIFOPR);
- an adaptative communication service that is able to trade-off coherency and cost in multimedia applications that can accomodate with variable coherency.

Our future work will focus on :

- experimentations on real WAN networks with :
  - our FIFOPR protocol instead of JGroups reliable FIFO;
  - integration of Δ-causality in the CAUSAL channel;
  - integration of asymetric network conditions in a multiparty application;
- study the impact of the fragmentation of large messages (e.g. compressed video frames that are larger than the MTU of the network) on a lossy network;
- perceptual validation of the use/lack of coherency in a real-world videoconferencing application;
- finding others applications of multimodal channel (e.g. simulations in the wide-area distributed computing).

References

[1] S.Pomares Hernandez, J.Fanchon and K.Drira, *Abstract Channels as Connectors for Software Components in Group Communication Services*, In *Fifth Mexican International Conference in Computer Science*, September 2004.

[2] Bela Ban *Design and Implementation of a Reliable Group Communication Toolkit for Java*, *Cornell University*, September 1998.

[3] R. Grigoraş, G. Padiou, Ph. Mauran and Ph. Queinnec, *Causal Ordering of Distributed Multimedia Streams*, In *Technical Report IRIT-ENSEEIHT*, 2003.

[4] Hugo Miranda and Luis Rodrigues, *Flexible Communication Support for CSCW Applications*, In *Internation Workshop on Groupware*, September 1999.

[5] Hugo Miranda, Alexandre Pinto and Luis Rodrigues, *Appia: A Flexible Protocol Kernel Supporting Multiple Coordinated Channels*, In *The 21st International Conference on Distributed Computing Systems*, April 2001.

[6] Ajay D. Kshemkalyani *Temporal interactions of intervals in distributed systems*, In *Journal of Computer and System Sciences*, April 1996.

[7] R. Baldoni, R. Prakash, M. Raynal and M. Singhal, *Efficient Δ-causal Broadcasting*, In *Journal of Computer System Science and Engineering*, September 1998.

[8] Friedemann Mattern, *Virtual Time and Global States of Distributed Systems*, In *In Workshop on Parallel and Distributed Algorithms*, October 1988.

[9] Colin Fidge, *Timestamps in Message Passing Systems that Preserve the Partial Ordering*, In *Australian Computer Science Communications*, February 1988.

[10] Leslie Lamport, *Time, Clocks and the Ordering of Events in a Distributed System*, In *Communications of the ACM*, July 1978.

Removing the MAC Retransmissions Times from the RTT in TCP

Eugen Dedu, Sébastien Linck, François Spies
{Eugen.Dedu,Sebastien.Linck,Francois.Spies}@pu-pm.univ-fcomte.fr
Laboratoire d'Informatique de l'Université de Franche-Comté (LIFC)
IUT de Belfort-Montbéliard, 4 place Tharradin, BP 427, 25211 Montbéliard, France

KEYWORDS
Wireless network, MAC retransmission, RTT, TCP.

Abstract

MAC retransmission of packets (for example in a video transfer) raises the RTT and leads to throughput decrease on TCP versions based on RTT for their sending rate. This is not the appropriate effect, because retransmissions are given by temporary interference, which often appear in wireless links. This paper treats effects of MAC retransmissions. A new TCP option is proposed, which fully takes into account the effect of MAC retransmissions. Wireless network cards have a timer which is initialised with the value of the option once the packet is sent the first time, and is put in this option each time the packet is retransmitted. The exact time of retransmissions is then known by the source. Simulations show that the proposed mechanism increases throughput.

INTRODUCTION AND MOTIVATION

It is now an evidence that TCP is not adapted to wireless links. The main concern is that losses/delays are given by *temporary* interferences, hence no congestion mechanisms should be taken.
Many studies treat effects of packet *loss* in wireless links (802.11) on TCP, but only few treat effects of packet *delay* during retransmission. It is worthwhile to note that successful MAC retransmissions appear much more frequently than MAC losses.
On wireless links, MAC retransmissions are the effect of radio interferences. Because these are temporary and appear randomly, they do not give any useful value, but mislead network mechanisms. Therefore, several works exist for avoiding their effect. But works treating the MAC retransmission concentrate generally on the influence the RTT has on RTO (retransmission timer) of TCP [11, 7, 6].
However, there are several TCP versions that use the RTT for their *sending rate*, for instance Vegas [1] uses all the RTT samples, and Westwood+ [4] and TIBET [2] use smallest RTT. RTP/RTCP over UDP, common for video streaming for example, is another case where RTT values are needed. Other mechanisms which need correct RTT may arise. [10] shows that in certain cases a gain of 10% in throughput appear if the real RTT is replaced by the adjusted RTT.
This article proposes a mechanism that allow senders to take into account the time taken by retransmissions. A TCP option is added, with one field. Also, a timer is added to the network card. Each time a packet is transmitted, the network card puts in its option the time taken by its retransmission, using the internal timer. The field is echoed back to the source in an ACK packet. The source is hence aware of the time lost in all the MAC retransmissions. A similar mechanism can be applied to non TCP flows too.

BACKGROUND

802.11 is a protocol which is used with two access modes: PCF (Point Coordination Function) and DCF (Distributed Coordination Function).
In DCF mode, each machine can access the network when it wants. However, in order to reduce collisions, machines must choose randomly a value, called *backoff*, in a given range, called collision window (Contention Window CW). CW doubles at each retransmission (with an upper bound equal to a power of 2 minus 1). The initial window lies between 0 and 31 ($= 2^5 - 1$) units. Units are called *time slots*, with $t_s = 20\mu s$ as defined by the standard.
In PCF mode the AP (access point) is the master of conversations. It gives the "access" in turn to all the mobiles. Mobiles can send packets only when the AP gives them the right. There is no backoff. It is worthwhile to note that in the mode PCF each period of time is divided in two parts: the PCF part and the DCF part. This allows machines which implement only DCF mode to have the possibility to access the network. The method proposed in this article works in both DCF and PCF modes.
802.11 is a protocol which relies on ARQ (Automatic Repeat Request). After sending a packet, the network card listens for an acknowledgement from the receiver's network card. If it does not arrive, then the sender's network card considers that the packet was lost and it will retransmit it. There is a limit in the number of retransmissions.
Packet loss is frequent in wireless networks. They appear generally when an external interference occurs in

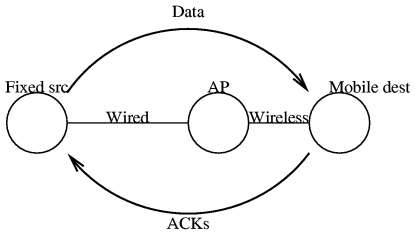


Figure 1: A TCP transmission between a wired machine and a wireless one.

the network, but also when the mobile machines exits from the covered area of the network. Interferences are known to be temporary and appearing at random times. A detailed explanation of the 802.11 standard can be found in [3].

PRINCIPLE

Due to frequent losses, 802.11 allows MAC retransmission. Hence, 802.11 transform a network with losses and predictible delay into a network with no loss and variable delay. On the other hand, packet losses are generally due to interferences. As they are *temporary*, the RTT should not be influenced by them. We therefore propose a mechanism to remove the delay taken by retransmissions.
In order to implement this idea we were faced to several choices. Is the number of retransmissions or the total retransmission times which is sent to the sender? Is this information put at the TCP or at the IP level? Is this information sent by the AP directly to the source, or passed to the destination which in turn sends it to the source? Several solutions exist, we present only the solution we judge the most appropriate.
Our solution is the following (see figure 1). An option is added to TCP, called `timelost`, containing a field, called `rets`. Each wireless network card has a timer. The timer is initialised to the value of the `rets` field of the packet for the first transmission of a packet (after the backoff). Each time it sends a packet on the wireless link, the value of the timer is put in the `rets` field. Thus the timer reflects the *exact* time loss due to retransmissions. When the source receives a packet, it takes the appropriate action, for example it subtracts the value of the `rets` field from the RTT of the packet.
The reminder of this section details this mechanism in TCP. A similar mechanism can be used for non TCP flows too, for example in an RTP/RTCP transmission, the receiver puts in RTCP packets going to the source the time lost in retransmissions of data packets.

Retransmission Time Computing

As specified, each network card has a timer. Each time a new packet is processed, its `rets` is used to initialise the timer. This allows to sum several times, for example when the packet is an ACK and it contains already a time, the lost time of the corresponding *data* packet. During each (re)transmission, the value of the timer is put in the `rets` of the packet.
The modification which needs to be done at the network card level is the addition of a timer. As nowadays timers exist on all the network cards (used for backoff for example), this is not an issue.
The MAC-level fragmentation does not influence our mechanism. Indeed, when an IP packet is fragmented the time loss is null if each fragment arrives at destination without retransmission.

Transmitting the Information to the Source

We propose an option, called `timelost`, to be added to TCP. This option has only one field, containing a time value. If the field has 2 bytes and the measurement unit is the time slot t_s given previously, then the field will overflow at a time $t = 65536 t_s = 65536 \times 20\mu s \approx 1.3s$. In the 802.11b standard the maximum CW is 1023, hence 2 bytes are sufficient. If the field has 4 bytes, it will overflow after $t \approx 65536 \times 1.3s \approx 1$ day, which is largely sufficient.
The source sets this field to zero. During the trip, the field may be modified by a wired-wireless machine. The receiver echoes back the value to the source in an ACK packet, exactly like the TCP timestamp option [5]. During the return trip, the field may further be increased. The source receives the ACK packet and computes the retransmission time.
This mechanism allows incremental deploying. It gives useful values only if the sender, the receiver and the AP know about it. If the sender is not aware of this option, nothing happens. Elsewhere, it adds the option and set the field to 0. If the AP and/or the receiver do not know this option, the sender either receives no option, or an option with value 0, which do not change anything either.

Actions Taken by the Source

It is up to the source to take appropriate actions. For instance, we suggest that a TCP Vegas source would use the new RTT in its formulae, or to a video server this would allow to better know the network jitter.

Drawbacks

In this mechanism, network cards have access to level 4 (TCP) information. It is worthwhile to note that several papers propose such methods.
Also, it needs to be further evaluated if the gains obtained by this mechanism for a higher range of transfers overtake the changes needed to implement it.

SIMULATIONS

The NS2 simulator [8] version 2.28 is used to simulate our mechanism. We have modified NS2. All the modifications to NS2 and the source of the tests shown in this article are available on-line[1].

Modifications to NS2

We add only a new field, called `rets`, to TCP header. It is easier to implement and does not change results. `rets` is of type `double`. We modify:

1. The sending part of TCP/Vegas, which initialises `rets` to 0.
2. The sending part of the MAC 802.11 protocol, which adds retransmission time to `rets` field.
3. The sending part of TCPSink (whose role is to send ACKs), which echoes `rets` field of the data packet in the ACKs it sends.
4. The receiving part of TCP/Vegas, which computes RTT as RTT minus `rets`. There is no other action to take at source, because it sees the modified RTT.

The second item is used/applied only when `useRets` is 1. `useRets` is an NS2 parameter which can be used in user's `.tcl` files. It defaults to 0.
At 802.11 level, each time a packet is retransmitted, its retransmission time is computed. The value used is the time between the packet sending time and the reception of its acknowledgement.

Shadowing propagation model in NS2

Currently, three wireless propagation models are implemented in NS2 [9]: Free space, Two ray ground and Shadowing. The two first models are of type "all or nothing": If distance d between the mobiles is smaller than a certain value, all packets are received. If d is greater, no packet is received. These are not appropriate in our case, since we need retransmissions from time to time.
In the shadowing model, packets are always received for $d < s1$, always lost for $d > s2$ and received with a probability for $s1 < d < s2$.

Simulations

The simulated network in given in figure 1: a wired TCP Vegas source, an AP, and a mobile TCP destination. Note that both AP and the mobile use retransmission times. The mobile destination moves away linearly from the AP, starting at $t = 0$. At $t = 150s$, it stops moving and stay there until the end of simulation. It is near

[1] At `http://lifc.univ-fcomte.fr/~dedu/publi/euromedia/`

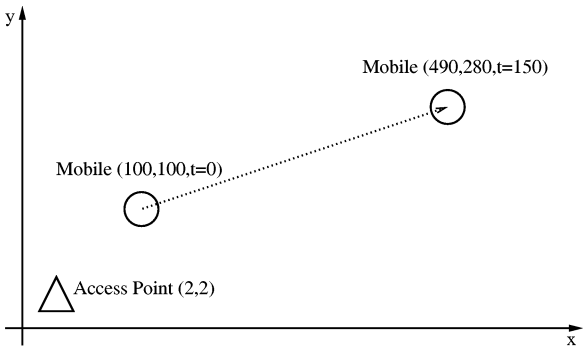


Figure 2: The movement of the mobile.

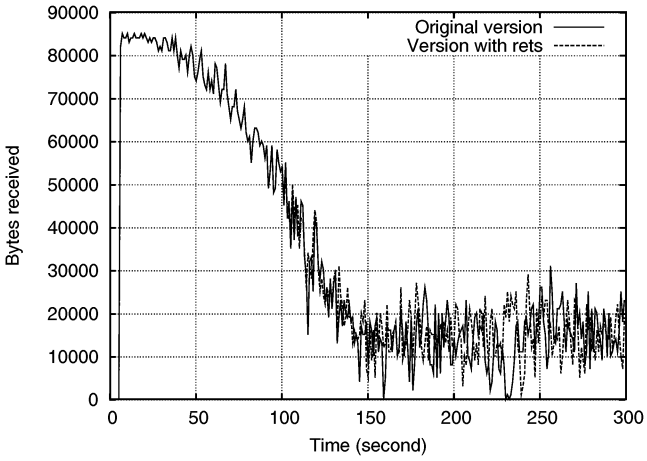


Figure 3: Simulation results.

the covered area of the AP, where most of the retransmissions appear. Figure 2 details the movement of the mobile.
Two simulations are executed: one with original version of TCP and one with modified version of TCP. The results of the simulation are given in figure 3. Several stages can be seen in the figure:

- During the first 5 seconds, the routing protocol initialises and no data exchange appear.
- In the shadowing model, no packet loss/retransmission appear until a certain distance. This is clearly seen in the figure, as the curves of two versions are identical for $t < 100$. Gains can be seen only on retransmissions.
- For $t > 150$, retransmissions are common. The mobile does not move. An average improvement of 4.5% is obtained.
- The maximum increase in the bandwidth over a 100 seconds interval is obtained for $150 < t < 250$: approximately 15%.

RELATED WORK

From the implementation point of view, there are several works on modifying the TCP specification and hence the sender and the receiver.

Several works treat packet loss in wireless networks.

[11] presents an analytical model that predicts the RTO of TCP for given network parameters. It concludes that delay variations are critical only when they are of order of seconds.

[7] presents a model where certain packets are artificially delayed at link-layer. The RTT increases, and the RTO of TCP with it, but the number of false TCP retransmissions decreases, leading to a higher overall throughput.

In [6], the wireless access point (AP) encapsulates each packet in a header containing a timestamp, which is echoed by the receiver in its ACKs. This allows AP to discover lost packets even if they are out-of-order. When such a loss is detected, the AP sets a bit (ERN, Explicit Retransmission Notification) in the ACK packet which inform the sender to retransmit the packet without taking congestion actions.

The negative effect of MAC retransmission on TCP is treated in [10], where the TCP connection between a wired machine and a wireless machine is divided in two TCP connections by the AP in the middle. The AP buffers data received from the wired end and retransmit it to the wireless end if it was not received. Also, the time spent in the AP is subtracted from the TCP timestamp option. Contrary to the method proposed in this article, the time spent at the AP is not accurate (the timestamp granularity depends on the source machine [5]), the AP needs to buffer data and it works only with TCP timestamp option.

CONCLUSIONS AND FUTURE WORK

This article presents a method to remove the effect of MAC retransmissions on the RTT computing on source-side in TCP. In adds an option to TCP containing a time value. This value is increased by wireless network cards whenever the packet is retransmitted. The field is propagated back to the source, like the TCP timestamp option [5]. This allows the sender to know exactly the time passed in retransmissions.

Simulations have shown an improvement of throughput with our method. We plan to do real experiments in order to validate our simulations. If successful, we plan to give a complete proposal for improving TCP in case of retransmissions on wireless links.

ACKNOWLEDGEMENTS

Authors are grateful to Dominique Dhoutaut for his insights in the 802.11 protocol.

References

[1] L. Brakmo and L. Peterson. TCP Vegas: end to end congestion avoidance on a global Internet. *IEEE Journal on Selected Areas in Communication*, 13(8):1465–1480, Oct. 1995.

[2] A. Capone, L. Fratta, and F. Martignon. Bandwidth estimation schemes for TCP over wireless networks. *IEEE Transactions on Mobile Computing*, 3(2):129–143, Apr.-June 2004.

[3] D. Dhoutaut. *Etude du standard 802.11 dans le cadre des réseaux ad-hoc : de la simulation à l'expérimentation.* PhD thesis, Ecole Nationale Supérieure de Lyon, Dec. 2003.

[4] L. A. Grieco and S. Mascolo. TCP Westwood and Easy RED to improve fairness in high-speed networks. In *IEEE/IFIP Seventh International Workshop on Protocols for High-Speed Networks*, pages 130–146, Berlin, Germany, Apr. 2002.

[5] V. Jacobson, R. Braden, and D. Borman. TCP extensions for high performance, May 1992. RFC 1323.

[6] Z. Jing and N. Zhisheng. A reliable TCP-aware link layer retransmission for wireless networks. In *International Conference on Communication Technology*, volume 1, pages 900–905, Beijing, China, Aug. 2000.

[7] N. Möller, K. H. Johansson, and H. Hjalmarsson. Making retransmission delays in wireless links friendlier to TCP. In *43rd IEEE Conference on Decision and Control*, Bahamas, Dec. 2004.

[8] Network simulator — ns-2. `http://www.isi.edu/nsnam/ns/`.

[9] The NS manual. `http://www.isi.edu/nsnam/ns/ns-documentation.html`.

[10] K. Ratnam and I. Matta. Effect of local retransmission at wireless access points on the round trip time estimation of TCP. In *The 31st Annual Simulation Symposium*, pages 150–156, Boston, MA, USA, Apr. 1998.

[11] M. Scharf, M. Necker, and B. Gloss. The sensitivity of TCP to sudden delay variations in mobile networks. In *3rd IFIP-TC6 Networking Conference*, pages 76–87, Athens, Greece, May 2004. LNCS 3042.

Inside the adaptation decision-taking engines : monitoring adaptabilities by learning

Romulus Grigoraş Pascal Dayre Vincent Charvillat Hadj Batatia

IRIT-ENSEEIHT UMR CNRS 5505

2 rue Camichel BP 7122 31071 Toulouse Cedex 7 France
{grig,dayre,charvi,batatia}@enseeiht.fr

Abstract

Both user modeling and multimedia adaptation communities have been actively carrying out research on adaptation, but the design of adaptation decision-taking engines is still an open issue for each application. In this paper we consider the decision-taking engines by using machine learning methods and propose an experimental adaptation platform. We also briefly describe actual implementations of this platform with 3 case-studies: supervised learning for teaching systems, reinforcement learning for multimedia streaming and unsupervised learning for adaptive hypermedia navigation support.

Keywords

multimedia adaptation, hypermedia, machine learning, dynamic adaptation, user modeling

1 Introduction : adaptabilities anywhere in our research communities

The Universal Multimedia Access (UMA) initiative is coming from a phenomenal increase :

- in the quantity and complexity of content (including {multi/hyper/rich}media content)
- in the number of different users with different preferences and intentions
- in the diversity of devices and digital storage (including mobile technologies)
- in the networks heterogeneities.

Interoperability through multimedia adaptation services is the key for enabling transparent use of multimedia content anytime anywhere and anyway.
Far from the previous Internet or TV paradigms, the emerging use cases are further complicated by the inherently dynamic nature of mobile communication and interactive services. Users and providers alike have to deal with rapidly varying bandwidth, load on terminals etc. There is, therefore, more need than ever before for context aware services and applications which are able to adapt dynamically to changing service provision conditions.
Standardization committees such as The Movig Picture Experts Group (MPEG), the Internet Engneering Task Force (IETF), the World Wide Web Consortium (W3C) are working on *ressource adaptation tools* providing scalable content, transcoders and/or transmoders of media stream or global scenes transformations.
The ensuing descriptors (see. MPEG-21 DIA [TH05], RDF and CC/PP [ccp]) try to represent the knowledge of usage constraints and users, servers and network contexts. Nevertheless these well-thought out tools also demand efficient mechanisms that allow delivering the aimed service in the best way. The design of adaptation decision-taking engines is still an open issue for each application even if elementary heuristic (CC/PP profiles matching for instance) may be directly inherit from the description framework.
In the same trend, (automatic) personalization has become an essential part in many adaptive Web-based applications. Many adaptive hypermedia systems offer: recommandation components, link annotation or hiding and/or canned text adaptation in the terminology of Brusilovsky's overview paper [Bru01]. There are several approaches for creating adaptive hypermedia systems. Both the navigation support and the content itself can be dynamically adapted based on visitor access patterns or on rules (pre)defined by an author. In adaptive hypermedia methods and techniques are often used to tailor a presentation to the individual user through a *user-model based adaptation*. The previously introduced evolution in Web technology and UMA now requires to provide adaptation for documents and to provide adaptation to more than just the users model (a learning style in an educational adaptive system) but also the properties of the user-interface device and the network connection.
In that sense, we claim that several communities between Web, multimedia, adaptive hypermedia system share some common objectives. It is no longer clear that the tools for *ressource adaptation* must be independently designed from *user-model based adaptation*. We bring two main contributions from that. This obvious yet complex conver-

gence is first analysed in the next section of this paper. We exploit the common features of all these adaptabilities to tackle the design of the adaptation decision-taking engines by using machine learning methods and propose an adaptation learning platform in section 2. The use of three types of learning (supervised, reinforced, unsupervised) are illustrated with three adaptive mechanisms (see the sections 3,4,5).

2 Adaptation decision by learning

2.1 Convergence elements

After observing the approches of various research communities that deal with adaptation it becomes clear that they converge on the following characteristics of an adaptive system:

- the *usage context*: contains elements from both user-modeling adaptation and ressource-based adaptation : user characteristics, preferences as well as available ressources (terminal, network etc.);
- the *context descriptors*: important standardization efforts have been made for defining interoperable context descriptors (e.g. UED of MPEG-21 [TH05]);
- the *dynamicity*: various statical adaptation strategies have been proposed, but there is a growing need for dynamic strategies, that can better capture the variability of the usage context. The *usage context* feedback in form of context descriptors is necessarily needed by dynamic adaptation;
- the *decision engine* is often described as functioning in two separate steps: decide/choose the adaptation strategy (specify the adaptation result) and effectively apply that strategy (produce the result);
- *application of a decision*: finding the appropriate adaptation action can be sometimes non-trivial. It may be necessary to search for/compose various adaptation services. For example, if the decision is to have a specified bitrate for a video sequence and no single service is able to produce the exact result, the system may choose to apply a combination of resizing/temporal decimation adaptation services.

2.2 Our adaptation platform

Motivated by the need of integrating and evaluating different dynamic adaptation systems that have a common view on the previously cited convergence elements, we propose an adaptation learning platform presented in figure 1.
The platform includes:

- a decision-taking engine (DE);
- effectors (E);
- mechanisms for observing the context (CL, Context Listeners)

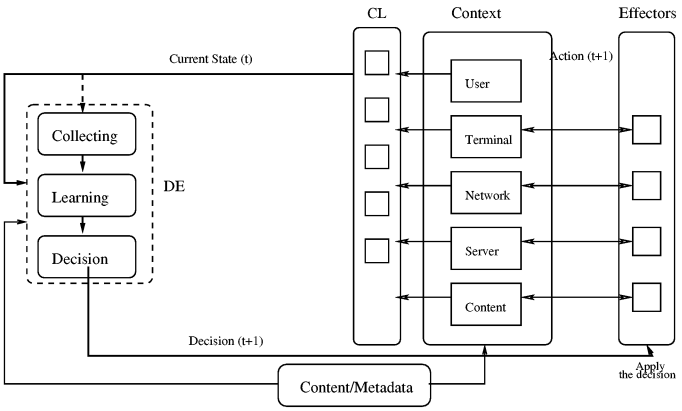


Figure 1: Our adaptation-learning platform

Constraints for adaptation are made available to adaptation engines from the usage environment (receiving terminal, servers, proxies and/or the network sensors). They include user characteristics and preferences, terminal characteristics (e.g. display, audio/video capabilities, storage, processing and power limitations), network and environment characteristics.
The adaptation is done in three steps:

- first, context information is retrieved by the decision-taking engine (DE) as a collection of context descriptors that we call *current state*, collected by the listeners (sensors);
- second, given a set of possible adaptation choices, the engine needs to decide what is the right choice, based on the state information, the content and its metadata; DE produces adaptation actions that are carried out by the effectors;
- third, the effectors put into practice the adaptation: effectors make the necessary actions on the components of the context. For example, a content effector would trancode the media (act on the content), a network effector would setup the network according to QoS adjustements.

The decision engine has three subcomponents:

- a *learning component*: it works in closed-loop and issues a decision model (for example, Markov, bayesian or neural networks models). This decision model usually has a functional form that takes as input the current context state and outputs the adaptation decision;
- a *decision component*: it uses the adaptation model and the current state of the context for choosing an adaptation action ;
- a *collecting component*: some decision engines need an initial set of training data (observations of the usage context) before issuing adaptation strategies; this may

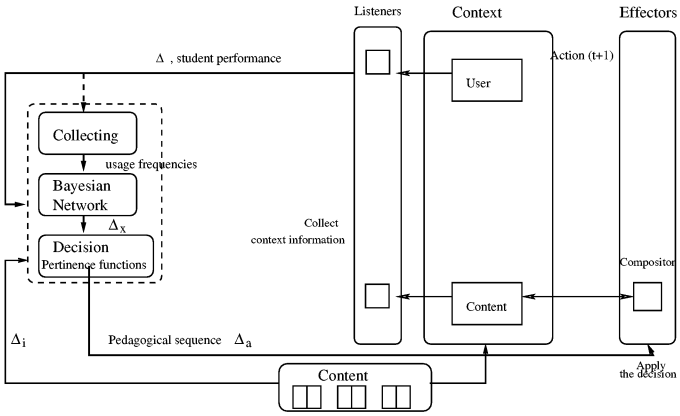


Figure 2: Adaptation learning in an educational context

include traces, navigation logs etc.; this prevents the decision engine from issuing quasi-radom adaptation models when not enough data has been incorporated (no convergence for the model).

Our experimental platform allows us to unify/compare various approaches for adaptation learning. We try to answer to a wide range of questions:

- on data: what past experiences can we rely on?
- on the assumptions: what can we assume about the context (user, content, network etc.) ?
- on context representation: how do we summerize the context state ?
- on the decision model: how the adaptation action can be predicted from the current state ?
- on estimation: how can we learn decision model's parameters ?
- on evaluation and on model selection: how well are we predicting ? perhaps we can do even better ?

In the rest of this paper we will present three adaptation systems that seamlessly integrate into our learning platform. These systems use different types of learning: supervised, unsupervised and reinforcement learning.

3 Supervised learning: an adaptive teaching system

Adaptive composition of pedagogical sequence. Personalizing instruction to individual students requires modelling student, content and pedagogical rules. Similarly to other application domains, personalization takes the form of adaptive navigation ([Bru03]) and adaptive presentation ([Cri04, FBHF04]). The former is commonly based on student knowledge model and the latter relies on student preference model.

A significant research effort has been devoted to knowledge models [CGV02].Work on preference models is less important. Most researchers consider preferences as static attribute-value pairs used during navigation [FBHF04].

We propose an adaptive student preference model based on a subject independent measure: the learning style [FFD02]. A student learning style is measured by a four-component vector $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4); \alpha_i \in [0, 1]$. Figure 2 shows the architecture of an adaptive teaching system that dynamically delivers pedagogical material depending on the student learning style. This architecture naturally fits our proposal (see figure 1). Content consists of pedagogical sequences described each by a vector $\Delta = (\delta_1 \ldots \delta_n)$; where δ_i is a discrete value representing the nature of the document (theory, example, exercises) , its presentation (text, audio, video etc.), its internal structure (top-down, bottom-up), its external structure (selective, linear etc.). The student is given a pedagogical sequence that responds best to the learning objective and to the current estimation of her learning style. At the end of the sequence, the terminal sends back various student performance indicators (scores, time etc.).

A Bayesian network model. A Bayesian network is used to model dependencies between learning styles, learning performance, and content descriptors [PBA05]. The structure of the network is static. However, conditional probabilities are learnt during a supervised session. During this network training session, students respond to a questionnaire that measures their learning style. They are then given a selected pedagogical sequence resulting in certain achieved performance. The statistical frequencies of content descriptors and measured styles allow the calculation of conditional probabilities. During system operation, context information is fed to the Bayesian network resulting in calculating a descriptor Δ_x of the reference (i.e. ideal) sequence to be presented to the student. The decision stage takes this ideal descriptor and fetches the content metadata for suitable content. This is done by selecting sequences that minimizes a pedagogical pertinence distance $\Pi(\Delta_x, \Delta_i)$. The resulting adapted metadata $\Delta_a = argmin(\Pi(\Delta_x, \Delta_i)\ \forall i$, is passed on to the Compositor effector which loads appropriate material and composes the sequence to be presented to the student.

4 Reinforcement learning for adaptive streaming

A streaming application. Hypermedia access from network-constrained platforms (e.g. mobiles) brings out the problem of latency reduction. In fact, following a link to a heavy content (an streaming audio/video sequence, for example), requires to download a playout buffer before the stream can play. This buffering process results in important delays, usually as long as a few seconds. The inconvenience caused to the user is all the more so in that, during the navigation (a sequence of clicks), the users experiences

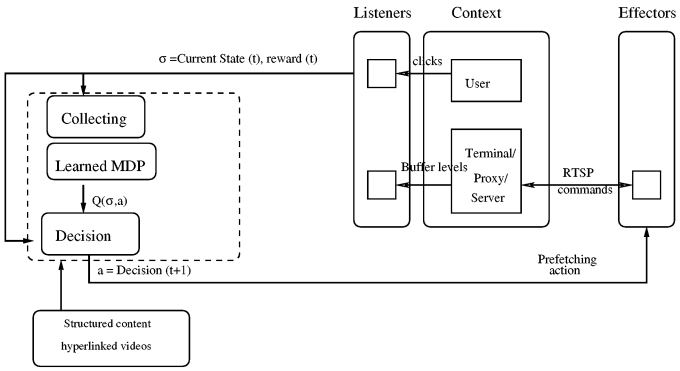
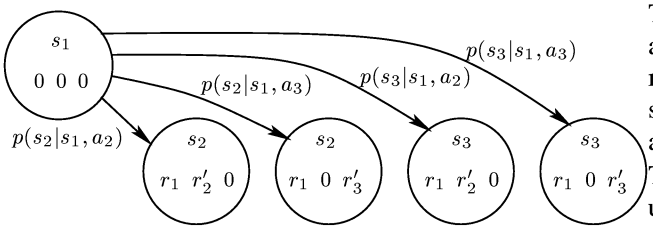


Figure 3: Reinforcement learning for adaptive streaming

the sum of these individual delays. The application could therefore make an effort to adapt the access by setting up a prefetching system that would reduce these delays. The streaming becomes adaptive. In this case the application has to decide what are the best prefetching actions for minimizing the latencies.

A Markov Decision Process (MDP). We model the problem as a MDP:

- the *current state* σ: the current hypermedia the user is viewing (e.g. the playing video s_1) and the quantized filling rate of the buffers (r_i) for any video (e.g. s_2 or s_3) the user has potentially access to
- the prefetching actions a_i: the system may choose to prefetch beforehand some accessible video object (e.g. a_2 = "prefetch s_2")



- the number of possible states is large and there is a probability ($p(s_i|s_j, a_k)$) associated to the transition between states (the probability of reaching s_i after s_j and having decided to prefetch a_k);
- the *reward* $r(\sigma, a)$ associated to a prefetcing action a is defined as the reduction of latency compared to the policy that does no prefetching at all. For a state $\sigma = (s_k, r_1, ..., r_n$ it is proportional with the filling level r_k.
- the *decisional policy* π: a function that associates an action a to each context state σ.

The aim of the *learning* component (see figure 3) is to operate a reinforcement ("trial and error") type of learning. The system tries to reinforce the decisions that resulted in good accumulation of reward and, inversely, tries to avoid unfruitful decisions. This implies an iterative update of its decisional policy that tends to an optimal policy, that is a policy that minimizes the latency experienced by the user. The *learning* component continually improves the decisional policy (adaptation model $Q(\sigma, a)$) and in each context state σ the *decision* component extracts the best associated prefetcing action a. According to the localization of the prefetching mecanism, the terminal, the proxy or the server puts into practice (through *effectors*) the prefetching by using RTSP play/stop requests. The *collecting* component may be used at bootstrapping, when no traces are present. An offline simulation algorithm such as "Q-learning' is then used to simulate user navigations and prefetching actions, and produce an initial adaptation model Q_0. This model is naturally refined when the system is operational by observing rewards and reinforcing good prefetching actions. Experimental results of using reinforcement learning for adaptive streaming can be found in [GCD02].

5 Unsupervised learning for adaptive navigation support

A hypermedia navigation aid system. Hypermedia has a non-linear structure of information creating an navigation space with redundant paths. Although several paths offer greater flexibility, the user can be disoriented. It is the well-known lost in hyperspace syndrome.

Adaptive navigation support has been proposed as means to aid users in finding their way through information spaces.

Our work consisted of modifying an existing "one-size fits all" approach of an SVG based aircraft handbook.

The produced NAVAIR handbook [DB05] now offers guidance to technicians who explore aircraft documents for maintenance procedures. By collecting user interactions, a sequence mining model is learned allowing dynamic links annotation and generation, content hightlighting etc.

The previously proposed architecture has been used to set up an adaptive navigation system based on sequence learning and SVG plus XHTML content adaptation. Figure 4 shows the architecture of the proposed Java-based system. Within the usage context, the user interacts with an XHTML document encapsulating various SVG components. A listener captures and routes user interactions to the decision taking engine. A prediction agent processes the probabilistic tree, previously learnt by the prefix-span model (see below). A DOM adaptation action is then produced from a predicted navigation and is forwarded to the adaptation effector. This last component is a DOM agent that acts directly on the content by annotating and highlighting. The next section presents the unsupervised prefix-span model that computes a decision model from the navigation logs (i.e. training data).

A sequence learning model. A sequence learning model can be useful to predict or generate future elements of a sequence, recognize a prefix or sequentially decide some

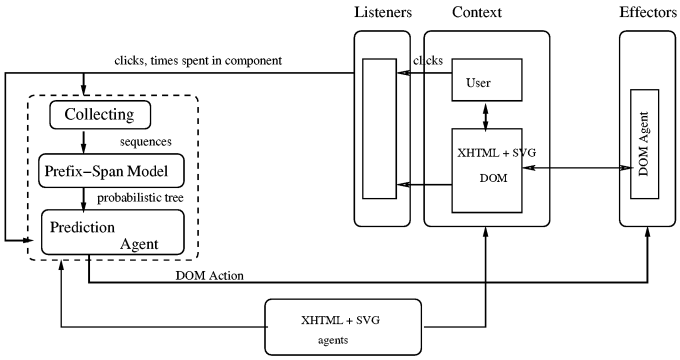


Figure 4: Unsupervised learning for adaptive technical documentation (SVG)

actions. In our navigation case we attempt to predict elements of a sequence on the basis of the preceding ones. The usual formulation of sequence prediction is : given the preceding elements s_i, $s_{i+1} \ldots s_j$, we want to predict s_{j+1}. When $i = 1$, we make predictions based on all the previously seen elements of the sequence. When $i = j$, we make predictions based on only the immediately preceding element.

Sequence learning is often based on models of feasible sequence patterns. These models can be parametric (Markov chains, hidden Markov models, recurrent neural networks) or non-parametric (structured trees or clusters of relevant sequences.)

Both models are usually learned from a training set. In our approach, we use an unsupervised learning method and a non-parametric model. The key idea is to performed an efficient mining of relevant/frequent sequential patterns. Following the prefixspan approach [Pei04], our complete training set (navigation sequences) is recursively projected into smaller projected subsets of reduced patterns , and sequential patterns are grown in each projected set by exploring only locally frequent fragments.

6 Conclusion and further work

Building on previous research of both user modeling and multimedia adaptation, we considered the design of the adaptation decision-taking engines by using machine learning methods and propose an adaptation learning platform. The strong points of our proposal are:

- natural integration of various learning types (supervised, reinforced, unsupervised) and learning models (bayesian networks, MDPs, sequence mining etc.);
- an experimental test-bed for comparing different adaptation techniques;
- combine matematical and cognitive models in order to better capture user's intentions or preferences;
- study the impact of distributing the adaptation components on the server, proxy and/or client side.
- our platform could also integrate negociation of adaptation services. This is another form of dynamicity, as negociation of services operates following a trial-and-error algorithm.

References

[Bru01] Peter Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction, Kluwer Academic Publishers*, 11(1-2), 2001.

[Bru03] Peter Brusilovsky. Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology*, 34(4):487–497, 2003.

[ccp] Composite capability/preference profiles (cc/pp). http://www.w3.org/TR/2004/REC-CCPP-struct-vocab-20040115/.

[CGV02] C. Conati, A. Gertner, and K. VanLehn. Using bayesian networks to manage uncertainty in student modeling. *Journal of User Modeling and User-Adapted Interaction*, 12(4):371–417, 2002.

[Cri04] A. Cristea. Adaptive course creation for all. In *In Proc. International Conference on Information Technology: Coding and Computing (ITCC'04)*, page 718, 2004.

[DB05] P. Dayre and H. Batatia. An original adaptive hypermedia aircraft handbook. In *In Proc. IADIS Applied Computing, Algarve, Portugal*, Feb. 2005.

[FBHF04] F. Frasincar, P. Barna, G. Houben, and Z. Fiala. Adaptation and reuse in designing web information systems. In *In Proc. International Conference on Information Technology: Coding and Computing (ITCC'04)*, page 387, 2004.

[FFD02] R.M. Felder, G.N. Felder, and E.J. Dietz. The effects of personality type on engineering student performance and attitudes. *Journal of Engineering Education*, 91(1):3–17, 2002.

[GCD02] R. Grigoraş, V. Charvillat, and M. Douze. Optimizing hypervideo navigation using a markov decision process approach. In *Proc. ACM Multimdia*, December 2002.

[PBA05] C. Piombo, H. Batatia, and A. Ayache. Reseau bayesien pour la modelisation de la dependance entre complexite de la tache, style d'apprentissage et approche pedagogique. In *In Proceedings of SETIT'05*, pages 563–570, Mars 2005.

[Pei04] Jian Pei et al. Mining sequential patterns by pattern-growth: The prefixspan approach. *IEEE Trans. on Knowledge and data engineering*, 16(11), Nov. 2004.

[TH05] Christian Timmerer and Hermann Hellwagner. Interoperable adaptive multimedia communication. *IEEE MultiMedia*, 12(1):74–79, 2005.

COMPOSING MULTIMEDIA CONTENT ADAPTATION SERVICES

Girma Berhe, Lionel Brunie, Jean-Marc Pierson
Laboratoire d'InfoRmatique en Images et Systèmes d'information (LIRIS)
Institut National des Sciences Appliquées de Lyon, France
Bâtiment Blaise Pascal (502), 20 Avenue Albert Einstein
69621 Villeurbanne, France
{Girma.Berhe, Lionel.Brunie, Jean-Marc.Pierson}@insa-lyon.fr

ABSTRACT

One of the problems with multimedia content adaptation is that, there is no a single software solution that can provide all the required adaptation tools (transformations). Generally speaking, content adaptation is multi-step process, for instance, adapting a 256-color depth jpeg image to a 2-color depth gif image can be carried out in two steps: converting 256-color to 2-color depth and converting jpeg format to gif format. To resolve the problem of a single solution, adaptation tools can be developed as software elements and deployed as services. Then, an adaptation need can be realized by combining several services. In this paper we present the construction of an adaptation graph in order to compose several adaptation services.

1. INTRODUCTION

Data access which was limited to desktop systems has been extended to several devices such as PDAs, Palmtop, Smart-phones, etc. Diversity and heterogeneity of devices poses new challenges for data delivery systems. Transferring high-resolution images to mobile devices is not feasible since most of them are unable to display high-resolution graphics. Sending text written in English to a user who can not understand English is useless. Therefore it is important to use techniques or tools that can tailor data according to the target client (Lei at al. 2001). These tools are called multimedia content adaptations. By content adaptation we mean that the system is able to scale the content being delivered to the client device based on user's preferences, device's capability and network conditions. The idea is, when the network is not good then there is no need to send big data. Some devices are unable to display high-resolution images due to their internal hardware capabilities such as supported color depth, buffer size, etc. in this case low-resolution images should be sent.

One of the problems with multimedia content adaptation is that there is no single software solution that can accommodate all the needed adaptation tools. To solve this problem, adaptation tools can be developed as software elements and deployed as services (e.g Web Services) (Buchholz 2002; Elson 2001; Rahman 2002). This new approach offers better performance and better adaptation results by integrating various adaptation services. However, there is a challenge to find an appropriate composition of the adaptation services given a composite adaptation task that can be carried out in a number of steps (adaptation tools) and given that there could be a number of possible adaptation service configurations. In this paper we present a content adaptation graph that takes into account user preferences and adaptation service qualities in order to find an optimal adaptation sequence.

There are ongoing standardization activities to develop an open framework of multimedia delivery and consumption. The work in progress by MPEG-21 is one of them. The aim of MPEG-21 is to enable the use of multimedia resources across a wide range of networks and devices (Burnett 2003). One part of MPEG-21 is the Digital Item Adaptation. DIA defines description tools for usage environment and content format features for the purpose of transparent access to the multimedia content. These tools enable the adaptation engine to make better decision in order to provide best content to the user. MPEG 21 proposes use of external adaptation tools; however, no formalism is defined for mapping the usage environment constraints and user preferences into required sequences of transformations. In our

approach, we use a rule-based constraint mapping to generate the needed transformations and then an adaptation graph is constructed by replacing each transformation with appropriate adaptation services. To find an optimal path from the graph, service qualities are used.

The remainder of the paper is organized as follows. Brief presentation of content adaptation framework is presented in section 2. In section 3 we present our methodology to construct content adaptation graph followed by prototype implementation (section 4). Our conclusion and future works is presented in section 5.

2. SYSTEM ARCHITECTURE

Figure 1 describes the architecture of our framework. Main components of the architecture and their function are summarized as follows:
Local proxies: access points to information delivery systems (content providers). They are in charge of retrieving and processing context profile (user, terminal, network), decide the type and number of adaptation processes, discover adaptation services and plan execution of the services. Core part of a local proxy is a Content Negotiation and Adaptation Module (CNAM). This module computes an optimal execution plan of adaptation process and invokes the adaptation services accordingly. The module consists of profile manager and an adaptation planner.

Client profile repository: stores user and device profile. Users can update and modify their profiles at any time. Dynamic profiles such as user location and networking condition are determined during request execution.

Adaptation service registry (ASR): it is like Universal Description, Discovery and Integration (UDDI) registry; it stores adaptation service profile (functional and non-functional) and allows for service look up. UDDI uses a Web Service Description Language (WSDL) (Web Service Definition Language 2001). A WSDL file describes the input and output of a web service. It also describes an interface to invoke a web service. However, a WSDL file provides no semantic information about what the service is used for, what its input stands for, what the effect/consequence it will bring about, how much it costs, or its execution time. The semantic information is important for automated service discovery and composition. In order to facilitate adaptation service look up we have used a very simple and structured service description that is adapted from WSDL. The service description consists of four parts (see Table 1): interface, preconditions, effects and service-quality.
Interface: contains information regarding the operation that can be invoked and its input and output parameters.
Preconditions: conditions for executing the service
Effects: results of the operation

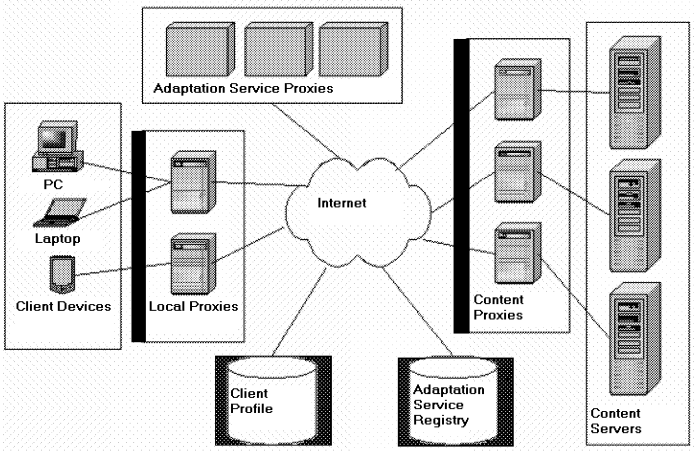


Figure 1: Distributed content adaptation framework (DCAF)

Service-Quality: provides information such as cost, execution time, etc.

Table 1: Service description

Service: jpeg to gif converter
Operation: jpegtogif
Input: inputfile, inputformat
Output: outputfile, ouputformat,
Pre-condition: inputformat=jpeg
Effect: outputformat=gif,
Service-Quality: time, cost

Adaptation service proxies (ASP): to keep the framework open and extensible, we consider external adaptation tools (adaptation services). The ASP hosts the adaptation tools.

Content proxies: provide access to content servers, formulate user request to source format, manage and provide content profile (meta-data).

Usage scenario of the architecture

Consider a physician that connects to a medical information system through the local proxy and request for patient's medical history with his handheld device (e.g. PDA). The local proxy forwards user request to content proxy and at the same time gathers client profile (user and terminal) from the client profile repository. CSCP (Held et al. 2002) like structure is used for profile description. The content proxy sends meta-data response to the request. The language and content type preference of the user is determined from the profile. In our example, we consider French as language preference and audio as content type preference of the user. One of the components of the medical record is text data. The meta-data response shows that the text data is very long and written in English. In this case, we need a minimum of three adaptation processes that are text summarization, language translation and text to audio translation to deliver the data in the current context.

tools. Therefore, we consider that the adaptation tools are developed as external software elements and deployed as services. This approach also allows building an open, extensible and interoperable framework.

3. CONSTRUCTING AN ADAPTATION GRAPH

In order to decide appropriate adaptation processes/operators, find and compose the available adaptation services, decision must be made by considering all the factors. On one hand we have context profile which consists of user preferences, device capabilities, and network conditions. On the other hand we have media profile or metadata. By matching the context profile with media profile and using a rule-based constraint mapping, we get a set of transformations called transformation prescript required to meet the constraints.

Definition: The transformation prescript is a Directed Acyclic Graph (DAG) where

- The start nodes of the graph are media sources, the intermediate nodes are operators and the end nodes are media sink or adapted media.
- The edges are media/data flow from one operator (or media source) to another operator (or media sink).

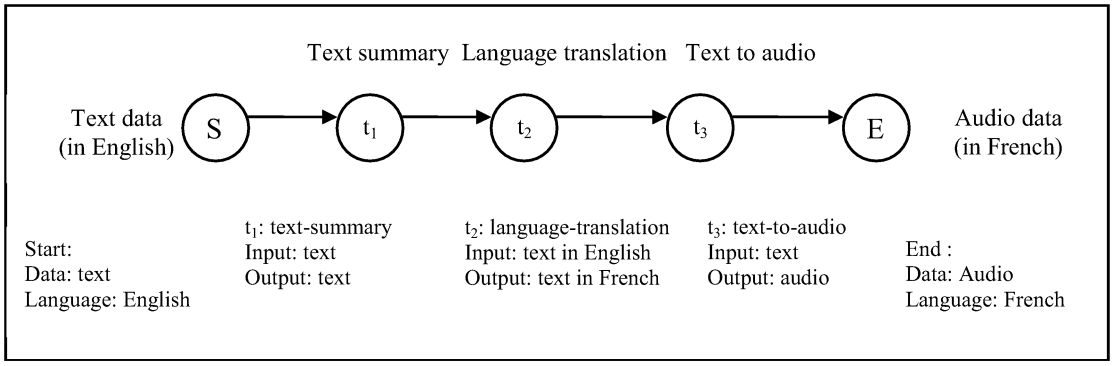


Figure 2: A transformation prescript graph

To handle such scenario the framework must be able to recognize device capabilities and contextual situations in order to use different adaptation tools such as image resizing (to display images on small devices), text-to-speech translation (to deliver in audio form), language translation (to deliver in different languages), summarization and so on. We can not assume that there exists a single software solution that can provide all the required adaptation

For the usage scenario described in section 2, the transformation prescript graph is shown in Figure 2. The transformation prescript graph provides a list of transformations need to deliver the data in the current context. It does not specify the actual services executing the transformation. In order to generate an execution plan we need to construct the adaptation graph from the transformation prescript by replacing

each transformation with appropriate adaptation service(s).

Taking a transformation prescript given in Figure 2, T= {t_1, t_2, t_3}, its adaptation graph is constructed as follows:

Step 1:
For each operator t_i, we find candidate adaptation services that can execute t_i: this will generate a set of adaptation services for each operator.
G= {<t_1, s_{11}, s_{12},>, <t_2, s_{21}, s_{22}, s_{23}>, <t_3, s_{31}, s_{32}>}, where t_1 (text summary), t_2 (language translation) and t_3 (text to audio) are realized by services {s_{11}, s_{12}}, {s_{21}, s_{22}, s_{23}} and {s_{31}, s_{32}} respectively.

Step 2:
We assume that each service accepts input data with a quality value (Q^{in}) and generates output with a quality value (Q^{out}). The quality parameters may include data format, resolution etc. To connect two services we consider quality consistency, that is, the output quality value of one service should match with the input quality value of the next service. In our prototype we limit service connectivity to data format. Figure 3 shows the adaptation graph.

Step 3:
Definition: an operator is called neutral if we can insert it anywhere in the graph but do not change the semantics of the graph (e.g. format transcodage). Neutrality of an operator is contextual, for example if the user is interested on graphic image but not on the quality of the image then converting BMP to JPEG is neutral otherwise not neutral.

Now, let us observe the adaptation graph in Figure 4. By looking at the graph we can see services s_{21} and s_{33} are connected only to one service. This is because the output (F_5) and input (F_8) formats of services s_{21} and s_{33} respectively do not match with the input and output formats of the next and predecessor services in the graph. It necessary to find one or more neutral operators so that s_{21} and s_{33} will be connected to one of {s_{31}, $s_{32,}$ s_{33}} and {s_{21}, s_{22}, s_{23}} respectively or remove them from the graph. Removing the unconnected services helps reduce the computational time of the algorithm to find the best path. But by doing this, we might exclude an interesting path through s_{21} or s_{33} in terms of global execution time from S to E.

Assuming that no neutral operator (s) is/are found to connect s_{33} then we remove s_{33} from the graph. On the other hand, suppose that the output of s_{21} is *Ms-Word* data and the input of s_{31} is *rtf* data. Let us say there are two services with neutral operators sn_1 (*Ms-Word* to *pdf*) and sn_2 (*pdf* to *rtf*) we can thus connect service s_{21} to s_{31} through a chain of the two services (sn_1 -> sn_2). By doing this we add the path {S, s_{11}, s_{21}, s_{n1}, s_{n2} s_{31}, E} that can be competitive with other paths in the graph. The complete adaptation graph is depicted in Figure 4.

Step 4:
We can see from Figure 4, there are at least four possible adaptation paths. To select an optimal adaptation path, QoS criteria defined in (Berhe et al. 2004) is used. We consider time and cost as quality criteria but other criteria such as availability, reliability, etc. can be easily included without affecting the algorithm.

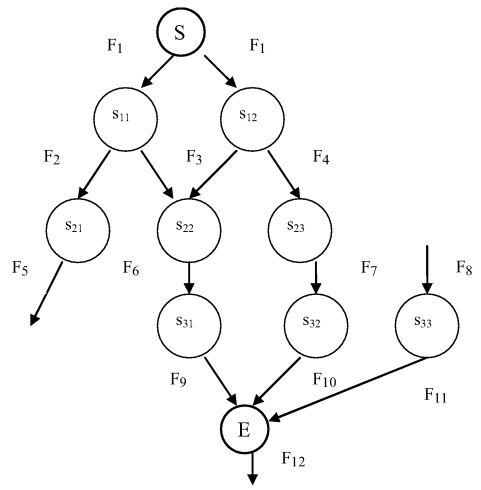


Figure 3: An adaptation graph

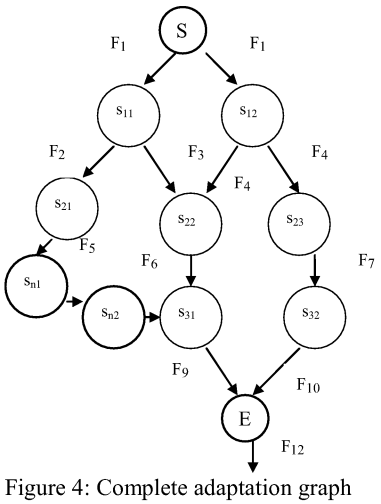


Figure 4: Complete adaptation graph

4. PROTOTYPING OF THE FRAMEWORK

We have developed a prototype implementation of the content adaptation framework in the context of medical application with some sets of adaptation services that allow us to test the viability of the ideas presented. The prototype has been developed in Java using MySQL.

For the purpose of testing the prototype, remote adaptation services such as image resizing, format conversion, color reduction, text to audio translation are developed with Web Service technologies (C# and ASP.Net). One of the external services used in the prototype is language translation service provided by WebServiceX.Net (http://www.webservicex.net).

Our experience with the prototype of the framework demonstrates openness, flexibility, and interoperability of the architecture. The architecture is able to use external services hosted in different platforms and new services can be incorporated easily through a mechanism of publishing on the service registry. Moreover, the framework can be used to seamlessly combine a set of adaptation processes in a way that guarantees conversion from arbitrary input formats to arbitrary output formats that is acceptable by the client.

6. CONCLUSION AND FUTURE WORK

In this paper we have presented a distributed content adaptation framework to deliver multimedia data in mobile environment. An important component of the framework is the construction of an adaptation graph. This graph determines the chaining of appropriate adaptation services to deliver data to a particular usage environment such as user preference, terminal characteristics, network conditions etc. To construct the graph a two-step solution is used: first determining the necessary transformation processes required to meet the usage environment, then finding the adaptation services available for each transformation. A complete adaptation graph is constructed based on input-output QoS level compatibility between successive services. The selection of an optimal adaptation path from the complete adaptation graph is done based on user's QoS criteria which includes time, cost etc.

The main advantages of the framework are openness, flexibility and interoperability which are demonstrated by the prototype. It supports integration of various multimedia transformation tools developed as Web Services.

One of the future works is to study the behavior of the path selection algorithm with respect to operator and service complexity. In addition, many enhancements can be made to the language that describes adaptation service capabilities such as input and output types. With the existing service description languages such as WSDL, there is no formal mechanism of input and output type naming. Currently each service is free to choose any name it wishes for the properties specified in its input and output formats, leading to the possibility of name space collisions. The same name may refer to different formats and vice-versa. A mechanism for avoiding such collisions is needed, such as creating a standard ontology for properties.

REFERENCES

Berhe, G., Brunie, L., Pierson, J-M. 2004. "Modeling Service-Based Multimedia Content Adaptation in Pervasive Computing". ACM Proceedings of the first conference on computing frontiers (CF'04) (Ischia, Italy, April 14 – 16), 60 – 69.

Burnett, I. Van de Walle, R. Hill, K. Bormans, J. Pereira, F (2003). "MPEG-21: goals and achievements". In Multimedia IEEE. vol.10.no. 4, 60-70.

Buchholz, S., Schill, A. 2002. "Web Caching in a Pervasive Computing World". In Proc. of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SC12002), (Orlando, FL, USA, July 14-18).

Elson, J., Cerpa, A. 2001. "ICAP-the Internet Content Adaptation Protocol. Internet Draft draft, the ICAP Protocol Group, (June) (http://www.i-cap.org/spec/icap specification.txt).

Held, A., Buchholz, S., Schill, A. "Modeling of Context Information for Pervasive Computing Applications". In Proc. of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SC12002), (Orlando, FL, USA, Jul 14-18).

Lei, Z., Georganas, N.D. 2001. "Context-based media adaptation in pervasive computing". In Proceedings of IEEE Canadian Conference on Electrical and Computer (CCECE'01), (Toronto, May).

Rahman, R., Menon, R.R. and Rafalow, L. 2002. "Enabling OPES to Use Web Service for Callout Service", Draft-rrahman-web-service-00.txt (work in progress), (May) (http://standards.nortelnetworks.com/opes/non-wg-doc/draft-rrahman-web-service-00.txt).

Web Service Definition Language. 2001. (March 15). (http://www.w3.org/TR/wsdl).

SEMANTIC WEB APPLICATIONS

WEB INTEGRATION FOR ENTERPRISE APPLICATIONS

Hristina Daskalova,
Institute of Information Technologies -BAS,
Acad. G. Bonchev, 2
1113 Sofia,
Bulgaria
E-mail: daskalovahg@abv.bg

KEYWORDS
Web services, Semantic Web Services, Integration types, Enterprise Integration

ABSTRACT

Approaches for web integration in enterprise applications are described in the paper. Different types of integration are considered. Using Web services for creating an integrated platform is discussed on three trends: communication protocols, service description and service discovery. Advantages of WebS technology are pointed out. Other industry standards towards Semantic Web are mentioned.

INTRODUCTION

Integration processes comprise different aspects of information systems (IS) starting from modeling of the architecture of complicate information systems, which allows to systemize methods for the design of corporate IS, to describe procedures for their realization and interoperation between the participating sides and finishing in using various technologies for solving the occurred problems. Software programs for industry purposes are turned to integration of enterprise application — EAI through consolidation of functional subsystems CPM (corporate performance management), CRM (customer relationship management), ERP (enterprise resource planning), PDM (product data management), SCM (supply chain management), with data bases, data repositories and other internal subsystems of enterprises. Almost all EAI projects present separate solutions in which subsystems are working with own data sets, they can not interchange data in real time and do not provide one system. The possibilities of up to now developed tools for integration of applications are connected with particular architecture; for example, Common Object Request Broker Architecture - CORBA, produced from Object Management Group; or with distributed version of Component Object Model (COM) – DCOM from Microsoft, etc. CORBA and DCOM are aimed to construct infrastructure of distributed calculations and organization of unified information field in which communication between objects is provided on object level. Objects and methods act only in framework of own architecture. Processes that are provided from the functioning of subsystems for control of business efficiency CPM and control for the delivery chains go outside the scope of the enterprise. Distributed systems of e-business B2B, e-Marketplaces and electronic commerce lead to the necessity (Vulkan 2003) of a significant expansion of corporate environment for ensuring unified information field for manufacture and trade and as a result of this expansion, the existing EAI systems probably will integrate with B2B systems.

WEB SERVICES FOR INTEGRATION

The web services (Fensel and Bussler, 2002) give a possibility for creating an integrated platform. They are tools for interoperation of information systems that are not human centralized. The web services' architecture is oriented to loosely coupled applications in which changes in given application do not reflect on others. The process of serving for requests in asynchronous regime of work between applications is done in defined time interval that could prevent web-storm. Message exchange is provided by large information blocks (*coarse grain*).
The web-services platform is based on SOAP, WSDL and UDDI standards and is divided into three parts – communication protocols, service description and service discovery. The access protocol to the objects - SOAP (simple object access protocol) supports interconnection between the web services providers. The WSDL language is a web-services description language.
The global catalog UDDI (universal description, discovery and integration) represents register for description of web-services providers.
Many organizations and committees are engaged in the development of web services standards. The leading role in this development has WS-I (Web Services Interoperability Organization) stated by Microsoft and IBM. It works in cooperation with World Wide Web Consortium and Internet Engineering Task Force. In WS-I initiative there are such firms as Oracle, BEA Systems and Hewlett-Packard
A group of developers from research laboratory in Hewlett-Packard suggest web services to be divided in the following generations: CGI (Common Gateway Interface) and Perl; Java; J2EE (Java 2 Platform, Enterprise Edition); platforms for web services (WS) (Sahai et al, 2002). The first three generations are oriented on the first place to

consumers as their requirements are defined by HTML. Arising of XML and its possibility to define syntaxes by XML Schema cause arising of web services of new type on realization platforms Sun ONE (Open Net Environment) and Microsoft .Net. The principal difference between WS and the fourth generation is in the insurance of interoperation between applications and it can be viewed as logical result of the development of distributed calculations. Web Services from fifth generations realize dynamic consolidation (unification) with semantic nets, agent technology etc. (McIlraith and Martin 2003; Trastour et al, 2003; Grosof and Poon 2004; Jennings et al 1996; He et al 2003)

To define a place of WS in modern view of processes of application integration some basic approaches and groups of technologies for solving the application integration problem which can be used separately or jointly are considered in the paper.

TYPES OF INTEGRATION

User interface integration

This means replacement of unique for every subsystem graphical or text interface with unified interface usually browser-based. Business logic remains distributed between the same ERP, CRM and SCM systems. For this replacement the functionality of information portals is enough.

Data integration

This integration supposes consolidation of data bases and data sources. It usually comes to data movement, and business logic remains distributed on initial subsystems as in user interface integration. Such a solution is ETL (Extract, Transform, Load), through which the data are extracted, transformed, filtered and stored in repository.

Business process integration

This kind of integration does not concern business logic of separate processes. There is special intermediate software which coordinates the work of separate applications by using messaging brokers that perform synchronization.

Function integration

Function integration assumes direct interoperation of cross-platform applications (application-to-application, A2A), working in the net. For these purposes the following approaches can be used:

- language tools (Cobol, C++, Java);
- API – interfaces;
- remote procedure call (RPC);
- distributed software of middleware;
- brokering architecture (CORBA, DCOM);
- remote method invocation, RMI;
- message oriented middleware, (MOM).

The full application integration supposes the use of all of these approaches.

WEB SERVICES

Besides the mentioned methods for function integration the Web Services can be used. Web Services is another technology which is applied to EAI and ensures design of integrated platform, but instead the traditional approaches it uses loosely coupled applications and is characterized by:

Simplicity – Web Services as well structured module solution is more preferable in designing, developing, adoption and monitoring than monolithic and strong coupled distributed technologies as DCOM or CORBA. There is a possibility for the blocks to be replaced and modified during the creation of general structure.

Open standards – Almost all projects in EAI are private, and Web Services is based on stack of standards.

Flexibility – Loose coupling between applications makes the project transparent, the "point-to-point" connection is avoided, which gives possibility for modifications.

Low price – Practice shows that EAI solutions based on messaging brokers usually are very expensive; meanwhile Web Services solution is expected to be more economical because of its simplicity and universality.

Scalability – Using Web Services allows the division of large systems into smaller modules and after that making selection and consolidation according the tasks.

Efficiency – Systems, build by module approach are more effective during exploitation.

Dynamics – The traditional methods of EAI are static by their nature; meanwhile Web Services can be more dynamic.

Web Services are realized by available tools on the base of stack of standards that are supposed to be changed in time. So the next generation of Web Services - Semantic Web Services (SWS) (Paolucci and Sycara 2003; Sycara et al 2003) is organized in such way that keeps the general model in which separate constituents can be replaced with new ones.

First of all the description of service functionality and the use of tools for discovery of appropriate service providers (service matchmaking) are significant (Paolucci et al 2002; Trastour et al 2001). Another stack of standards is suggested too. The service describing tools on the base of WSDL are not sufficient for the performance of software agents in Semantic Web.

The DAML (DARPA Agent Markup Language) is aimed to allow using in documents a wide spectrum of semantic concepts even equivalent and unique. DAML expands XML and RDF. Version DAML + OIL (Ontology Inference Layer) adds DAML with ontological function and allows to use common concepts and to give a sense of data in a document. It is supposed that a standard of kind of DAML-S (DAML Services: DAML-based Web Service Ontology) will replace WSDL.

Ontologies are fundamental structures for systems that operate with information and are base for Semantic Web. Such an ontology and conceptual model for description of different aspects of Semantic Web Services is so called Web Service Modeling ontology (WSMO) (Arroyo et al 2004). Having Web Service Modeling Framework

(WSMF) (Fensel & Bussler 2002) as a starting point the formal ontology and language are developed. According the philosophy of WSMF the ontologies, goals, mediators and web service are formulated. The syntaxes of the logical language is defined, which is used in WSMO. As the major component of axiom, logical expressions are used almost everywhere in the WSMO model to capture specific nuances of meaning of modeling elements or their constituent parts in a formal and unambiguous way. The semantics of this language will be defined formally by the Web Service Modeling Language (WSML) working group. Clarifying and development of WSMO are still in process.

An alternative of the registry UDDI for describing of WebS may be ebXML in the field of technology for exchange of corporate information which use Internet as transport environment. This standard is developed by OASIS (Organization for the Advancement of Structured Information Standards) and it is oriented to industry. Development of specification ebXML for e-commerce is made in joint with UN/CEFACT (United Nations Center for Trade Facilitation and Electronic Business).

Besides that IBM and Microsoft present WS-Inspection (Web Services Inspection language), which supplements UDDI, and logically continues it, completely combined with it and expands possibilities for searching in UDDI. The standard WS-Inspection for exchange information about services that are provided from different providers covers the requirements for processing and integration of information of international data bases for enterprises and e-commerce.

CONCLUSION

The full application integration supposes the use of all of the mentioned approaches. Advantages of using web service technology could lead to design and implementing of integrated platform for EAI. Adding semantics to web services description and developing of new standards in this direction is necessary for the integration of enterprise applications.

ACKNOWLEDGEMENT

This work is carried out under EU Project INFRAWEBS - IST FP62003/IST/2.3.2.3 Research Project No. 511723.

REFERENCES

Arroyo, S., Lara, R., Gomez, J. M., Bereka, D., Ding, Y., Fensel, D. 2004. Semantic Aspects of Web Services. In: *Practical Handbook of Internet Computing*. Munindar P. (Ed.) Chapman Hall and CRC Press, Baton Rouge

Fensel, D. and Bussler, C. 2002. The Web Service Modeling Framework WSMF. *Electronic Commerce: Research and Applications*, 1, 113-117.

Grosof, B. and Poon, T. 2004. SweetDeal: Representing Agent Contracts With Exceptions using Semantic Web Rules, Ontologies and Process Descriptions. To appear, *International Journal of Electronic Commerce*

He, M., Jennings, N.R. and Leung, H. 2003. On Agent Mediated Electronic Commerce. *IEEE Transactions on Knowledge and Data Engineering* 15(4) 985-1003.

Jennings, N.R., Faratin, P.,Johnson,M.J., O'Brien,P. and Wiegand, M.E. 1996: Using Intelligent Agents to Manage Business Processes. *Proceedings of the First Int. Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology,* 345-360

McIlraith, S. and Martin, D. 2003: Bringing Semantics to Web Services. *IEEE Intelligent Systems*, 18(1) 90-93

Paolucci, M. and Sycara, K. 2003: Autonomous Semantic Web Services. *IEEE Internet Computing*, (September) 34-41

Paolucci, M., Kawamura, T., Payne, T.R. and Sycara, K 2002: Semantic Matching of Web Service Capabilities. *Proc. International Semantic Web Conference,* 333-347

Sahai, Akhil et al, 2002. *The unfolding of the web services paradigm*. Hewlett-Packard Laboratories.

Sycara, K., Paolucci, M., Ankolekar, A. and Srinivasan, N. 2003: Automated Discovery, Interaction and Composition of Web Services. *Journal of Web Semantics* 1(1), Elsevier

Trastour, D., Bartolini, C. and Gonzalez-Castillo,J. 2001: A Semantic Web Approach to Service Description for Matchmaking of Services. In *Proceedings of the Semantic Web Working Symposium*, Stanford, CA, USA, July 30 - August 1.

Trastour, D., Bartolini, C. and Preist, C. 2003. Semantic Web Support for the Business-to-Business E-commerce precontractual lifecycle. *Computer Networks* 42 (5), August, 661-673.

Vulkan, N. 2003: *The Economics of E-Commerce*. Princetown University Press, Princetown, New Jersey

AUTHOR BIOGRAPHY

HRISTINA DASKALOVA was born in Plovdiv, Bulgaria and went to the Polytechnic Institute, at Saint Petersburg, Russia, where she studied "Automation and Telemechanics" and obtained her degree in 1960. She completed her Ph.D. Thesis "Construction of optimal receiver of discrete information at intersymbol interference and additive noise", at the Institute of Information Theory and Automation - Chech-Slovacian Academy of Sciences, Prague in 1974. She worked for Bulgarian Academy of Sciences, where she received Patents and Certificates for Authority in the field of information transmission.

EXISTING APPROACHES FOR MODELING AND INTEGRATION OF BUSINESS LOGICS AND PROCESSES

Hristina Daskalova
Vladislava Grigorova
Tatiana Atanasova
Institute of Information Technologies
"Acad. G.Bonchev" 2, 1113 Sofia,
Bulgaria
E-mail:daskalovahg@abv.bg, v.grigorova@abv.bg, atanasova@iit.bas.bg

KEYWORDS
Business process, Web services, Orchestration, Choreography, Grid-environment

ABSTRACT

Web-based modeling of business processes is considered. The technological chain for generating loosely coupled systems, using the technology of Web-services is described. Requirements that concern the language describing the sequence of actions in the business process, as well as the infrastructure that will realize these actions, are determined. Standards for orchestration and choreography are discussed. Integration in enterprise applications using development of software technologies on the basis of a distributed service-oriented architecture is shown on the example of IBM solutions.

INTRODUCTION

The methods for modeling business processes may be divided with respect to the moment of their release – either created before the development of Web-services or directly based on Web-services and service-oriented architectures. The Web-services give a possibility for creating an integrated platform. They are tools for interoperation of information systems that are not human centralized. There exists an opinion that the first actual languages for models description, such as ebPML, BPML, XLANG, WSFL, BPEL4WS, EDOC, XPDL and UML 2.0 will lose their independent existence, but will continue to be used in combination with the second ones. After the development of version 1.1 of the specification of BPEL4WS language and its renaming into BPEL, its positions have considerably improved in comparison with the remaining languages.
A completely new step in modeling is the release of Process Modeling Notation (BPMN), which advantages are recognized by BPEL developers as well. BPMN graphical objects are supported by a large selection of attributes, which can be directly presented in BPEL language. Future consolidation of the two mutually complementary approaches to modeling – BPMN notation and BPEL language, is expected. The purpose is to offer such form of presenting the models in business, which will be comprehensible both for the developers and for the persons who monitor the business processes. The technological chain for generating weak-connected systems, using the technology of Web-services and BPMN may include: design of the processes with the help of BPMN tools; modeling and optimization of the processes; description of the services by BPEL tools and orchestration of the services.
The elaboration of the business processes on the basis of Web-services uniting has determined two aspects of interaction and the respective introduction of the notions "orchestration" and "choreography". Orchestration concerns the process, which is realized and can interact with external and internal Web-services, while choreography is responsible for the succession of the messages generated by the participants in it. The standards of orchestration and choreography must meet some technical requirements, which make possible the application of the Web-services (Peltz 2003) in the development of business processes.

REQUIREMENTS TOWARDS BUSINESS PROCESSES DEVELOPMENT

These requirements concern the language describing the sequence of actions in the business process, as well as the infrastructure that will realize these actions:

- In order to guarantee security and universality of the computing environment, a possibility is required for asynchronous demanding of the services and the realization of asynchronous Web-services needs a mechanism for requests correlation. There exists a concept that the asynchronous services could be the core of weak-connected architectures, which use messages exchange. During the design of the interface, messages exchange is foreseen, at that not only in transmitting data to Web-services providers, but also in receiving the corresponding data. In order to react to the input message, it is necessary to know the address, where to send the generated message with some information about its correlation with the input initiated message (there is a WS-Addressing specification in BPEL for this purpose). The large integration processes suppose the participation of a human operator in the automatic process and a process including a human response is not synchronous.
- The architecture of the business process must provide control in extreme situations and preserve the transactions as well. Besides processing of errors and time-outs, the orchestrated Web-services must guarantee accessibility to the resources during the execution of time-

consuming distributed transactions. When using different methods of compensating transactions, the transactions coordinator may cancel some operations (due to the process or the user).

- The orchestration of the Web-services must be dynamic, flexible and adaptive in order to meet the altering requirements of business. For this purpose the logic of the process of the Web-services used is divided, which is enabled by the mechanism of orchestration. With the help of this mechanism, the working flow of the business process is managed, the respective Web services are invoked and the steps for their realization are determined. The approach makes possible the replacement of the services in the working flow.
- The design includes also a possibility to combine the services at a high level, based on the existing orchestrated processes. The recursive composition is realized presenting the processes as interfaces to Web-services.

STANDARDS FOR ORCHESTRATION AND CHOREOGRAPHY

XLANG (XML-based language) developed by Microsoft, was initially designed to describe successive, parallel and multi-variant working flows for BizTalk Server (Microsoft 2003). There is not any principal possibility in BizTalk Server to express the business process by traditional programming languages, but the process is defined by graphical means, which enable the determination of the logical sequence of the steps aimed at solving a certain business problem and also the description of the conditions, cycles and other characteristics in the behaviour of the business process with the help of special forms. When describing the interfaces of the Web-services, XLANG uses Web Services Description Language (WSDL), suggested by World Wide Web Consortium (W3C). The main task of XLANG consists in defining the business processes and in the organization of the messages exchange among the Web-services. This language has tools for processing extreme situations and supports long-term transactions.

WSFL (Web Services Flow Language) of Microsoft allows the description of public and private processes. It determines the data exchange, the succession of the execution (a flow model) and the expressing of each step in the flow in the form of specific operations (global model). WSFL supports WSDL interface, which enables the solution of recursive composition problems, possesses means for dealing with exceptional situations, but does not support transactions.

The United Nations Organization center for co-operation in the area of commerce and electronic business (UN/CEFACT) has developed the Electronic Business Extensible Markup Language – ebXML, on the basis of XML. This standard provides a set of program documents, facilitating the co-operation between business partners. It includes the Business Process Specification Schema (BPSS), which makes possible the determining of the choreography and the communication protocols among the Web-services.

Hewlett-Packard Company has offered a standard for modeling sequences of interactions of the Web-services – Web Services Conversion Language (WSCL), released in March 2002. This standard is discussed by the working group in choreography and Web-services of W3C.

The companies Sun Microsystems, SAP, BEA and Intalio have developed a specification of the language describing the interfaces ˜ Web Services Choreography Interface (WSCI), which expands WSDL language with the purpose to organize shared work (Arkin et al 2002). In the general case WSCI determines the choreography or the exchange of messages among the Web-services. The specification ensures correlation between the messages, rules for ranking, processing of extreme situations, transactions and dynamic interaction. It supports base as well as structural actions.

BPML (Business Process Modeling Language) is a language designed to describe the business processes on the basis of XML. Its specification is developed by the Business Process Management Initiative organization. At first BPML was intended to present the course of actions of a system controlling the business processes. Later on a possibility was added to it for WSCI support. BPML has constructions depicting the actions of the working flow of the business process similarly to BPEL (Shapiro 2002). It allows structural actions, successive and parallel processing, cycles and synchronization of the processes, base actions for transmitting and receiving of messages, requests invoking. It enables planning of the tasks and contains functions for data long-term storing. The data exchange among the participants is realized in XML format, using roles and definitions of the partners, similarly to BPEL constructions. BPML supports a recursive composition, intended to form integrated processes from components, and provides both long-term and short-term transactions. The developers can actuate some of the built-in processes and transactions. BPML has a mechanism to process extreme situations.

UML (Unified Modeling Language) is used for specification, visualization and documentation of the design process of object-oriented software systems. Diagrams are used to introduce three types of models: functional, objective and dynamic. The graphic notation in UML has textual equivalents in the object-oriented programming languages, for example JAVA.

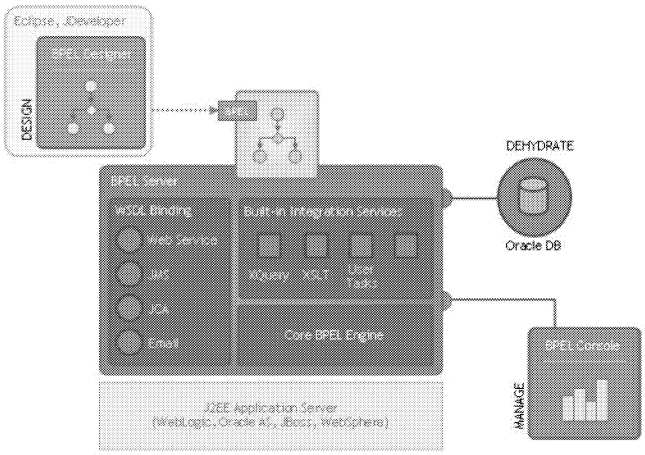
In May 2003 Microsoft, IBM, Siebel, BEA Systems and SAP have developed together a version of the specification of BPEL4WS language. This specification, called BPEL, enables the modeling of the behaviour of the Web-services in the interaction of the business processes (Weerawarana and Francisco 2002). Its constituent part is a grammar on the basis of XML, describing the logic of control in the co-ordination of the Web-services, participating in the working flow of the business process. The mechanism of orchestration in conformance with the grammar can coordinate the actions and compensate the process as a whole in case of errors occurrence. BPEL is an extension of WSDL, which determines the list of the possible operations and BPEL describes the order of their execution. WSDL indicates the public points for input and output of every BPEL process and the types of data in WSDL determine the information exchanged among the requests of the business process. WSDL allows referring to external services, necessary for BPEL process, which supports executable and abstract business processes also. The process, which is

executable, models the behaviour of the participants under a certain interaction, i.e. a given working flow. The abstract process or the business protocol is not an executable process, it defines the public messages among the participants, but it does not reflect the inner details of the working flow of the business process (the executable processes model the orchestration, while the abstract – the choreography of the Web-services).

BPEL4WS specification supports both base and structural actions. The base action is an instruction for interaction with a factor, external with respect to the process. The base actions serve the reception of requests, responds and Web-services invoking. The structural actions control the working flow of the business process as a whole, determining the sequence of Web-services demands. These actions support also the execution of the cycles and the dynamic branches. In their essence they define the main logic of programming in BPEL. Two other important elements in BPEL are the partner connections and the variables. The variables identify certain data in the messages flow. When BPEL process receives a message, it assigns a value to the respective variable, so that the next requests may refer to this data. The variables are used to control the long stored data in the course of processing the requests for Web-services. The partner link may denote any request, with which a given process interacts. Each partner relation has its corresponding role in the business process. BPEL and BPML allow defining of the executable business process, while WSCI is mainly oriented towards interaction and it requires each one of the participants to determine its WSCI interface. The protocol for joint work concerns the organization of the messages exchange among several participants in the business process, while the executable process is a given working flow and it is under the control of the separate participant. BPEL supports the executable processes and the organization of the interactions among them as well. BPML and WSCI can work together in such a way, that BPML models the execution of the business process and WSCI – the choreography of the Web-services.

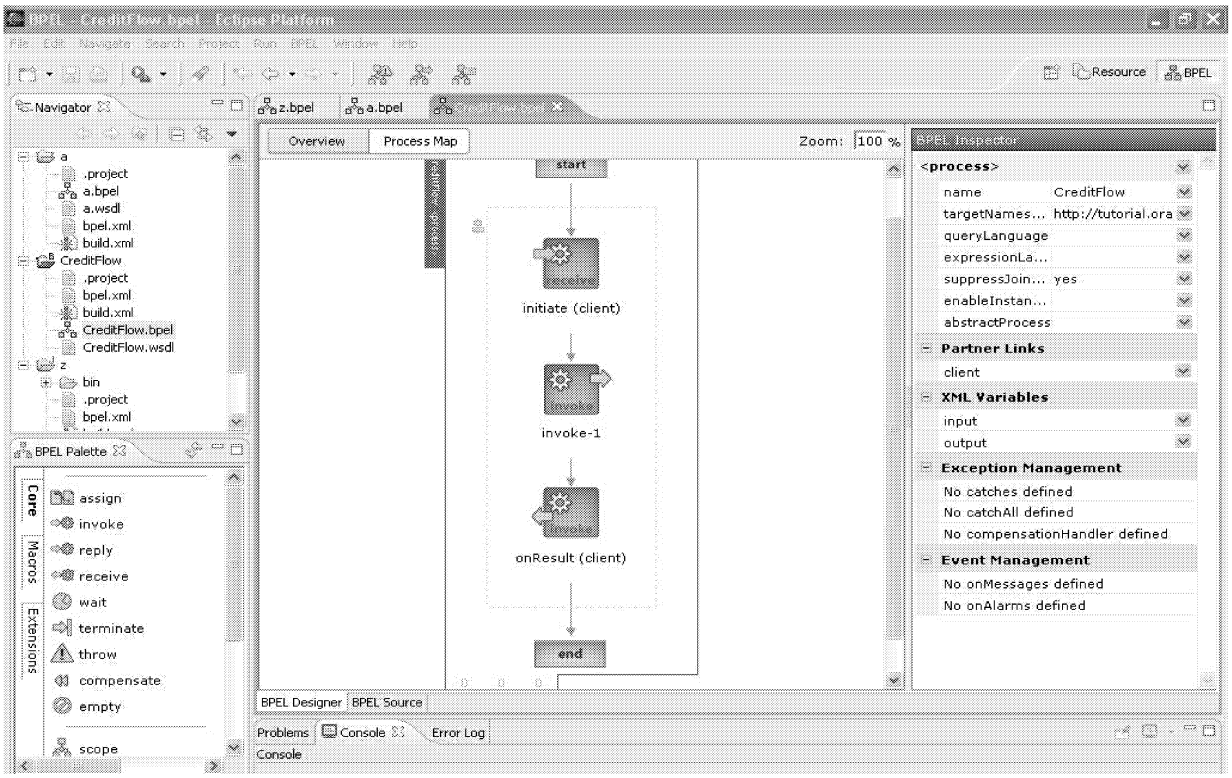
In 2004 IBM and BEA have issued a common document "BPELJ: BPEL for Java", which suggests another language – BPELJ, being a combination between BPEL and Java. The purpose is to place closer the description of the processes to software products development. BPELJ enables the including of BPEL fragments in Java-codes, called Java snippets.

The uniting of Web-services with Business Process Execution Language for Web Services increases the efficiency of the applications integration. The Web-services allow the transfer of the integration processes from a software level to the level of messages exchange among applied functions, formed as requests. Microsoft and IBM companies have a leading role in the standardization of the technology of the Web-services and they offer their own realization of the mechanisms for applications integration (for example the last version of BizTalk Server 2004). The construction of a working environment, based on Web-services passes through two stages of work with the services publishing and orchestration. The publishing of services gives opportunities to use services with the help of interface protocols. The orchestration combines the services in a united business application and accomplishes their co-ordination. The necessity in products, similar to Oracle BPEL Process Manager (Oracle 2005) is grounded on the fact that the logic of the orchestration is sufficiently complicated and must correspond to the surrounding environment. It is an appropriate tool, which gives the organizations the possibility to model and implement business processes using BPEL. The latter is recognized as an outstanding one in the organization of the orchestration and services execution in service-oriented architectures, which causes diminishment of the complexity and speeds up the realization of the integration projects. Oracle BPEL Process Manager is the first version of a BPEL-machine, completely corresponding to the standard and ready for release.



Figures 1: Oracle BPEL Process Manager

An existing functionality is used for publishing and this functionality becomes accessible through the network. There are some standards describing how to create an interface to this functionality (WSDL) and a data model (XML and XML Schema), supporting different protocols. The service thus issued may be considered as a main element (block), which can exchange input and output XML messages with its environment. Oracle BPEL Process Manager (Figures 1) comprises a BPEL Designer (Fig.2), allowing graphical interface to BPEL-processes (forming a BPEL machine), built-in integration services for connection to SOAP and other protocols, and a BPEL Console, supporting Web interface.



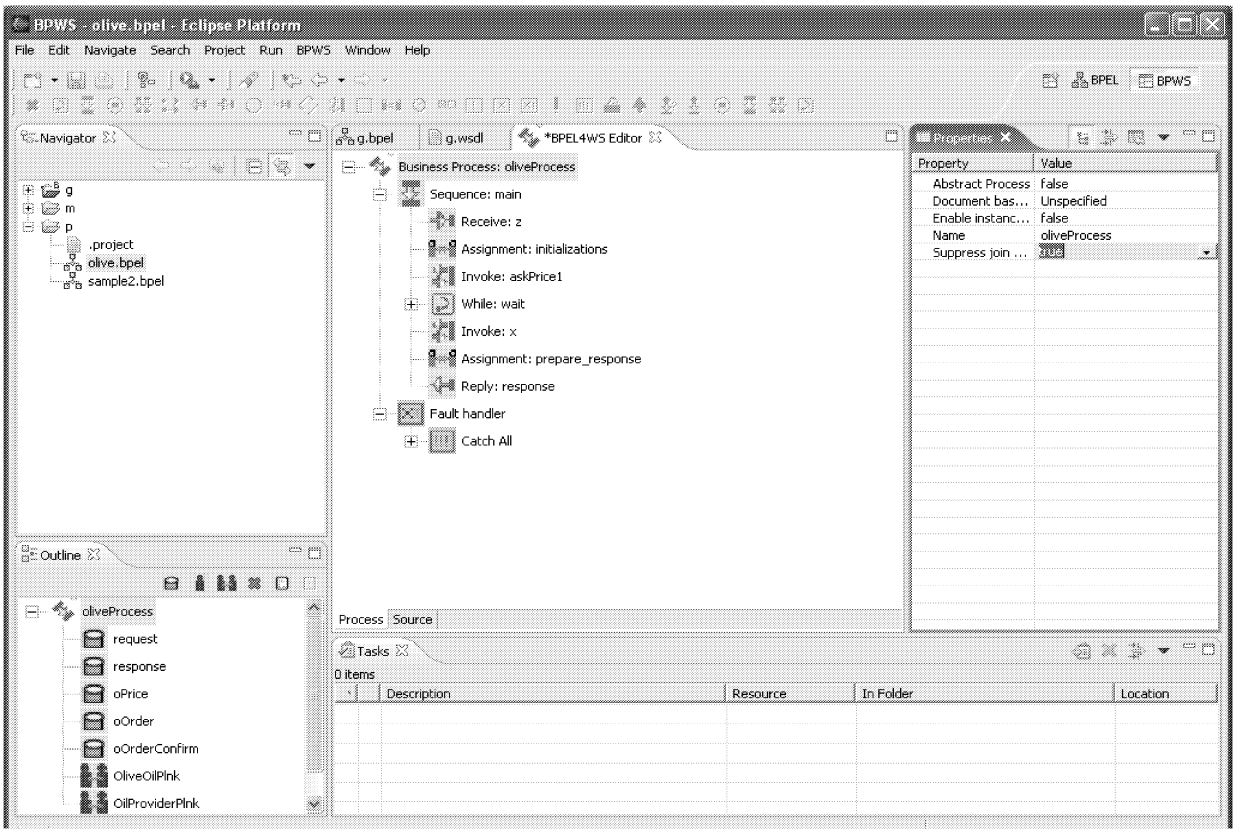
Figures 2: BPEL Process Model Designer

IBM SOLUTIONS FOR BUSINESS INTEGRATION

The strategic solution of the integration problems is the use of some software platforms tested in time and practice at industrial level, EAI (Enterprise Application Integration). The platform of IBM, WebSphere proved to be one of the leading and best functional instruments of EAI. All products of WebSphere family may be referred to the category of the integration software insurance, but in this case the solutions realized within the frames of IBM Business Integration attract special interest.

The main problems of business integration include: integration of the users experience; information integration; applications integration; integration of the business partners; processes integration. The business integration uses a set of possibilities for service-oriented architectures which comprise: organization of interactions and control of the business operations, application of the information resources in the enterprises, connection with the business partners, and availability of a common network for co-operation among decision makers solving specific problems in business. The software platform of IBM – WebSphere Business Integration (Schmidt 2003, Herness et al 2004) possesses these possibilities; it suggests a method for creating solutions in business integration that applies the modeling tools to determine the requirements towards the solution and also the necessary means for generating and connecting the service-oriented software components. IBM has declared as its priority the development of its software technologies on the basis of a distributed service-oriented architecture (SOA). The solutions of WebSphere Business Integration Server Foundation makes possible the creating and integrating of the applications within the frames of SOA. This is the first IBM product with built-in support of Business Process Execution Language – as a standard specification for realization of business logic. With the help of WebSphere Business, manifold usable services can be created on the basis of the existing Web-services and ready applications, and also combine the services connecting the business processes with the software applications (Klopmann et al 2004). IBM division Global Services proposes services connected with SOA. IBM Assessments for Service Oriented Architectures analyzes the functional and technical aspects of the planned introduction of SOA. Strategy and Planning for Service Oriented Architectures enables the determining of the business and technological possibilities necessary for the realization of the advantages of the Web-services oriented computing technologies. The Application Renovation and Integration for Service Oriented Architectures intends to re-build the old applications and make them a part of SOA. Besides this, the subdivision Component Business Modeling, uniting IBM Research and Business Consulting Services, suggests the classification of the organizations activities into activities, performed by associates, by processes or by systems.

The software platform WebSphere Business Integration integrates the internal systems of the enterprise, the systems for business partners and also the control of the business operations; it supports the solutions with integrated tools as a result of the integration of metadata in Eclipse, including Ecliplse Modeling Framework, realizing Meta Object Facility (Mukhi 2002). Thus the integration of the meta-models is maintained, which enables the exchange of solutions elements also with the help of instrumental tools.



Figures 3: BPEL4WS Editor

Ontologies are fundamental structures for systems that operate with information and are base for Semantic Web. Ontology-based Web Services for Business Integration is a semantic Web services for the industrial sector that shows service discovery, composition, and business process transformation. It supports the infrastructure for next-generation, semantic Web services middleware that enables business process integration, including business-to-business (B2B) exchange. It is a demonstration of technology that can facilitate enhanced discovery, composition, and transformation of business processes through semantic enablement and intelligent agents.

FEATURES OF THE GRID-ENVIRONMENT CONNECTED WITH BUSINESS

Primarily, there is difference between the computing grid-environment and this for business, concerning the Web-services and formed as a concept of Grid services. Standardization is being elaborated in this area within the frames of Global Grid Forum. Architecture that combines the main Grid technologies with Web-services is suggested (IBM grid 2005), intended for formation of components of an infrastructure for integration, virtualization and control of various resources in virtual organization. Despite the difficulties in the technical realization, the main characteristics of the grid-architecture and the requirements towards it are already outlined:

- Grid – this is methodology and technology uniting a large part of modern computer infrastructure (servers, data stores, applications, control systems, etc.);
- From the viewpoint of technologies, grid is not a ready product, nor a set of products, nor a component of a product or service, but a way to organize an infrastructure, based on distributed architecture and on dynamic distribution of the loading of the complete set of applications;
- From the viewpoint of the user, grid is a service, providing the necessary computing resources on a request. The Grid-system is not adjusted after altering the list of tasks, neither after altering the loading in a given problem. The complexity of the solutions is not visible for the user and he pays for the service only, not for the infrastructure built.

The Grid is built on already existing technologies and consists of a set of services, registered in repositories, which indicates where each service is accessible and where the service used must be paid for. It must be based on industrial standards that enable the operation with components of different manufacturers. The services proposed must be quick and protected. It is expected that a grid for business will be realized if the economic effect achieved gives good reason for the technical complications and expenses.

CONCLUSION

The realization of an infrastructure is expected, in which the central place will be occupied by the process directed towards business problems solution, characterized by the following: all stages of the business process are supported in the stage of their creation as well as during their execution; the components and the functions of the process

are presented as self-defining services; there is a possibility to integrate different necessary sources of information and applied functionality in the process, independently on the way of their realization and their physical location; there is a possibility to automate the flow of information and messages for events within the frames of the process; there is a possibility to specify and monitor some agreements for the level of serving all services, actuated in the process; each stage of the process may be added, removed or altered without any influence on the other stages, etc.
The tendency towards automation of the business processes with the help of weak-connected components will give the infrastructure the possibility to react with respect to the changes in business and will develop the idea for an adaptive enterprise and also make considerable steps in its realization.

ACKNOWLEDGEMENT

This work is carried out under EU Project INFRAWEBS-IST FP62003/IST/2.3.2.3 Research Project No. 511723.

REFERENCES

Arkin A. et al., “Web Service Choreography Interface 1.0”, 2002, http://www.sun.com/software/xml/developers/%20wsci/wsci-spec-10.pdf

Herness E., R. High, J. McGee, “WebSphere Application Server: A Foundation for On Demand Computing”. *IBM Systems Journal*, Vol. 43, No. 2, 2004.

IBM Grid computing, http://www.ibm.com/grid

Kloppmann M., D. Konig, F. Leymann, G. Pfau, D. Roller, “Business Process Choreography in WebSphere — Combining the Power of BPEL and J2EE”. *IBM Systems Journal, Vol. 43, No. 2, 2004.*

Microsoft, “Microsoft BizTalk Server 2004”. *Enterprise Integration*. December 2003.

Mukhi, N. “Creating processes with the BPWS4J editor”, http://www-128.ibm.com/developerworks/webservices/library/ws-bpelcol4/index.html

Oracle BPEL Process Manager, http://www.oracle.com/technology/products/ias/bpel/index.html

Peltz Ch., “Web Services Orchestration and Choreography”, *IEEE COMPUTER*, October 2003. IEEE Computer Society, 2003.

Schmidt M.-T., “WebSphere Business Integration Programming Model”. *IBM Software Group Technical Report*, March 2003.

Shapiro R., “A Comparison of XPDL, BPML and BPEL4WS” / draft document, Cape Visions, May 2002, xml.coverpages.org/Shapiro-XPDL.pdf.

The Business Rules Group. “Organizing Business Plans: The Standard Model for Business Rule Motivation”, Version 1.0., November 2000.

Weerawarana S., C. Francisco, “Business Process with BPEL4WS: Understanding BPEL4WS”, Part 1/ research report, *IBM developerWorks*, Aug. 2002, www-106.ibm.com/developerworks/webservices/library/ws-bpelcoll/

AUTHOR BIOGRAPHY

HRISTINA DASKALOVA was born in Plovdiv, Bulgaria and went to the Polytechnic Institute, at Saint Petersburg, Russia, where she studied "Automation and Telemechanics" and obtained her degree in 1960. She completed her Ph.D. Thesis "Construction of optimal receiver of discrete information at intersymbol interference and additive noise", at the Institute of Information Theory and Automation - Chech-Slovacian Academy of Sciences, Prague in 1974. She worked for Bulgarian Academy of Sciences, where she received Patents and Certificates for Authority in the field of information transmission.

INFRAWEBS SEMANTIC WEB UNIT FOR DESIGN AND COMPOSITION OF SEMANTIC WEB SERVICES

Tatiana Atanasova
Institute of Information Technologies
Acad. G. Bonchev 2, 1113 Sofia
Bulgaria
atanasova@iinf.bas.bg

Gennady Agre
Institute of Information Technologies
Acad. G. Bonchev 29-A, 1113 Sofia
Bulgaria
agre@iinf.bas.bg

Joachim Nern
big7.net GmbH & Aspasia
Knowledge Systems
Germany
E-mail: nern@aspasia-systems.de

KEYWORDS
Semantic Web Services, Ontologies, Integrated System, Distributed Decision Support, Semantic Web Unit

ABSTRACT

A framework for design, composing and execution of Semantic Web Services with decision support – Semantic Web Services Unit (SWU) is presented. All phases of Semantic Web Service usage process, namely: discovery, composition, mediation, execution, monitoring, and compensation are reflected in SWU. The architecture of SWU enables the user to be "broker, provider and client" at the same time within one integrated system. This can be achieved by realizing SWU as a multi-agent system that is enriched with ontology for Semantic Web Services description from particular domain. A motivation of SWU design is outlined. Then the analysis of requirements for the developing of such Semantic Web Unit is presented. Specifications on the base of the general requirements for INFRAWEBS SWU are described.

INTRODUCTION

The World Wide Web is emerging for a variety of resources that are available as Web services. Web services are distributed, platform independent, reusable pieces of software that are accessible via the Web, providing any type of functionality that software can deliver (Arroyo et al 2004). With the growing popularity of Web services there arise issues of finding relevant services, especially with the possibility of the existence of thousands of Web services. Current standards in Web service technology like UDDI WSDL and SOAP have focused on operational and syntactic details for implementation and execution of Web services, which limits the service search mechanism to keyword-based searches. Realization of the full potential of Web services requires further technological advances in such areas as services interoperation, service discovery, service composition and orchestration. A possible solution to all these problems may be converting Web services to Semantic Web services by application of Semantic Web technologies. Semantic mark-up can be used to automate the tasks of discovering, executing and composing Web Services and interoperating with them (Fensel et al 2002). The achievements of semantic web technologies can keep the intervention of the human user to the minimum.
The state of the art of technologies around Semantic Web Services (SWS) represents that the full semantic usage of Services has many obstructions. At present the establishment of semantic based software components is reserved to specialists and programming experts. There is a lack of semantic based systems that use decision supported techniques, which allow automate process of web service designing, composing and maintaining.
In this paper an architecture and specification of semantic web unit for design and composition of semantic web services is introduced.

MOTIVATION OF SWU DESIGN

Web Services provide the basis for the interaction and coordination with distributed applications. Automated support is necessary in order to decrease time, effort and cost of development (and maintenance) of web services design and composition. But automated knowledge processing and adequate services design, discovery, selection, composition and execution can not be achieved without adding semantic mark-up to web services description and using advanced knowledge processing techniques. This semantic mark-up enriches web services description with information on what the service can do and how. In such way the additional information facilitates service discovery, selection and invocation.
The semantic description of Web Services has to cover (Verma et al 2000):

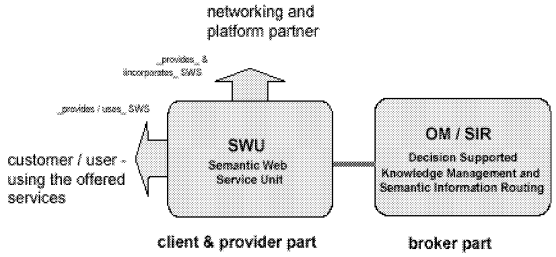
- **Data/Information Semantics**
  - What: definition of data in input and output messages of a web service
  - Why: for discovery and interoperability
  - How: by annotating input/output data of web services using ontologies

– **Functional Semantics**
  - representing capabilities of web service for discovery and composition of Web Services by annotating operations of Web Services as well as provide preconditions and effects
– **Execution Semantics**
  - representing the execution or flow of a services in a process or operations in a service for analysis (verification), validation (simulation) and execution (exception handling) of the process models
– **QoS Semantics**
  - represent qualitative and quantitative measures of Web process
  - Qualitative includes security, transactions
  - Quantitative includes cost, time etc.
  - Business constraints and interservice dependencies (Domain and application ontologies)

It is reasonable to construct platform for design, composing and execution of semantic web services associated with particular domain in separate unit – Semantic Web Services Unit (SWU). Each of these units is relying on set of application ontologies, Semantic Web Services with associated metadata and business logic from the organizational memory.

ANALYSIS OF REQUIREMENTS

The SWU unit bases its operation by providing semantic web services of a "library of knowledge", representing the domain (organizational) phenomena from different sides. The idea of linking business process to organizational rules and business objectives in a formal manner has attracted the attention of numerous researches (Sirin et al 2003), (Narayanan and McIlraith 2002). The technology development puts these investigations on a new level.
The current INFRAWEBS research project (Nern et al 2004) may be seen as an important step in the direction of the semantic services creation, accession and reusing within the Semantic Web since one of its objectives is: *Development of a full-life-cycle software toolset for creating and maintaining Semantic Web Services supporting specific applications.*



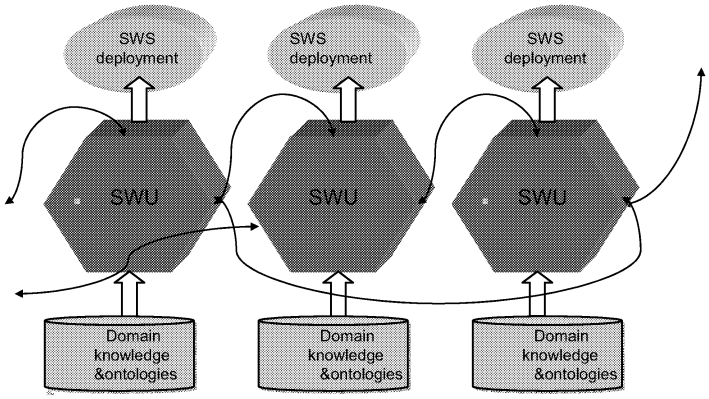
Figures 1: SWU in the framework of INFRAWEBS

One of the main components of INFRAWEBS project is SWU – Semantic Web Services Unit. In the framework of the project it is intended that SWU *acts in two directions: it provides web services to the platform and collaboration partner and to the customer of the entity. The main B2C / B2B process is conducted via this channel* (Figures 1).

The following challenges for developing of SWU have to be taken into account:

- Converting Web Services from available descriptions and Domain knowledge (organizational memory) to the semantic ones;
- Composition of Web Services, combining and orchestrating them in order to deliver added-value services;
- Dependencies that arise when a service integrates with external services and becomes dependent on them;
- Integration Processes - processes of several business partners situated on different locations have to be integrated with each other;
- Flexibility that reflects on ever-changing business environment, the SWU needs to deal with changes in the functionality, displacement or use of web services;
- Security - Semantic Web Services might have several security mechanisms depending on the environment where they are deployed.

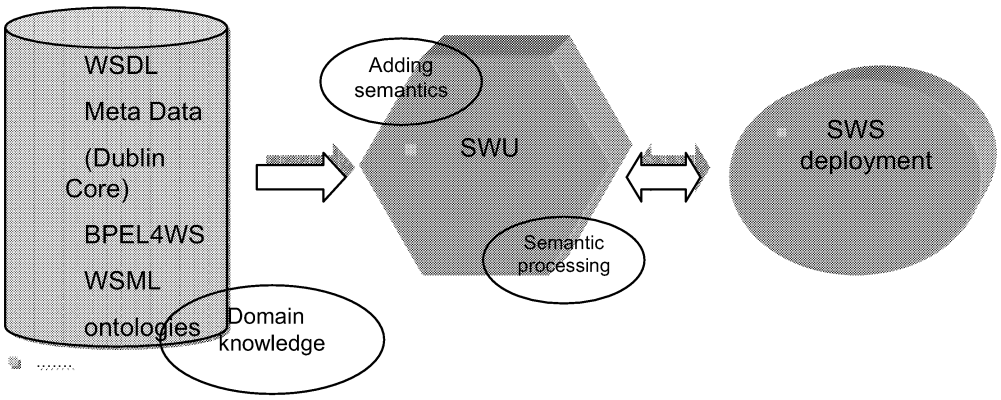
These challenges are not solved for traditional Web Services. The appropriate solution may be found using semantic approach to description and processing of Web Services (Figures 2).



Figures 2: Interoperation between SWU Modules in INFRAWEBS

Different domains as data sources provide different types of Web Services, and every SWU is associated with its domain ontology, so (Figures 3):

- sophisticated tools for selective and context-sensitive Web Services extraction and SWS composition are necessary. Such tools have to be able to bridge the gap in structure and semantics of the respective Web Services and knowledge sources using domain-specific ontologies.
- it is desirable to process the SWS in a distributed fashion. This calls for efficient decision making algorithms with performance in a distributed setting.
- it is need to access to autonomous Web Services with appropriate privacy and security policies. Thus there is a need for decision making algorithms in such environments.



Figures 3: SWU input/output

- algorithms that can update the results of design, composition and execution of dynamic Web Services have to be provided.

General requirements for SWU may be summarized as follows:

- heterogeneity integration (mediation),
- general management functionality,
- scalability,
- support for rules that govern the selection, integration and composition of Semantic Web Services,
- communicating with Semantic Web Services in a specific manner
- security.

EXISTING APPROACHES

Web service protocols and data are machine readable, but they do not provide the necessary level of "understanding" of the web content.

At present:

- Composition of Web Services is static; it is made in design-time.
- Dynamic service discovery, run-time binding, analysis and simulation are not supported directly.

Beside that most of the existing approaches focus on the service side management of web services. They allow developers to build and deploy Web Services and also provide some management capabilities such as load balancing, concurrency, monitoring, error handling, etc. Very few efforts provide support for the client applications that want to integrate different third-party web services and manage them.

Framework for design and (semi) automatic composition of SWSs

A framework for design and (semi)-automatic composition of SWSs at a language-independent and knowledge level on the base of Problem-Solving Method approach is presented in (Corcho et al 2003). This framework is based on a stack of ontologies that (1) describe the different parts of SWS; and (2) contain a set of axioms that are really design rules to be verified by the ontology instances. Based on these ontologies, design and composition of SWSs can be viewed as the correct instantiation of the ontologies themselves. Once these instances have been created they will be exported to SWS languages such as OWL-S. In the ODE SWS framework (ODE SWS 2004) a service is described at the knowledge level using the Problem-Solving Method-based approach for SWS modelling.

WSMO and OWL-S

WSMO (WSMO 2004) and OWL-S (OWL-S 2004) are main initiatives to describe Semantic Web Services, aiming at describing the various aspects related to Semantic Web Services in order to enable the automation of Web Service discovery, composition, interoperation and invocation. WSMO covers aspects (e.g. mediation) that are not even considered in OWL-S and that increase the applicability of WSMO in a real setting. Web Service Modelling Framework (WSMF) provides the model for developing and describing web services and their composition on the base of WSMO. OWL-S is feature-driven rather than concept-driven; WSMF provides central conceptual pillars, but no conceptual architecture (Priest 2004).

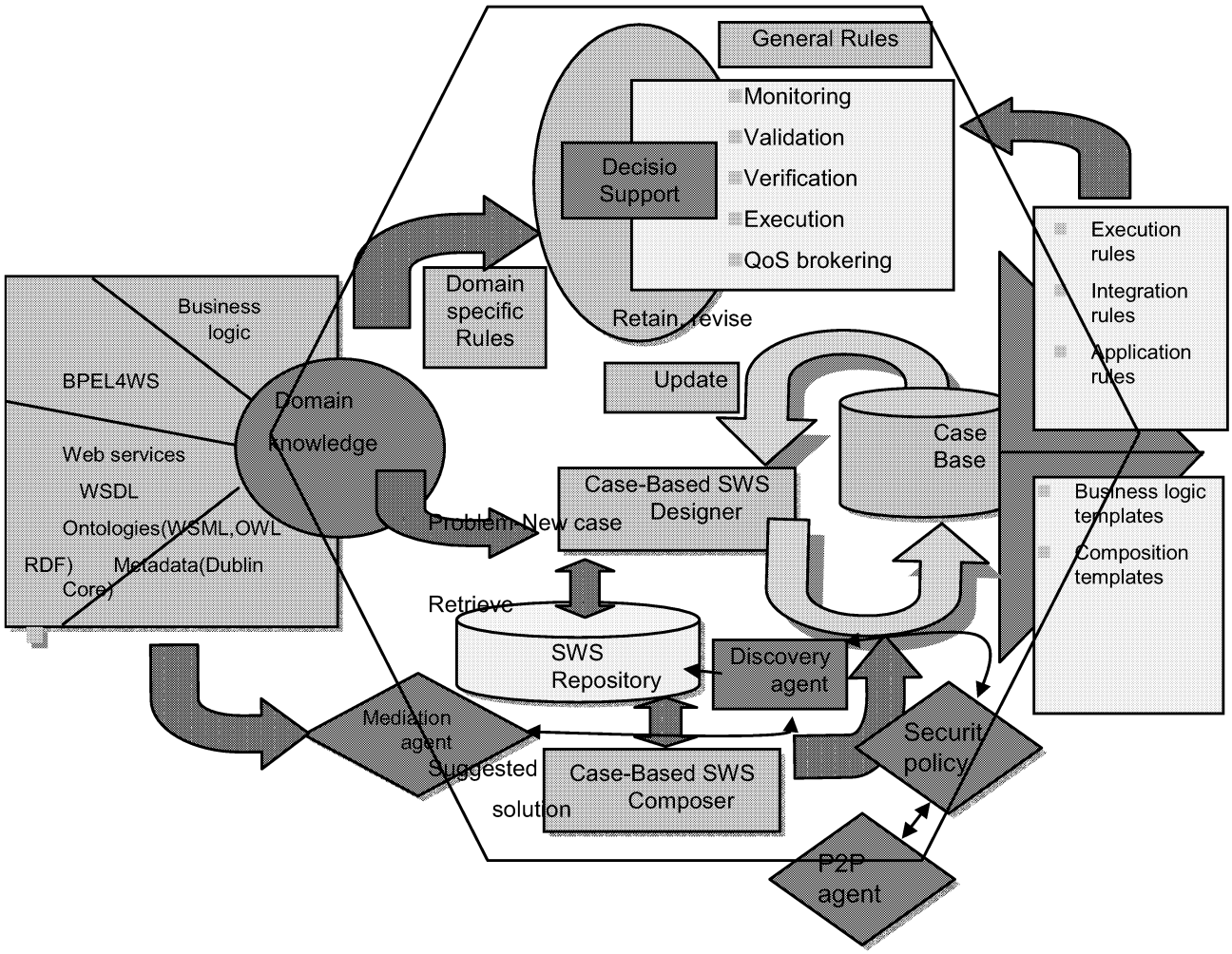
DESIGN OF SEMANTIC WEB UNIT

Within the framework of INFRAWEBS project *SWU has to be designed to enable the user to be "broker, provider and client" at the same time – within one integrated system. There will be not necessary to use different software components or packages – the user will be embedded in an environment, which allows to change the action type and modus according to the given user specific demands. The designing and composing process has to be a comprehensively decision supported, especially under the aspect of "learning from the past"* (Nern et al 2004).

SWU is a semantic and ontology based environment and it is aimed to be a base for operating with semantic web services. SWU is an infrastructure that ensures handling and maintaining of web services. It is a client & provider component, which is intended to be designed as a workflow and knowledge sharing component.

Semantic Web Services are provided in the network, and then they may be incorporated in other services by composing process. It is assumed that the web services should be able to play as the role of the requester as well

the role of the provider. The roles may switch during the same transaction.
Technically SWU is a multi-agent ontologically-based interoperable environment. Semantic description techniques are part of agent based computing and will help external agents to understand both functionality and internal structure of SWSs to be able to discover, compose, and invoke SWSs (Hendler 2001).
The conceptual architecture of SWU with decision support is shown on Figures 4. It consists of several main blocks: Case-Based SWS designer, Case-Based SWS Composer, Execution module with rule-based decision support, SWS repository and agent system. SWU has its own case-based memory, ontology of specialized domain knowledge and library with general situations about errors during planning associated with possible recipes for recovering. SWU is based on service-oriented knowledge.
Case-based memory is a registry of services, which contains abstract service templates. The registry is hierarchically designed. On the upper level it is a dynamical memory.
When a client send a request then the searching is started first in the own SWU repository that contains local specialized registry of SWS. It is proposed to realize the designer and the composer of SWS as Case Based Reasoning SW services development tools with decision-making capabilities. Storage and retrieval of designing/composing experience patterns are provided by means of self-organizing case base.



Figures 4: Conceptual Architecture of SWU

ARCHITECTURE OF SWU

All six phases of Semantic Web Service usage process (Arroyo et al 2004), namely: discovery, composition, mediation, execution, monitoring, and compensation are reflected in SWU that serves as client and provider of SWS:

SWU as a Platform to configure of semantic web services in provider application

Provider of services designs SWS from the domain knowledge. All knowledge artifacts from the organizational phenomena influents on constructed semantic web service fulfilling service non-functional properties, associated ontologies, post- and pre-conditions, assumptions and effects.

SWU as a Platform to control the integration and configuration of web services in client application.

It is necessary to define a library of reusable aspects that would allow the application developer to dynamically instantiate and configure the needed aspects to deal with different SWS concerns. These reusable aspects can be seen as generic templates that can be customized and integrated "on demand" to accommodate to service requirements.
For ensuring the functionality of broker, client and provider the SWU has to consist of the following modules:

- **Semantic repository**
- **Module for SWS design**
- **Module for SWS composition**
- **Distributed Decision Support (DDS):**
  - Monitoring Module is required as several properties of services might need observation over time.
  - Module for analysis (verification),
  - Validation Module (simulation)
  - QoS-Broker
- **Execution Module,**
- **Security & privacy**
- **P2P agent**
- **Selection Module** considers different service properties.
- **Discovery agent**
- **Mediator** architectures for automated composition, control flow and data flow based on requirements.

DDS module contains service oriented rules. These rules are derived from the application requirements and are based on the non-functional properties of services and QoS. They govern the selection of services specifying under which conditions the services have to be selected. The nonfunctional properties of services also have to be analyzed to instantiate the necessary management aspects.
SWS Design and Composition modules SWS design consists in specifying the descriptive, functional, and structural features of a service. SWS composition, moreover, deals with the combination of different semantic services to obtain a new service.

CONCLUSION

Using the semantic web service approach, in conjunction with knowledge representing ontologies, SWU accomplishes automated knowledge processing and adequate services design, discovery, selection, composition and execution.
The requirements for the SWU can be fulfilled by:

- case-based reasoning for SWS design and composition,
- using knowledge of conversation patterns supported by the service,
- using Rule Base for execution monitoring,
- simulation techniques, validation and verification,
- adding of functionality/ preconditions/ effects of the participating services,
- privacy policy.

Some problems need to be solved: how to efficiently discover Web services – based on functional and operational requirements – and how to facilitate the interoperability of heterogeneous Web services.
SWU is aimed to assist fully networked enterprises in which Semantic Web Services has to be discovered, invocated, negotiated, selected, composed, executed, and validated with integrated knowledge and inference. Agents have to execute tasks on behalf of a decision process. Supporting Semantic Web Processes on multi-enterprise require addressing heterogeneity/integration, scalability, dynamic change and performance challenges. Semantics is seen as the key solution to address these challenges.

ACKNOWLEDGEMENT

This work is carried out under EU Project INFRAWEBS-IST FP62003/IST/2.3.2.3 Research Project No. 511723.

REFERENCES

Arroyo, S., Lara, R., Gomez, J. M., Bereka, D., Ding, Y., Fensel, D. 2004. "Semantic Aspects of Web Services". In: *Practical Handbook of Internet Computing*. Munindar P. (Ed.) Chapman Hall and CRC Press, Baton Rouge.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., and Lama, M. 2003. "An Environment for Development of Semantic Web Services". *Proc. IJCAI-2003 Workshop on Ontologies and Distributed Systems*, 13–20. Acapulco, México

Fensel, D., C. Bussler, Y. Ding, B. Omelayenko. 2002. "The Web Service Modelling framework WSMF". *Electronic Commerce Research and Application*, 1(2).

Hendler, J. 2001. "Agents and the Semantic Web". *IEEE Intelligent Systems,* 16 2:30–37

Narayanan, S. and McIlraith, S. 2002. Simulation, Verification and Automated Composition of Web Services. *Proc. 11th Int. World Wide Web Conference WWW-2002*, 77-88. Hawaii, USA

Nern, H.-J., G. Agre, T. Atanasova, J. Saarela. 2004, "System Framework for Generating Open Development Platforms for Web-Service Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems - INFRAWEBS II. *WSEAS TRANS. on INFORMATION SCIENCE and APPLICATIONS*, 1, Vol. 1, 286-291.

Preist, Ch., 2004 "WSMO-Full – Some thoughts", 15-03-2004, http://www.wsmo.org/

Sirin, E., Hendler, J., and Parsia, B. 2003. "Semi Automatic Composition of Web Services using Semantic Descriptions". *Proc. ICEIS-2003 Workshop on Web Services: Modeling, Architecture and Infrastructure.*

Verma, K., A. Sheth, F. Curbera. 2000. "Service Oriented Architectures and Semantic Web Processes" Technical Report 03-008 LSDIS Lab, Dept of Computer Science, University of Georgia.

MIGRATING FROM WEB SERVICES TO SEMANTIC WEB SERVICES: INFRAWEBS APPROACH

Gennady Agre[1], Tatiana Atanasova[1], Joachim Nern[2]
[1]Institute of Information Technologies – BAS, Acad. G. Bonchev 2, 1113 Sofia, Bulgaria
{agre, atanasova}@iinf.bas.bg
[2] big7.net GmbH & Aspasia Knowledge Systems, Germany
nern@aspasia-systems.de

KEYWORDS: Semantic Web Services, Web Service Modeling Ontology (WSMO), Multi-Agent Systems

ABSTRACT
A *semantic* Web service may be created either by converting it from an existing *non-semantic* Web service or by composing of several existing *semantic* Web services. The paper describes the architecture of the INFRAWEBS SWS Designer – a tool for semiautomatic converting of non-semantic Web services to semantic ones. A general conceptual architecture of the Designer, which does not depend on a specific framework used for describing semantic Web services, is specified. It is also proposed a functional architecture of the Designer and the mapping between both architectures is discussed.

INTRODUCTION

Web services define a new paradigm for the Web in which a network of computer programs becomes the consumers of information. However, Web service technologies only describe the syntactical aspects of a Web service and, therefore, only provide a set of rigid services that cannot adapt to a changing environment without human intervention (Fensel & Bussler 02). Realization of the full potential of the Web services requires further technological advances in the areas of service interoperation, discovery, composition and orchestration. A possible solution to all these problems is likely to be provided by converting Web services to *Semantic* Web, which are self-contained, self-describing, semantically marked-up software resources that can be published, discovered, composed and executed across the Web in a task driven semi-automatic way (Arroyo et al. 04). Semantic Web services (SWS) can constitute a solution to the integration problem, as they enable dynamic, scalable and reusable cooperation between different systems and organizations. There are two major initiatives aiming at developing world-wide standard for the semantic description of Web Services. The first one is OWL-S (OWL-S 04) aiming to enable automation of web service discovery, invocation, composition, interoperation and execution monitoring by providing appropriate semantic descriptions of services. The second one is Web Service Modelling Ontology (WSMO) (Roman et al. 04) intending to create an ontology for describing various aspects related to SWS and to solve the integration problem.

The next, more technology-oriented step in the process of development of SWS, is proposed in the current INFRAWEBS European research project (Nern et al. 04). The primary project objective is to develop an ICT framework, which enables software and service providers to generate and establish open and extensible development platforms for Web Service applications. Such a platform consists of linked INFRAWEBS units, whereby each unit provides tools and adaptable system components to analyse, design and maintain SWS within the whole life cycle. One of the concrete goals of the project may be formulated as *developing a full-life-cycle software toolset for creating and maintaining SWS supporting specific applications based on WSMO framework.* An important part of an INFRAWEBS unit is a Semantic Web Unit (SWU) – a collaboration platform and interoperable middleware, which is designed to be a semantic and ontology-based module handling and maintaining SWS (Atanasova & Agre 04). The present paper describes the architecture of a part of SWU - SWS Designer, which aims at *semi-automatic design-time construction of WSMO-based SWS from available descriptions of existing Web services and domain knowledge.* The structure of the paper is as following – the next section presents an analysis if the existing approaches to the problem of designing SWS and infers some general requirements for INFRAWEBS SWS Designer based on such analysis. The Section 3 describes basic design solutions and general directions for developing the tool. The next two sections are devoted to description of the Designer architecture. The last section is a conclusion.

EXISTING APPROACHES

Although up to now there are no fully operational tools for converting Web services to WSMO-based semantic services, there are several approaches addressing the problem of adding semantic to existing WSDL-based Web services in order to convert them into SWS. WSDL description of a Web service (WSDL 03) provides a way for service providers to describe the basic format of Web service requests over different protocols. WSDL description supports information exchange between Web services, but it is not reach enough to specify the *semantics* of the service composition or the interaction protocol needed for such composition. In general case, the semantics to be added to a Web service, may be classified into the following categories:

- *Data Semantics* – semantics of input and outputs of the Web service providing a relative mapping to existing ontologies;
- *Functional Semantics* – semantics describing what does the service do;
- *Execution Semantics* – semantics describing how the service may be verified and executed;
- *Discovery Semantics* – semantics of service capabilities used for service discovery;
- *Selection Semantics* – QoS and other non-functional parameters associated with the service that may be used for selecting the particular service among other similar services.

The existing approaches to the problem described above may be analysed from different points of view. We will use two of them – the first one reflects what types of semantics an approach tries to automatically provide for; the second – what methodology is used for description of semantic Web service.

The earliest approach to the problem of semi-automatic translation of a WSDL-based Web service to DAML-S–based SWS is described in (Paolucci et. al. 03). Their *WSDL2DAMLS* converter automatically generates *Grounding* ontology of the service (Execution semantics) and partially creates the *Process Model* (Functional Semantics) as a set of atomic processes. The designer of a SWS should manually construct the complete Service profile based on process control and data flow extracted from text or other files accompanying the WSDL description of the service. The converter also perform a preliminary step to the addition of Data Semantics by converting existing complex XSD types into corresponding DAML-S concepts, which are not related to ontologies available in the Semantic Web. The designer of a SWS should relate such concepts to existing ontologies manually again. She should also complete the *Service Profile* by selecting its inputs and outputs, adding its pre-conditions and effects (Discovery Semantics) as well as to describe non-functional properties of the service (Selection Semantics).

OWL-S Editor (Scicluna et al. 04) is a tool for designing SWS, which facilitates the process of describing the Functional Semantics of a service. The tool extends WSDL2DAMLS by means for graphical constructing OWL-S Process Model. The editor implements a visual composer, which uses the UML Activity Diagrams (UML 03) to enable a visual composition of service functional model and the automatic generation of the process markup. This abstracts away the underlying structural details of the *Process Model* ontology enabling a fast and easy creation of a composite service. The user still has to know the basic key features behind OWL-S, but there is no need for an in-depth understanding of how these concepts are expressed in the markup.

OWL-S API (Sirin 04) provides a Java API for programmatic access to read, execute and write OWL-S service descriptions. The API supports the reading of different versions of OWL-S descriptions. The API provides an *Execution Engine* that can invoke *Atomic Processes* that have WSDL groundings, and *Composite Processes* that use control constructs *Sequence, Unordered*, and *Split*. Executing processes relying on conditionals such as *If-then-else* and *Repeat-until* is not supported in the default implementation; however they can be extended to handle these constructs if the application that uses the OWL-S descriptions has a custom syntax and evaluation procedure for the conditions. The API is incomplete in the areas where the OWL-S specification is incomplete, such as the definition of preconditions and effects. The names for packages, interfaces and methods in Java are chosen to match the names of ontologies, classes and properties in OWL-S ontologies.

An interesting approach for semi-automatic adding the Data Semantics to WSDL descriptions is proposed in (Patil et al. 04). It deals with the problem of relating concepts of Web services, mentioned in WSDL files, to domain specific ontologies. The proposed algorithm tries to match each concept from WSDL to all concepts from a user-defined set of ontologies (*Ontology Store*) and returns the degree of similarity between them, which is calculated based on structural similarity of the compared concepts as well as on linguistic similarity of their names. The linguistic similarity reports on equality of names, equality of synonyms and results of NGram algorithm (Damashek 95). The designer of a service can accept or reject the proposed mappings based of the calculated degree of similarity. The constructed SWS is categorized by the ontology, which has received the highest overall similarity with concepts used in WSDL-description of the service. Furthermore, the SWS is described according Meteor Framework for Semantic Web Process Composition (Meteor-S 04), in which the concepts of Service Profile and Service Model of *DAML-S* are extended with facilities to specify the composition structure and to identify the sub-services' properties, capabilities and QoS constraints.

In the ODE SWS framework (ODE SWS 04) a service is defined at the knowledge level by means of Problem-Solving Methods the service is associated to. Such methods are claimed to be the knowledge components that are reusable among different domains and tasks. The ODE SWS ontology is built on the top of the OWL-S ontology and uses all its concepts. The first software release of ODE SWS is the ODE SWS Designer that provides a user-friendly graphical interface with which the user is completely unaware of the instantiation of the SWS ontology.

WSDL2Agent is a tool facilitating the migration of existing Web services to the SWS environment (Varga 04). The tool automatically generates a Protégé project, which is further enhanced by the service designer with the needed semantic information (Data, Functional, Discovery and Selection Semantics). Then the project can be exported into OWL format, from which the skeletons of the elements of the Web Services Modeling Framework (Fensel et al. 02) can be derived.

SWWS Studio (SWWS 04) is a research tool for creating all elements of Web Service Modeling Ontology (WSMO). It comprised of two components: a graphical designer for

modelling WSMO specifications (Service Editor) and a graphical tool for modelling compositions of services (Service Composer).The Service Editor provides a graphical interface for creating definitions of all WSMO elements - ontologies, goals, mediators and services. The Service Editor can be also used as a front-end browser to a remote WSML-based service registry. The Service Composer is used for defining service orchestrations, e.g. the interaction with other (*non semantic*) services in order to achieve the service functionality. The mappings between outputs and inputs are specified with the help of so called containers - state variables that can be used for storing data values from output ports and retrieving the values into the input ports. At present the orchestration definitions created by SWWS Studio can be exported as an OWL-S Process Model.

REQUIREMENTS TO SEMANTIC WEB SERVICE DESIGNER

The analysis of the existing approaches for the problem of converting existing Web services to semantic ones has shown the following:

- The process of adding *Data Semantics* to a Web service can be significantly facilitated by application of algorithms for semantic annotation;
- The process of adding *Functional Semantics* to a Web service can be automated when the BREL4WS description of the web service is available. Otherwise, user-friendly tools for describing and translating the service business logic are necessary;
- The process of adding *Execution Semantics* to a Web service can be fully automated based on service WSDL description;
- The process of adding *Selection Semantics* to a Web service may be partially automated by using the service provenance information contained in UDDI (UDDI 00);
- The process of adding *Discovery Semantics* to a Web service still require the service designer intervention and some tools for facilitating this process should be invented.

These conclusions have allowed us to formulate the following *general requirements* to INFRAWEBS Semantic Web Service Designer:

- *Maximal automation* of the Web service converting process by means of fully utilization of all available information about the service;
- *Supporting conceptual level of operation* by focusing the service designer on general concepts of semantic service methodology used and by abstracting away service implementation details;
- *System driven interaction with the designer* preventing from possible errors and allowing to be concentrated on the relevant part of knowledge to be acquired;
- *Building syntactically consistent and complete models* by applying context-sensitive syntactical and completeness checks at each step of the semantic service design;
- *Maximal facilitating the process of designing semantic service description* by utilizing own experience of the designer in the process of semantic Web service design.

BASIC DESIGN SOLUTIONS

The INFRAWEBS SWS Designer is aiming to convert an existing Web service into semantic Web service described by *Web Service Modelling Ontology* (WSMO). Since WSMO is in a stage of intensive development, two basic design solutions have been accepted. The first declares that the *constructing a working functional prototype of the Designer independently of the current state of WSMO implementation*: since the orchestration in WSMO is not defined yet, the functional and execution semantics of a semantic service in the first version of the Designer prototype will be described by BPEL4WS-like style and will be able to be exported to OWL-S Process Model and Grounding. The second design solution concerns a method chosen for implementing the requirement of ability of the system to improve its behaviour by utilizing its own experience. Such re-usability can be achieved by application of the case-based reasoning (CBR) approach (Aamodt & Plaza 94).

CONCEPTUAL ARCHITECTURE OF INFRAWEBS SWS DESIGNER

A conceptual architecture of INFRAWEBS SWS Designer able to satisfy the mentioned above requirements is shown on Fig. 1. Such architecture does not relate to any specific methodology for describing SWS – it describes only the flow of data processing and transformation that should be done in order to convert one (or several) Web (non-semantic) service(s) into a semantic Web service.

The *Execution Semantics Editor* translates WSDL description of an existing Web service into an initial description of a SWS that contains a complete Service Grounding, a service specific ontology corresponding to XSD types (if needed), and an initial description of service functionality as a set of non-linked atomic sub-services.

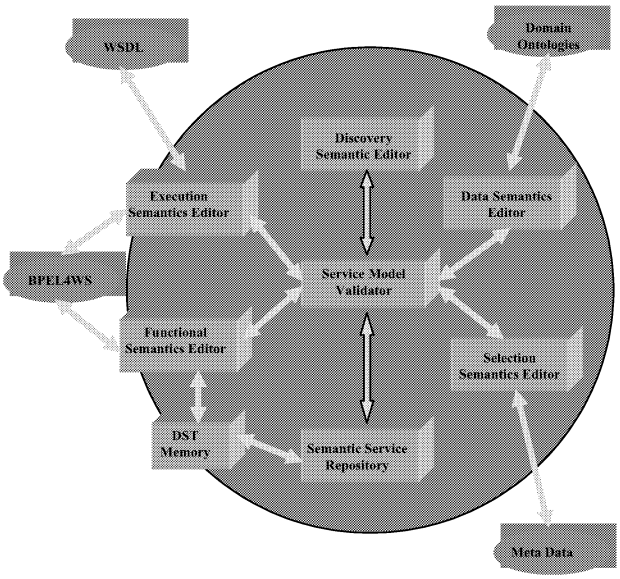


Figure 1: Conceptual Architecture of the Designer

The generated service specific ontology of inputs and outputs of atomic sub-services goes to the *Data Semantics Editor*, which accomplishes semi-automatic mapping of

such artificial concepts to the concepts from user-supplied existing ontologies.

The *Discovery Semantics Editor* is responsible for creating a consistent declarative description of service capability in terms of its inputs (pre-conditions), outputs (post-conditions), assumptions and effects. Such a description is created via a system-driven ontology-based dialog with the service designer.

The *Selection Semantics Editor* prepares a part of service description concerning QoS and other non-functional properties associated with a service. A part of this information can be acquired automatically based on the information contained in UDDI where the Web service is registered. If such information is not available, it is acquired from the user via system-driven dialog.

The *Functional Semantics Editor* aims at constructing the complete functional service description explaining how the service works. The graphical model of service functionality can be generated automatically from BPEL4WS description of the service if it is available; otherwise it is constructed semi-automatically based on Service business logic acquired from the user via system-driven dialog. The process of constructing the functional service model may be simplified by using the Design Service Templates (DSTs) retrieved by the system from DST Memory based on CBR mechanisms. Such a template represents functional description of an abstract (unrelated to some concrete WSDL description) service (or part of it), which may be re-used as a building block for creating other services. The final description of the semantic service may be stored in a Service Repository associated with the Designer.

The conceptual architecture does not prescribe any rigid sequence of steps in the process of semantic Web service design. For example, the designer may start with describing a desired capability of the service under construction or may begin with creating of its abstract functional description rather than creating the service grounding. In order to ensure the syntactical correctness of the created service model at each stage of its construction, the conceptual architecture of INFRAWEBS SWS Designer includes the Service Model Validator. This module monitors the current state of the service development and is responsible for applying various context-dependent checks for service model correctness.

FUNCTIONAL ARCHITECTURE OF THE INFRAWEBS SWS DESIGNER

The Functional Architecture of the Designer is a set of functions needed for designing a SWS using a selected framework for service description. According to the basic design principles, a SWS created by INFRAWEBS SWS Designer is described (partially) by WSMO (the service capability part) and partially by OWL-S framework (the service interface part), which can be further converted into WSMO-based representation if such standards will be developed. According to WSMO (Roman et al. 04), a semantic Web service is described by:

- *Nonfunctional core properties* (according to Dublin Core Standard) and some additional properties related to the quality aspect of a web service (QoS) such as performance, reliability, security and so on.
- *Used mediators,* describing which ontologies should be imported (or processed). Such ontologies contain concepts, relations and axioms used in definition of the concrete service.
- *Capabilities* defining the semantic service by means of its functionality. The capabilities are specified by their non-functional properties, pre-conditions, post-conditions, assumptions, effects (PPAE), and used mediators. PPAE are described by their non-functional properties and *axioms*, which are represented by some logical expressions written on a language used for implementing WSMO (WSML 04)).
- *Interface* describing how the functionality of the service can be achieved. Since up to know there is no clear agreement in WSMO consortium about specifications of a semantic Web service interface, in the INFRAW|EBS Designer it will be implemented by using OWL-S Process Model and Service Grounding.

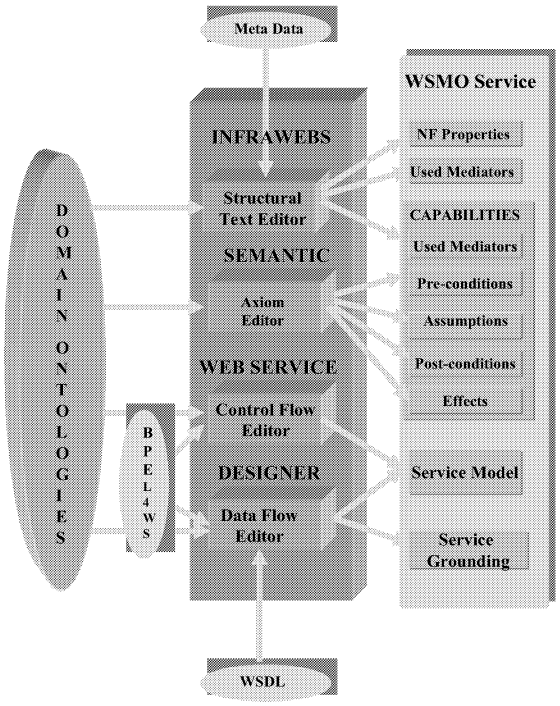


Figure 2: Functional Architecture of the Designer

The analysis of WSMO Web service definition shows that the creation of such a service may be realized by the following main functional components (Fig. 2):

- *Structural Text Editor*, which is responsible for creating definitions of QoS and other non-functional properties of the service. This module can be seen as a WSMO-based implementation of the Selection Semantics Editor from the Conceptual Architecture of the Designer. It is also planned to implement an extension of such editor allowing utilizing the available service metadata stored in UDDI.

- The PPAE part of the service description is created by *Axiom Editor* - an ontology-driven context sensitive editor of logical expressions, which creates a correct logical description of the service capability consistent with all entered data about the service. The Axiom Editor may be seen as a WSMO-based implementation of some parts of the Discovery Semantics Editor and the Service Model Validator from the Conceptual Architecture of the Designer.
- The *Control Flow Editor* will realize the Functional Semantic Editor from the Conceptual Architecture. The module uses its own representation of internal service functionality, which may be exported to different frameworks for describing SWS. It will be augmented by means for creating, storing and retrieving Design Service Templates (re-usable parts of service definition) as well as means for validating the designing Service Model against available description of other parts of the service.
- The Execution and Data Semantic Editor from the Conceptual Architecture will be realized as the *Data Flow Editor.* The module uses its own internal representation of service data flow, which may be exported to different frameworks for describing SWS (in that moment it will be OWL-S Grounding). Such an editor will be extended by a validation part responsible for accomplishing all necessary consistency checks.
- It is also planned to implement a new module creating both Process Model and Process Grounding from available BPEL4WS description of a Web service.

The Semantic Service Registry will be implemented as a Distribute Service Repository - a separate part of the Semantic Web Unit (see (Atanasova & Agre 04)).

CONCLUSION

A semantic Web service may be created by two different ways – either by converting it from an existing *non-semantic* Web service or by composing of several existing *semantic* Web services. The architecture of the INFRAWEBS SWS Designer – the tool for semiautomatic converting of non-semantic Web services to semantic ones - has been described. After analysing the existing approaches to the problem, some general requirements and basic design principles of such a tool have been formulated. These specifications have been further used for developing a general conceptual architecture of the Designer, which does not depend on a specific modeling framework used for describing semantic Web services. Then we have described the functional architecture of the Designer and discussed the mapping between both architectures.

ACKNOWLEDGEMENT

This work is carried out under EU Project INFRAWEBS - IST FP62003/IST/2.3.2.3 Research Project No. 511723.

REFERENCES

A. Aamodt and E. Plaza (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communication 7(1)*, 29-59.

T. Atanasova and G. Agre (2004). Semantic Web Unit for Design and Composition of Semantic Web Services. *IIT Working Papers, ISSN 1310 - 652X, IIT/WP-197.*

Arroyo, S., Lara, R., Gomez, J., Bereka, D., Ding, Y., Fensel, D. (2004) Semantic Aspects of Web Services. In: *Practical Handbook of Internet Computing.* Munindar P. (Editor).

Chapman Hall and CRC Press, Baton Rouge. BPEL4WS Consortium. *Business Process Execution Language for Web Services.* (2004) Available at: http://www.ibm.com/developerworks/library/ws-bpel.

M. Damashek (1995). Gauging Similarity with N-Grams: Language-Independent Categorization of Text. Science, Vol. 267, 843-848.

D. Fensel, C. Bussler, Y. Ding, B. Omelayenko (2002). The Web Service Modelling framework WSMF. *Electronic Commerce Research and Application*, 1(2).

D. Fensel and C. Bussler (2002). The Web Service Modelling Framework WSMF. *Electronic Commerce Research and Applications*, **1(2).**

METEOR-S: Semantic Web Services and Processes; http://lsdis.cs.uga.edu/projects/METEOR-S/ (2004).

H.-Joachim Nern, G. Agre, T. Atanasova, J. Saarela (2004). System Framework for Generating Open Development Platforms for Web-Service Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems - INFRAWEBS II. *WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS*, ISSN 1790-0832, Issue 1, Volume 1, 286-291.

ODE SWS - A Toolset for Design and Composition of Semantic Web Services, available at: http://kw.dia.fi.upm.es/odesws/, accessed 2004.

The OWL Services Coalition: *OWL-S: Semantic Markup for Web Services, version 1.0*; (2004); available at http://www.daml.org/services/owl-s/1.0/owl-s.pdf.

M. Paolucci, N. Srinivasan, K. Sycara, T. Nishimura (2003). Towards Semantic Choreography of Web Services: from WSDL to DAMLS. In: *Proceedings of the First International Conference on Web Services (ICWS'03)*, Las Vegas, Nevada, USA, 22-26.

A. Patil, S. Oundhakar, A. Sheth, and K. Verma (2004). METEOR-S Web Service Annotation Framework. In: *Proceedings of the 13th conference on World Wide Web.*

D. Roman, U. Keller, H. Lausen (2004): *Web Service Modeling Ontology (WSMO), version 0.1*; http://nextwebgeneration.com/projects/wsmo/2004/d4/d4.1/v01/index.html.

J. Scicluna, C. Abela, M. Montebello (2004). Visual Modelling of OWL-S Services, In: *Proceedings of the IADIS International Conference WWW/Internet*, Madrid, Spain.

E. Sirin (2004). *OWL-S API: Mindswap - Maryland Information and Network Dynamics Lab Semantic Web Agents Project*, http://www.mindswap.org/downloads/.

SWWS Studio (2004), http://stronghold.sirma.bg/swws/.

UDDI Consortium. *UDDI Specification* (2000). Available at: http://www.uddi.org/specification.html.

Unified Modelling Language. UML 2.0 (2003), available at: http://www.uml.org/

L. Varga, A. Hajnan, and Z. Werner (2004). An Agent Based Approach for Migrating Web Services to Semantic Web Services, *LNAI 3192,* 371-380.

R. Chinnici, M. Gudgin,J. Moreaum, S, Weerawarana (2003). Web Services Description Language (WSDL), Version 1.2. *W3C Working Draft*, available at: http://www.w3.org/TR/wsdl20/.

Web Service Modelling Language (2004), available at: http://wsmo.org/wsml/index.html,

CASE-BASED SEMANTIC WEB SERVICES DESIGNER AND COMPOSER

Gennady Agre
Institute of Information Technologies
Acad. G. Bonchev 29-A, 1113 Sofia
Bulgaria
agre@iinf.bas.bg

Tatiana Atanasova
Institute of Information Technologies
Acad. G. Bonchev 2, 1113 Sofia
Bulgaria
atanasova@iinf.bas.bg

Joachim Nern
big7.net GmbH & Aspasia
Knowledge Systems
Germany
E-mail: nern@aspasia-systems.de

KEYWORDS

Semantic Web Services, Web Service Modelling Ontology (WSMO), BPEL.

ABSTRACT

Realization of the full potential of the Web services requires further technological advances in the areas of service interoperation, service discovery, service composition and orchestration. A possible solution to these problems can be provided by application of Semantic Web technologies. A *semantic* Web service may be created either by converting it from an existing *non-semantic* Web service or by composing of several existing *semantic* Web services.

INTRODUCTION

A semantic Web service may be created by two different ways – either by converting it from an existing *non-semantic* Web service or by composing of several existing *semantic* Web services. Case-based designer of semantic Web services is a tool for semiautomatic converting non-semantic Web services to semantic ones within Semantic Web Services Unit (SWU) using Case-Based Reasoning (CBR) approach (Agre and Atanasova 2004). SWU is an important part of an INFRAWEBS – a collaboration platform and interoperable middleware, which is designed to be a semantic and ontology-based module handling and maintaining SWS (Nern et al. 04). The idea of case-based approach is in using of own experience. The reasoning approach considers aspects of "learning from the past" through storing experience patterns and structures.

Case-based reasoning is an artificial intelligence methodology that uses specific encapsulated prior experiences as a basis for reasoning about similar new situations (Wilson 2001). CBR systems rely on various "knowledge containers", such as the case-base of prior experiences and similarity criteria for comparing situations and retrieving the most relevant cases.

CASE-BASED SEMANTIC WEB SERVICES DESIGNER

Two of the most important principles based in developing the architecture of Case-Based Designer (SWS-D) are (see Fig. 1):

- *Maximal automation* of the Web service converting process by means of fully utilization of all available information about the service, and
- *Maximal facilitating* the process of designing semantic service description by utilizing own experience of the designer in the process of semantic Web service design.

The first principle is implemented via designing of such SWS-D sub-modules, that are able semi-automatically to process available information on functional behavior of the semantic service under construction (represented as a BPEL4WS file (BPEL4Ws 2004)), as well as available meta-data (represented as a file in UDDI) which describes non-functional properties of the service. The Designer has the proper modules that allow acquiring mentioned above information from the user via context-sensitive ontology-driven dialog when such information is not available in an explicit form.

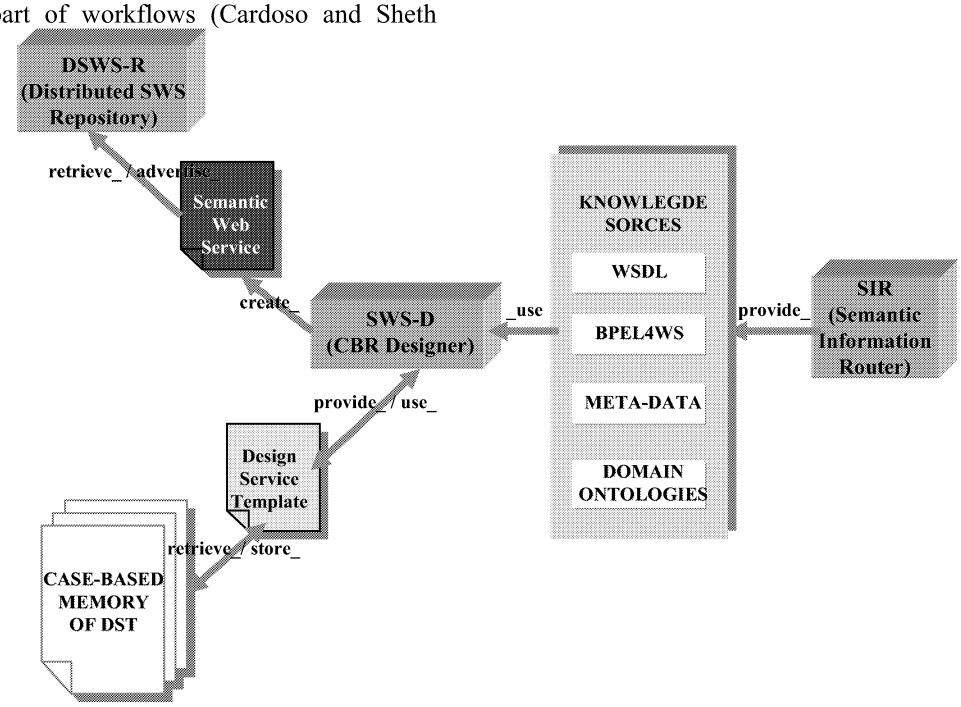
The second principle is implemented via Design Service Templates (DST), which can be used by the user during designing a semantic Web service. A DST represents a functional model (or a part of it) of a semantic Web service, which has been designed by the user in the past. The template is automatically indexed by ontology-oriented semantic information extracted from the semantic description of the SWS, which has been created by semi-automatic mapping of WSDL description of the service to the corresponding domain ontology (semantic annotation).

Creating, storing, and retrieving of similar DST are organized using CBR approach. A retrieved template can

be further used or adapted by the user for designing the desired functional model of a new semantic Web service and/or to be stored in case-based memory for later re-use.

CASE-BASED SEMANTIC WEB SERVICES COMPOSER

It is natural to expect that Semantic Web services will be integrated as a part of workflows (Cardoso and Sheth 2002). A workflow is an abstraction of a business process. It comprises a number of logic steps, dependencies among tasks, routing rules, and participants. In a workflow, a task is associated with a simple (atomic) or complex (composed) Semantic Web services.



Figures 1: Case-Based SWS Designer

The main objective of Case-Based Semantic Web Service Composer (SWS-C) as a *provider's tool* is semi-automatic design-time construction of a WSMO-based SWS by composing descriptions of existing WSMO-based SWSs and using domain knowledge (Fig. 2). The SWS-C uses the previous service compositions that form the general tasks. Such compositions are represented by Service Composition Templates (SCT), which semantically describe control and data flows between several sub-services representing as *proxies.* On the base of this information, the composer can present similar composition or ranks a list of compositions according to additional information.

Retrieval, Indexing and storing of SCT are organized according to CBR approach. A retrieved template may be further adapted by the user. A ready-to-use service template is transformed to a static Semantic Web service by discovery of appropriate existing Semantic Web services, replacing the proxies participated in the template. Such sub-services are discovered by Service and User Agents (SUA) module either from provider's SWS repository or from external repositories of WSMO-based SWS.

There are different solutions for location and matching of web services (Schirmer 1998), (Limthanmaphon and Zhang 2003). But they have to be developed further to deal with matching and discovery of Semantic Web Services.

The result of matching may give a number of matches. The selection of one of them could be based on some domain-specific selection rules and constraints provided by Distributed Decision Support module (DDS).

The current template has to be changed if no real sub-services could be found to match exactly some of template proxies. Making of such changes could be facilitated by some domain-specific adaptation rules provided by DDS.

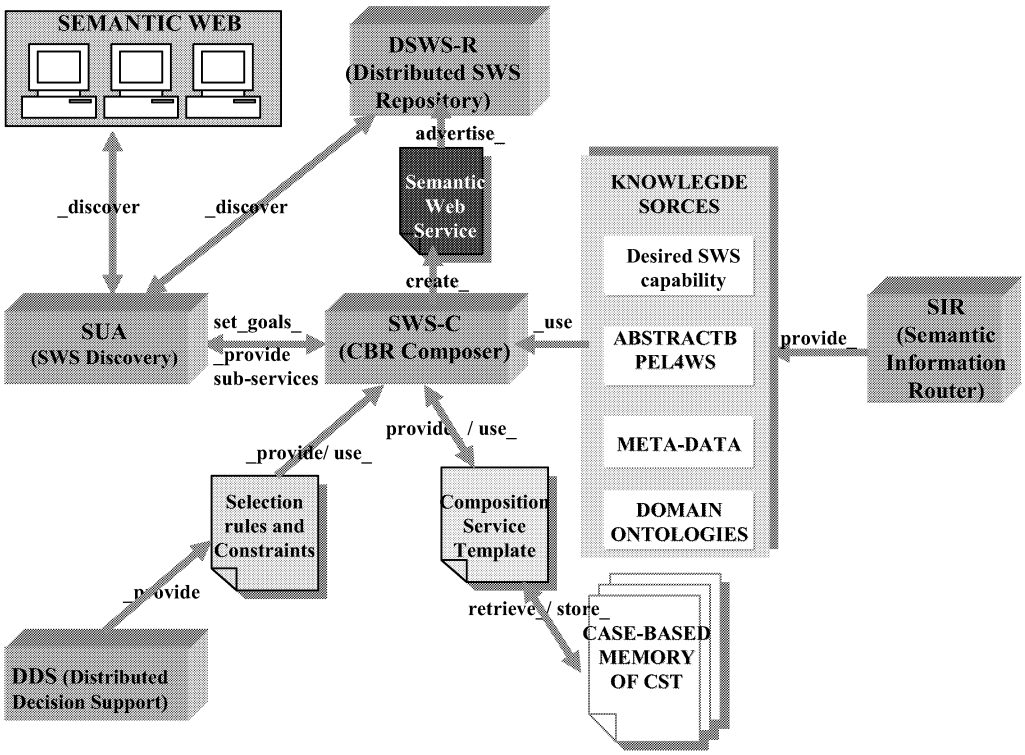
The user has a possibility to store the new composed semantic Web service as a dynamic service implementing a task for discovering the proper existing sub- service for each service proxy in run-time (or do it only for some proxies). However, even in this case it is necessary to perform the described above procedure for discovery of sub-services in order to guarantee, that the description of such dynamic service is correct, i.e. that it could be found at least one existing sub-service per proxy that matches exactly all specified constraints.

The main idea of SWS Case-Based Composer operating as a provider's tool for semi-automatic design-time composing of semantic web services may be resumed as follows:

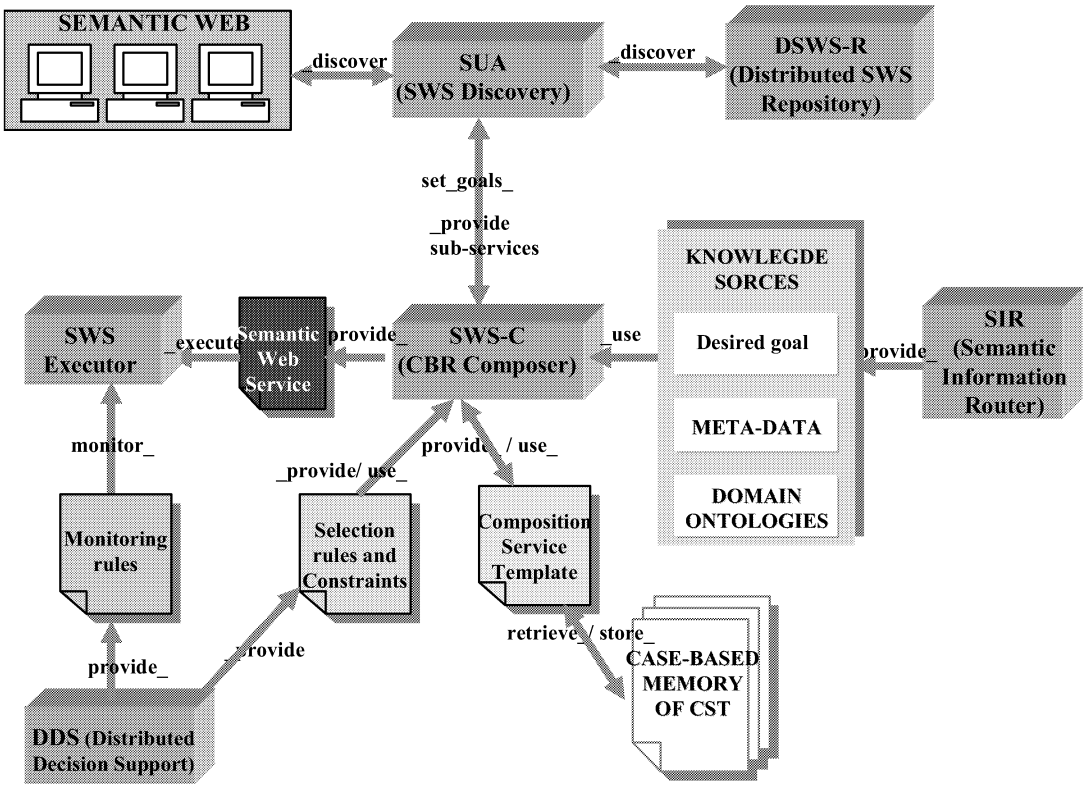
- Similarity-based retrieval of an appropriate semantic service template based on the description of capability of the desired service and some (partial) description of its functionality (e.g. as abstract BPEL4WS process)

- Semi-automatic adaptation of service functional model based on the results of discovery of sub-services matching the template proxies
- Advertising the created service and its generalization and storing as new template for later re-use.

The main objective of Case-Based SWS Composer as *a client's tool* is discovery and *execution* of a Semantic Web service satisfying a of WSMO-based dynamic SWS by composing client goal. Since it is very probable that such a service does not exist, this objective can be realized by semi-automatic run-time construction existing WSMO-based SWSs and using domain knowledge.



Figures 2: Case-Based SWS Composer as Provider's Tool



Figures 3: Case-Based SWS Composer as Client's Tool

The main differences in functionality of SWS-C as a client's tools are the following:

- The lack of knowledge of capability and functionality of the requested service that should satisfy the specified goal.
- The composed service matching the client's goal should be executed
- The composed service does not need to be advertised after its execution

Composing of a dynamic Web service satisfying the desired user goal can be realized as a sequence of the following main steps (Fig. 3):

1. *Formulating a WSMO goal:* this task may be accomplished via the Axiom Editor described in (Agre and Atanasova 2004).
2. *Creating of composed design-time service:* it can be done by retrieving and adaptation of service composition patterns and discovering and selecting SWS services that match the pattern;
3. *Execution of the created composed service:* the process of service execution is monitored according to the Monitoring rules provided by DDS unite.
4. *Storing the goal and the created service composition pattern* for later re-use.

CONCLUSION

SWU is associated with particular domain where typical tasks within this domain exist that leads to existing of typical solutions. So collecting web services templates according to this typical tasks and generalization of experience (learning from the past) could help developing effective methods of their applications.

Case-based designer of semantic web services aims to plan and discover Semantic Web Services within SWU using Case-Based Reasoning (CBR) approach.

ACKNOWLEDGEMENT

This work is carried out under EU Project INFRAWEBS - IST FP62003/IST/2.3.2.3 Research Project No. 511723.

REFERENCES

G. Agre and T. Atanasova. Adding Semantics to Web Services. *IIT Working Papers IIT/WP-196*, ISSN 1310-652X, December, 2004

T. Atanasova and G. Agre. Semantic Web Unit for Design and Composition of Semantic Web Services. *IIT Working Papers, ISSN 1310 - 652X, IIT/WP-197,* December 2004.

BPEL4WS Consortium. *Business Process Execution Language for Web Services,* 2004.. Available at: http://www.ibm.com/developerworks/library/ws-bpel,

J. Cardoso and Amit Sheth, Semantic e-Workflow Composition, *Technical Report# 02-004*, LSDIS Lab, Computer Science Department, University of Georgia, Athens GA, July 2002.

H.-Joachim Nern, G. Agre, T. Atanasova, J. Saarela. System Framework for Generating Open Development Platforms for Web-Service Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent Systems - INFRAWEBS II. *WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS*, ISSN 1790-0832, Issue 1, Volume 1, July 2004, 286-291.

B. Limthanmaphon and Yanchun Zhang, Web Service Composition with Case-Based Reasoning, *Proceedings of the Fourteenth Australasian database conference on Database technologies* 2003 - Volume 17.

A. Schirmer, *Case-Based Reasoning and Improved Adaptive Search for Project Scheduling*, April 1998, www.bwl.uni-kiel.de/bwlinstitute/prod/mab/schirmer.

D. C. Wilson, Case-Base Maintenance: The Husbandry of Experience, *Ph.D. Dissertation*, Indiana University, 2001.

SEMANTIC WEB SERVICE EXECUTION FOR WSMO BASED CHOREOGRAPHIES

James Scicluna
Axel Polleres
Digital Enterprise Research Institute (DERI Innsbruck)
Institute of Computer Science, University of Innsbruck,
Technikerstrasse 21a, 6020 Innsbruck, Austria
E-mail: {james.scicluna, axel.polleres}@deri.org

KEYWORDS
Semantic Web, Web Services, WSMO, Web Service Execution, Choreography, Orchestration, Workflow.

ABSTRACT

The Semantic Web is slowly gathering more importance as both academic and industrial organizations are realizing the potential benefit that might be obtained from it. This is especially true in the areas of tourism in which Semantic Web Services can provide a drastically new way on how to find and book related services such as hotel, flights and taxi transfers. However, many aspects of Semantic Web Services are still under development. This short paper presents issues related to choreography and orchestration representation in the Web Service Modelling Ontology (WSMO) and also how such ideas can be applied to an e-tourism use case.

INTRODUCTION

The Web Service Modelling Language (WSMO) is based on the Web Service Modelling Framework (Fensel and Bussler 2002) which specifies the key elements for describing Semantic Web Services. Such elements include Ontologies, Goals, Web Services and Mediators. Ontologies are the key to describe all of these elements through concepts, relations, functions, instances and axioms. Goals specify what the user wants from a particular web service. Web Service descriptions specify what a service can provide in terms of a capability and any number of interfaces. Mediators link heterogenous elements between each other in order to enable heterogenous components to interoperate.

Details of these elements are beyond the scope of this document and we refer the reader to (Feier 2005). Rather, here we want to focus on the Interface of a web service description which enables to model a choreography interface and orchestration for a web service. We will then provide an example of to define such an interface based on an etourism use case (Lopez 2004).

Our use cas is called *Stream Flows! System* (SFS). It is a frequent flyer program designed to be accessible from the internet by its customers. Using this system, customers can buy packages or contract new ones in exchange of points rather than using a credit card. Packages may include flights, hotel booking, car rentals and others. The system is responsible for the continuous maintenance of its packages, that is, once a service is no longer available, the package has to be updated or removed from the system.

A Web Service description in WSMO is defined in terms of non functional properties, a single capability and any number of interfaces. *Non Functional Properties* define information regarding the usage of the service without affecting its functional aspect. In addition to the core non functional properties for the WSMO elements, a web service defines further service specific issues such as quality of service, financial range, robustness and others. A *capability* defines what the service can provide in terms of preconditions, assumptions, postconditions and effects. A *precondition* defines what the web service expects for enabling it to provide its service. An *assumption* describes the expectation of the service on the state of the world when starting an execution of the service. A *postcondition* describes the state of the information space that must be reached by executing the service and an *effect* describes the state of the world that must be reached after executing the service. All these features are defined using logical expressions based on the Web Service Modelling Langue (WSML) (De Bruijn 2005) .

The *interface* of a service defines two pieces of information, namely, a choreography and an orchestration. WSMO allows to define multiple interfaces for a service and hence there could be more than one way in which a user (or an agent) may interact with it. The *choreography* exposes how a client (whether a user or an agent) should interact with the service. The *orchestration* describes how the service uses other web services in order to achieve the required functionality.

CHOREOGRAPHY

WSMO defines a choreography as how services interact with a client which may either be a user or a service. This adheres to the W3C definition. WSMO is concerned with the formal modeling of a choreography which would also enable to perform reasoning over the description (such as to identify whether two web services can talk to each other).

Abstract State Machines as a Formal Model

Abstract State Machines (formerly knows as Evolving Algebras) have been used to describe and validate a wide variety of computing systems. In WSMO, they provide the basis for describing both the choreography and orchestration specifications. In a nutshell, an abstract state machine is defined in terms of a state signature, a state and a number of guarded transitions. A state signature defines the invariant elements of the state description. A state is described by a set of instance elements and guarded transitions are rules that express the changes of states. For further details about how

these aspects are defined in a choreography, we refer the reader to (Roman et. al., 2005). Rather, we will provide some examples on how these aspects can be applied In our use case.

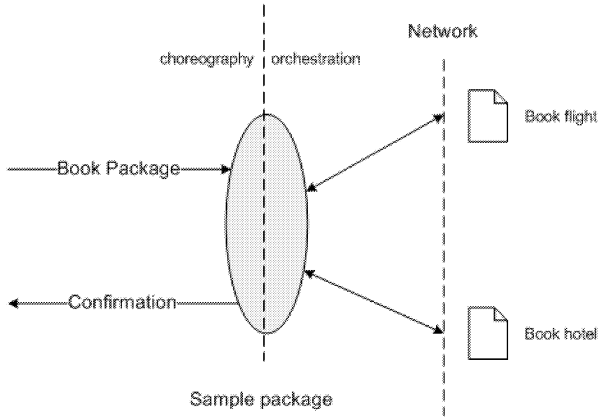


Figure 1: A simple package illustrating the WSMO Choreography and Orchestration Perspectives

Figure 1 depicts a very simple package which defines in its choreography a booking of a flight and then that of a hotel. We will now provide a simple way on how to specify this in terms of an abstract state machine in the choreography of some interface defined in the service. For simplicity reasons and to keep this document concise, we will just describe the choreography definition in the web service. Hence, we will assume that the needed ontologies (for flight, hotel etc) already exist in some ontology. The classes needed are as follows:

- Package Booking (receive a request to book the package)
- Package Confirmation (send a confirmation that package has been booked)
- Flight
- Hotel

In order to identify which concept is an input or an output, we require to extend the concepts needed in order to support the attribute mode whose value determines whether a concept is an input or an output of a choreography. This extended ontology imports all the ontologies which define the classes to be extended in order to be used by the Web Service. Below is an abstract of the extended concepts to be used by the web service.

```
bookPackage subClassOf sfs#packageBooking
  nonFunctionalProperties
    mode hasValue in
    grounding hasValue sfsWsdl#bookPackage
  endNonFunctionalProperties

packageConfirmed subClassOf sfs#confirmedPackage
  nonFunctionalProperties
    mode hasValue out
    grounding hasValue sfsWsdl#bookPackage
  endNonFunctionalProperties
```

A state is described by the attribute values of the instances in the choreography vocabulary (that is, the ontology defining the extended classes needed). The transition rules in the choreography express the current state in the condition and also the updates required in order to move to a next state. In both cases, this is defined in terms of attribute values of the instances used by the choreography as shown below.

```
choreography simplePackageChoreography

  guardedTransitions simplePackageTransitions

    if(bookPackageInstance memberOf
       sfs#packageBooking[
         book hasValue true
       ])
    then
      (confirmedPackageInstance memberOf
        sfs#confirmedPackage[
          confirm hasValue true
        ]).
```

Notice that the concepts bookPackage and packageConfirmed are grounded to a single WSDL Operation. The mode attribute thus identifies whether the concept is an input or an output of the operation. In our very simple example, the choreography requires only one transition since as soon as the booking request is received, what's left is to send back a confirmation. The booking request defines a boolean value which determines whether the booking should be performed or not, if this is so, the confirmed package instance will be sent to the user and its *confirm* value set to true.

ORCHESTRATION

WSMO Orchestration is based on the same Abstract State Machine model. However, the guarded transitions are extended to support links to mediators when performing updates. These mediators can either link to a goal (wgMediator) or to another Web Service (wwMediator). The orchestration thus models what other Web Services are to be used in order to achieve the requirements of the Web Service defining the orchestration.

```
if(flightInstance memberOf sfs#flight[
     booked hasValue true
   ]
then
  _"http://example.org/bookFlightMediator"

if((flightInstance memberOf sfs#flight[
     booked hasValue true
   )]
   and
   (hotelInstance memberOf sfs#hotel[
     reserved hasValue true
   ])
then
  _"http://example.org/bookHotelMediator"
```

Notice that the second rule specifies a condition on both the flight and hotel instances in order to ensure that it is not infinitely fired.

The roles of the mediators linked by the orchestration differ depending on the type of mediator used. A *wgMediator* must:

1. Mediate data between the ontologies used by the source service and the target goal.
2. Link the capability of the source web service to the one in the target goal such that it reduces the

information space needed by the orchestration at a given state

The purposes of a *wwMediator* are as follows:

1. The normal data mediation which is needed to resolve heterogeneities between the ontologies used by the source and targets
2. Link the orchestration to the target choreography of the web service such as to enable invocation of the target web service
3. Perform Process Mediation in case the orchestration and the target choreography don't match at the protocol level

LANGUAGE FOR A WORKFLOW SPECIFICATION

The language for choreography and orchestration in WSMO formally models such aspects. However, such a model is not executable in terms of concrete Web Service calls (such as via SOAP). Therefore, a more expressive workflow language is needed in order to enable actual execution of an orchestration. This language must enable to perform execution at the semantic level in order to enable reasoning about parameters (such as simple subsumption) and also reasoning for service composition purposes. An example of such a workflow language is outlined below.

```
inputs
  bookingRequest impliesType sfs#packageBooking

outputs
  confirmation impliesType sfs#confirmedPackage

initial_state
  bookingRequest = EMPTY

flow
  get(bookingRequest)

  if(bookingRequest.book == true)
    then
      confirmation.confirm hasValue true
    else
      confirmation.confirm hasValue false

  put(confirmation)
```

First, the inputs and outputs are declared. All of the elements needed are instances of ontologies. Where necessary, variables are used but this is not the case. The initial state is declared by setting the input to get to EMPTY. The <*flow*> element defines the behavioral flow of the service. To receive data, the method *get* is used. This method required that execution stops until the required data is obtained. Once the booking request is received, it is checked if the *book* attribute is set to true and the *confirm* attribute of the confirmation element is set appropriately. Once the package has been booked, the confirmation element is sent to the client using the *put* statement.

Another approach in defining "semantic" workflows is that of providing semantic information for existent technologies and specifications. An example for such an approach is the one taken in (Mandell and McIlraith) providing a description of how BPEL4WS can be extended in order to provide the semantic information needed for automatic discovery and composition.

FUTURE WORK

Our main future efforts will involve in a better specification for Choreography and Orchestration in WSMO. Furthermore, a more clear and executable workflow language will be specified and mapped to the ASM model defined in WSMO. Also, the mediators which are used to link from an orchestration to a goal (*wgMediator*) or another Web Service (*wwMediator*) need to be further defined in order to enable interoperability between their source and target components.

CONCLUSION

We hereby provided a very short description of the current work which addresses the choreography and orchestration specifications in WSMO based on Abstract State Machines and how we intend to provide a workflow language in order to enable execution of the web service. Also, a simple example was provided to show how these aspects can be applied in an eTourism use case. Our main focus was how WSMO can model such scenarios and enable also actual execution of the underlying service.

ACKNOWLEDGEMENT

The work is funded by the European Commission under the projects DIP, Knowledge Web, InfraWebs, SEKT, SWWS, ASG and Esperonto; by Science Foundation Ireland under the DERI-Lion project; by the Vienna city government under the CoOperate programme and by the FIT-IT (Forschung, Innovation, Technologie - Informationstechnologie) under the projects [EQUATION] and TSC.

REFERENCES

De Bruijn J. 2005. "The WSML Family of Representation Languages" *WSML Working Draft*, March 2005.

Feier Cristina. 2005 "WSMO Primer". *WSMO Working Draft*, March 2005..

Fensel D.; Bussler C. 2002 "The Web Service Modeling Framework". January 2002.

Gurevich Y. 1993 "Evolving Algebras: Lipari Guide, 1993".

Lopez O. 2004 "Requirement Profile 1 & Knowledge Objects". *Infrawebs Working Draft* December 2004

Mandell D. and McIlraith S. "Adapting BPEL4WS for the Semantic Web: The Bottom-Up Approach to Web Service Interoperation"

Roman D.; Feier C.; and Scicluna J. "Ontology-based Choreography and Orchestration of WSMO Services" *WSMO Working Draft.* March 2005

DISTRIBUTED DECISION SUPPORT IN MULTI-AGENT ENVIRONMENT FOR SEMANTIC WEB SERVICES

Tatiana Atanasova
Institute of Information Technologies
Acad. G. Bonchev 2, 1113 Sofia
Bulgaria
atanasova@iinf.bas.bg

Joachim Nern
Aspasia Knowledge Systems
D-40210 Duesseldorf
Germany
E-mail: nern@aspasia-systems.de

KEYWORDS
Distributed Decision Support, Semantic Web Services, Semantic Web Unit, Multi-Agent Systems.

ABSTRACT

Distributed decision support (DDS) for monitoring and execution of Semantic Web Services within Semantic Web Services Unit (SWU) is discussed. SWU is aimed to assist fully networked enterprises. Semantic Web Services Unit is a multi-agent ontologically-based interoperable environment. DDS module contains service oriented rules. These rules are derived from the application requirements and are based on the domain dependent and domain independent knowledge. They govern the monitoring and execution of services specifying under which conditions the services have to be selected for execution. The non-functional properties of services also have to be analyzed to instantiate the necessary management aspects. Semantic description is considered as a part of agent based functioning of networked enterprises. Agents perform activities in SWU that are aimed to exploit key features of Semantic Web-enabled Web Services in semi-automatic way.

INTRODUCTION

One of the components of the current INFRAWEBS project (Nern et al, 2004) is Distributed Decision Support (DDS) module for Semantic Web Service (SWS) handling and maintaining in Semantic Web Services Unit (SWU).
Within the framework of INFRAWEBS project Semantic Web Services Unit has to be designed to enable the user to be "broker, provider and client" at the same time – within one integrated system. For this purpose it is intended to implement Semantic Web Services Unit as a semantic and ontology based environment which is aimed to be a base for operating with Semantic Web Services. All six phases of Semantic Web Service usage process (Arroyo et al 2004), namely: discovery, composition, mediation, execution, monitoring, and compensation are reflected in SWU. Beside that SWU ensures design of SWS from non-semantic web services and their deployment to the net.
When Semantic Web Services are provided to the network they may be incorporated in other services by composing process. It is assumed that the Semantic Web Services should be able to play as the role of the requester as well the role of the provider. The roles may switch during the same transaction.
SWU has its own shared ontology of specialized domain knowledge and library with general situations about errors during executing associated with possible recipes for recovering.
Handling and maintaining of Semantic Web Services in SWU has to be ensuring by Distributed Decision Support infrastructure that is intended to be designed as a workflow and knowledge sharing component. Distributed Decision Support has to cover monitoring of SWS execution, verification and validation as well as Quality of Service (QoS) brokering for e-business applications.
Business processes usually are described by BPEL tools, in which local flows may be independently developed by different participants. Besides that the components are not reusable because they combine business logic and data interactions. The local flows may be failure in the result of different goals of the participants in given business process or they can not be adaptable to changes at runtime. The flows are parts of SWS orchestration and choreography. So monitoring of service execution is necessary because of these properties of services that need observation over time. Verification and validation modules have to deal with analysis and simulation of SWS behaviour.
Web services enriched with semantic description of their capabilities determine agent's abilities in SWU. A common semantic markup for Web Services is a base for effective communication between agents in the multi-agent environment.

MONITORING OF SWS EXECUTION

The autonomy of Semantic Web services does not allow for users to identify their operational metrics at design time or before their actual execution. Thus the Distribution Decision Support unit has to monitor SWS execution and provide compensation mechanism to repair failed services. It would be useful to see the progress of execution when a complex service is being executed. For this aim the Semantic Web Service is specified at knowledge level from the execution point of view by specifying execution semantics through a rule set.
Rules allow inferring at runtime thus the dynamic behaviour may be achieved. It is proposed to develop this set of decision rules on two levels: as domain independent (global) rules and domain dependent (local) set of rules.

The global set of rules for service execution is mirrored into every SWU and concerns some general planning and monitoring issues. The local rules are domain dependent and are specific for every particular set of semantic web services, connected to given ontology and business logic process management. Distributed rule base contains experience about successful or failed execution of SWS.
So the Rule-Based Reasoning (RBR) approach is proposed for implementation of DDS. In RBR, rules can be thought of as if-then-statements; the behaviour of the system is dictated by a process that iteratively searches for a rule whose if-part (antecedent) matches the input, and implements the prescriptions of the corresponding then-part (consequent). In (Schirmer 1998) it is mentioned that some aspects of CBR implementation are closely related to the Rule-Based Reasoning approach. In CBR, cases are constants, they are retrieved based on partial matching, and require partial adaptation to get a solution. In RBR, rules are patterns, they are retrieved based on exact matching, and consequently adaptation is not required to get a solution.
Such Rule-Based Patterns for SWS monitoring in DDS are constructed from the multi-level models of e-business templates for:

- Business logic
- Integration
- Composition
- Application
- Execution

Effective decision support for SWS configuration should contain tools for modelling QoS, which includes their timeliness, quality of products delivered, cost of service, and reliability. This may be implemented by validation and verification of SWS execution behaviour (Wilson 2001), (Cardoso 2004).
Interoperability between different partners in distributed network environment has to be considered too.

INTEROPERABILITY - MEDIATION

To achieve interoperability, it is necessary to address the problem of semantic integration – the identification of semantically similar objects that belong to different systems and the resolution of their schematic differences. When Semantic Web Services are put together, their interfaces (inputs and outputs) need to interoperate; therefore, structural and semantic heterogeneity needs to be resolved in e-business applications.
Structural heterogeneity exists because Semantic Web Services may use different data structures to define the parameters of their interfaces. Semantic heterogeneity considers the intended meaning of the terms employed in associated with particular ontology input and output parameters. The data that is interchanged among Semantic Web Services has to be understood. Semantic conflicts occur when a Web service output connected to another service input does not use the same interpretation of the information being transferred. The general approach to semantic integration is to map the local terms onto a shared ontology. Even though a shared ontology ensures total integration, constructing such ontology is costly, if not impractical; autonomous systems are required to commit to a shared ontology, and compromises are difficult to maintain when new concepts are added.

MULTI-AGENT APPROACH FOR SEMANTIC WEB UNIT IMPLEMENTATION

Semantic web services provide computational resources but agents are those that coordinate the framework. Evolution from web services to semantic web services insists on evolution of coordination mechanism that allows software agents to be loosely coupled to ensure the behaviours they possess.
If agents use semantically described Web Services, then the semantic service descriptions become the basis for determining the agent's abilities. A common semantic markup for Web Services will facilitate effective communication between agents. Social agents that have access to an ontology-based semantic description of their behaviours should be better able to proactively coordinate themselves.
Agents can be viewed as independent applications that provide services to one another through loosely coupled, asynchronous message exchange.
Agents execute tasks on behalf of a decision process. Numerous agents in the multi-system can individually and concurrently search for relevant SWS at appropriate levels. A hierarchy of agents may be formed from different groups of agents: info broker, DS agent, mediation agent, etc. For realising a higher goal, the activities of cooperative agents should be coordinated. Several coordination mechanisms exist such as competitive coordination, cooperative coordination, temporal sequencing or no coordination.
The generic agent architecture can be instantiated by adding specific types of knowledge to support SWS functionalities and behaviour required. Depending on the choice of these requirements, an agent is created for a specific context by including the appropriate types of knowledge.
Agents on the web made Semantic Web Services available and usable for any purpose by enabling expressive service capability advertisements. So it is intended to realize a SWU as multi-agent system that is enriched with ontology for Semantic Web Services description from particular domain.
Agents perform activities in SWU that are aimed to exploit key features of Semantically-enabled Web Services in semi-automatic way:

- selection of a required service from a set of services.
- invocation of composite, generically-built services.
- monitoring of execution.

Discovering, invocation and selection of SWS on the base of their functionality and internal structure according semantic description techniques are part of agent-based computing (Hendler 2001) in SWU.
SWS services are located using discovery process based on the services ontologies which describe services in terms of their capabilities (through a description of a service which

an agent uses to advertise, compare, or look up that service), how the service works (through the workflow of how an agent would interact with a service), and how to invoke it (through defining the format of the messages that an agent and the service exchange and includes practical information, such as the address and port of the machine where the service is hosted).

A first step to design this multi-agent system (MAS) is to provide an ability to communicate with sufficiently precise shared meaning via Semantic Web (Varga 2004). Semantic markup of Web services enables a wide variety of agent technologies.

Each agent is autonomous, goal-driven, cooperative, coordinated, rational and able to communicate with other agents to fulfil the users' needs. Different agents use very different internal representations for reasoning about and planning sequences of actions (El-Ramly 2002). So agent can be viewed as customizations of reusable, high-level generic procedures. These generic procedures are customized with respect to constraints, using deductive machinery for discovery, execution, composition and interoperation. They can be stored in sharable generic procedures ontologies for accessing to them from multiple users in networked enterprises.

CONCLUSION

Semantic Web Unit is aimed to assist fully networked enterprises. SWS have to be discovered, invocated, negotiated, selected, composed, executed, and validated with integrated knowledge and inference in SWU. Agents have to execute tasks on behalf of a decision process. Supporting Semantic Web Processes require addressing heterogeneity/integration, scalability, dynamic change and performance challenges. The Distributed Decision Support unit is aimed to assist the SWS handling and maintaining in SWU.

The Distributed Decision Support infrastructure ensures three dimensions of distribution: computation, control, and data. The data is managed via a shared RDF-based repository, while the control of the SWS execution is driven by asynchronous message exchange between the agents. The message exchange pattern for the control messages is derived from a rule set.

There are unsolved problems for supporting Semantic Web Services Processing, for example, how to facilitate the interoperability of heterogeneous Web services and how to provide compensation mechanisms in run-time based on functional and operational requirements of SWS.

ACKNOWLEDGEMENT

This work is carried out under EU Project INFRAWEBS - IST FP62003/IST/2.3.2.3 Research Project No. 511723.

REFERENCES

Arroyo, S., Lara, R., Gomez, J. M., Bereka, D., Ding, Y., Fensel, D. 2004. "Semantic Aspects of Web Services". In: *Practical Handbook of Internet Computing*. Munindar P. (Ed.) Chapman Hall and CRC Press, Baton Rouge.

Buhler, P. and J. M. Vidal, "Towards adaptive workflow enactment using multiagent systems," *Information Technology and Management Journal: Special Issue on Universal Enterprise Integration*, Vol 6, No 1, 2005, pp.61 - 87.

Cardoso J., Amit Sheth, John Miller, Jonathan Arnold, and Krys Kochut, 2004, Quality of Service for Workflows and Web Service Processes, *Journal of Web Semantics* (to appear), Elsevier.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., and Lama, M. 2003. "An Environment for Development of Semantic Web Services". *Proc. IJCAI-2003 Workshop on Ontologies and Distributed Systems*, 13–20. Acapulco, México

Fensel, D., C. Bussler, Y. Ding, B. Omelayenko. 2002. "The Web Service Modelling framework WSMF". *Electronic Commerce Research and Application*, 1(2).

Hendler, J., 2001, "Agents and the Semantic Web". *IEEE Intelligent Systems,* 16 2:30–37

Nern, H.-J., G. Agre, T. Atanasova, J. Saarela. 2004, "System Framework for Generating Open Development Platforms for Web-Service Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems - INFRAWEBS II. *WSEAS TRANS. on INFORMATION SCIENCE and APPLICATIONS*, 1, Vol. 1, 286-291.

Schirmer A., 1998, "Case-based reasoning and improved adaptive search for project scheduling", Manuskripte aus den Instituten fur Betriebswirtschaftslehre 472, Universitat Kiel, Germany.

Varga L., A. Hajnan, and Z. Werner, 2004, "An Agent Based Approach for Migrating Web Services to Semantic Web Services", *LNAI,* 3192, 371-380.

Verma, K., A. Sheth, F. Curbera. 2000. "Service Oriented Architectures and Semantic Web Processes" Technical Report 03-008 LSDIS Lab, Dept of Computer Science, University of Georgia.

Wilson D. C., 2001, "*Case-Base Maintenance: The Husbandry of Experience*", Ph.D. Dissertation, Indiana University.

THE SUA-ARCHITECTURE WITHIN THE SEMANTIC WEB SERVICE DISCOVERY AND SELECTION CYCLE

Csaba Fülöp, László Kovács, András Micsik

MTA SZTAKI
Computer and Automation Research Institute
of the Hungarian Academy of Sciences
Department of Distributed Systems
H-1111 Budapest XI. Lágymányosi u. 11. Hungary
{csaba.fulop,laszlo.kovacs, micsik}@sztaki.hu

KEYWORDS
Semantic Web Services, middleware

ABSTRACT

As participants of the INFRAWEBS project aiming at establishing an open and extensible development platform for web service applications, we are investigating the tasks and methods in the process of web service discovery, selection and invocation, from the viewpoint of user applications. Current architecture of INFRAWEBS environment is presented, and decisions taken on various web service deployment approaches are detailed.

INTRODUCTION

The primary objective of INFRAWEBS is to develop an ICT framework, which enables software and service providers to generate and establish open and extensible development platforms for Web Service applications. Generated in this way, the open platform consists of coupled and linked INFRAWEBS units, whereby each unit provides tools and adaptable system components to analyse, design and maintain Web Services realized as Semantic Web Services within the whole life cycle.
In the first step the INFRAWEBS units are used to establish an open development and collaboration platform for SMEs and strong industrial suppliers for software and services. This offers a new dimension of collaborative work and service production, service provision and service maintenance in run-time environments (Nern et al. 2003).
The software environment of INFRAWEBS provides ontologies and similarity based organizational memory to enhance the description of Semantic Web Services. The rich descriptions of Semantic Web Services are placed into a distributed SWS Repository. With the help of this repository a middleware layer called SUA (Service and User Interface Agent Environment) provides facilities for discovery, selection and invocation of Semantic Web Services. There is also an executor unit responsible for the choreography and orchestration of web services.
In this paper the generic problems of the middleware layer are discussed.

THE WEB SERVICE INVOCATION ENVIRONMENT

The SUA unit is a basic layer of the INFRAWEBS software environment. Its architecture is shown on Figure 1. The Communicator is responsible for building up the communication channels between users and various other modules based on security and privacy templates provided by the S&P module. The Communicator selects the most appropriate level of security and privacy for the given method of communication. The User circle in this diagram denotes the application acting on behalf of the user: for example a GUI interface or an intelligent agent. The SWS Composer is a user tool for the creation of new semantic web services as a composition of existing semantic web services. Both components need to discover and select existing web services with specified capabilities. The Discovery Agent supports this task, while the Service Agent supports the execution of the selected web service in cooperation with the SWS Executor module.
Both the SWS Executor and the Discovery Agent request information about web services such as capability or choreography from the SWS Repository. The discovery process is planned as a hybrid approach combining text processing and reasoning. Therefore, this process needs other external information and tools as well: the domain ontologies contain the background information about service domains and organizational memories gathered from SMEs in the INFRAWEBS framework and transformed into Semantic Web compatible format. Controlled vocabularies are different from domain ontologies both in their focus and implementation: they support the processing of natural languages, for example word stemming or translation.
In this environment the SUA component is the direct communication channel with web services. In this aspect we are investigating the process of transforming semantic web service descriptions (such as capabilities and choreography) into user interface dialogs. Normally, this would be the task of the user application, and therefore the user application should have the same or even more semantic knowledge about the web service than the SWS environment. The transfer of this knowledge and transformation into user-friendly multilingual dialogs is another area of our interest.

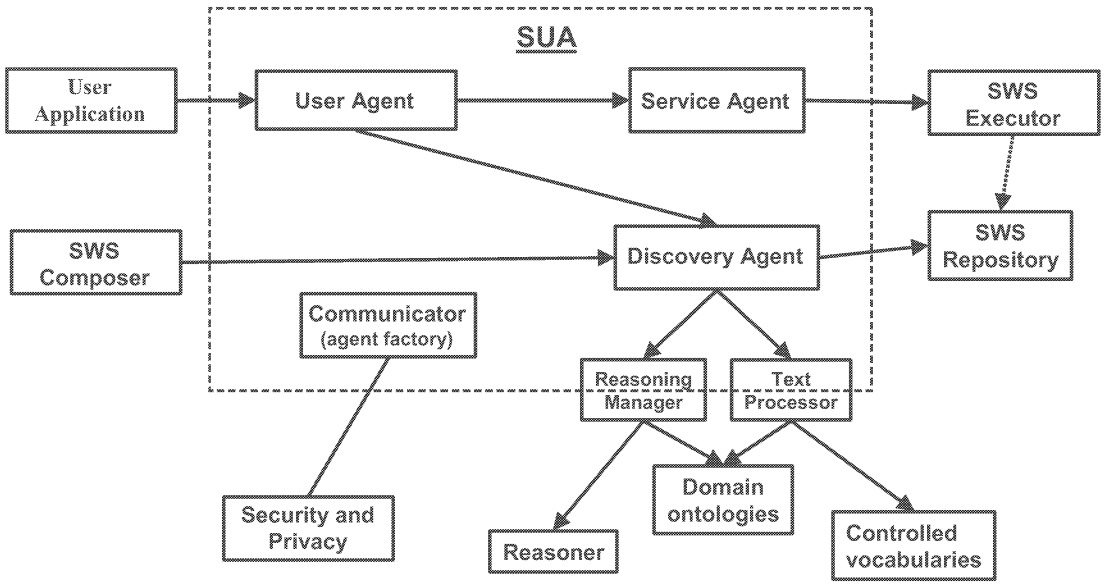


Figure 1: Internal Architecture of Middleware Layer and its Dependencies with Other Modules

GENERIC MIDDLEWARE FOR SEMANTIC WEB SERVICES

The SUA unit within INFRAWEBS acts as a middleware layer for user applications, connecting these applications to the functionality available in the form of Semantic Web Services. An important decision is whether to hide the semantic approach and accompanying logic-based framework of SWS from the user applications or not. What level to choose for communication? On the level of plain web services, simple XML data structures are exchanged, which are easy to generate and process, but lack the possibilities of the semantic approach. On the level of semantic web services, facts and rules are exchanged in the form of logical expressions. As the latter case puts extra requirements for the user application, we decided to find a middle course between the two solutions. Communication is kept on the semantic level, but its form is either hidden or aided with special functionalities and automatic translations.

The usual steps of SWS deployment include:

1. Construction of a goal: The goal describes what the user wants to achieve by SWS execution in terms of logical expressions.
2. SWS discovery: a list of appropriate web services is provided for the goal.
3. SWS selection: the user or the application has to select a service if there are more services available for the given goal.
4. SWS invocation: the selected web service is started for the given goal, optionally with facts as additional input.
5. SWS execution: the invoked web service executes its communication pattern with the invoker (choreography) and it may also communicate with other web services (orchestration) to produce the result.
6. SWS result: the result of service execution is manifested as new facts that fulfill the postcondition of the goal, for example, the reservation number and departure time of the flight are given.

Design considerations for the previously described SUA component are detailed in the following with respect to these steps.

Goal Construction

The creation of logical expressions as goals is an unusual and challenging task for ordinary applications. Users are neither expected to type in goal expressions using proper syntax. Therefore, use of goal templates was decided, which cover the more frequent service goals occurring within the applications. Missing data within the goal template can be provided to the necessary extent either by the application or by the user. Textual transcriptions of the templates are presented for the users where input boxes mark the missing parts. Goal templates are stored in the SWS repository. These templates are searchable by contained ontology terms and they can be clustered by these terms as well, thus comfortable selection mechanisms can be provided for users. If a goal is not covered by any template, an “axiom editor” with an easy-to-use graphical user interface can be used to construct the new goal template.

Discovery

Several variations for SWS discovery exist (Keller et al. 2004; Benatallah et al. 2003; Sycara et al. 2004). The simplest way is based on keyword matching within the descriptive metadata of web services (similar to the UDDI approach). The most complex way is to logically prove that the web service is able to fulfill the given goal. Currently, response time of discovery process is a significant tradeoff for these approaches.

The keyword-based method can be improved if the goals and capabilities are used for keyword generation instead of metadata. Goals and capabilities are described using ontology terms, so more homogeneous and precise keywords can be extracted this way.

The logic-based methods can differ in aspects of goals and capabilities considered. It is simpler and faster to match only by postconditions, but then matching services might not be executable because of missing or unacceptable information (input). There are experiments for building hierarchies and other data structures acting like an index for semantic web

services, enabling quick location of matching pre- and postconditions. (Li and Horrocks 2003).
Our approach is to split the discovery into two phases. In the first phase, the keywords of capabilities are used to filter the possible web services. Then, logical matching is applied only for the filtered services.
The list of matching web services is returned to the user application enriched with descriptive and qualifying metadata. Qualifying metadata is automatically collected in the INFRAWEBS system, and measures aspects such as reliability or availability of the web services, helping the selection between services.

Selection and Invocation

Selection of the best web service is not equivalent to the selection of the best service. The terms service and web service should not be confused (Keller et al. 2004). Web service is the access point to the real service: its quality determines the quality of access, but might be independent of the quality of service (e.g. automatic translation or flight reservation). Therefore, an iterative selection process (similarly to iterative query refinement in information retrieval) guides the selection of the best service in INFRAWEBS. This is based on a simple two phase workflow agreement (or business logic) between the web services and the middleware layer.
The first phase is for collecting service offers, and the second phase is for executing the selected service. A service offer characterizes the details of the service execution. In case of flight reservation, a service offer might contain the details of the flight, the price and the payment method. Such details cannot be advertised in the web service capabilities, but are essential for the users' decision.
Service offers are collected in parallel by the SUA. During this phase, no irreversible commitments take place, services only inform the user about the expected result of the service execution. Additional information might be required by certain web services (e.g. can payment be done using credit card?). This kind of question answering is part of web service choreography, which is, at the moment, rather poorly specified on the level of semantic web services. In our design, questions will ask for the values of instance attributes defined in the input ontology. These values can be given by the user application or the SUA itself, based on the previous answers of the same user session.
After the first set of offers received, the user might refine the goal and/or change the selected services and re-collect the offers. Thus, selection is a looping process, which ends when a single service is chosen.

Execution and Result

Execution of the selected service is the combination of internal processing, choreography and orchestration. These are performed within the "SWS world", according to the well-defined rules of SWS execution. This happens in the SWS Executor component of INFRAWEBS, with the exception of choreography events concerning the user, such as additional questions and request for new information. Translation of these questions into logic-neutral form which can be interpreted by user applications is target of ongoing research. In case of simple attribute queries, which occur most often, SUA offers a default translation method.
The SWS Executor also collects information about the execution and forwards them to a Monitoring Agent. This agent maintains quality metrics about web services, and makes those available for helping the web service selection process.
Finally, the result of the execution has to be translated for the user application. This is a conversion of facts from logical format into object-oriented data structures. The conditions of a successful conversion algorithm are the use of a common ontology and a required level of simplicity for facts, which can be easily guaranteed in most e-business scenarios.

SUMMARY

The SUA component of INFRAWEBS presented in this paper serves as a generic middleware for deployment of Semantic Web Services. The SUA provides support for the usual steps of goal construction, discovery, selection and execution of semantic web services. This support is achieved through a neutral interface, which hides the complexities of Semantic Web Services, therefore applications can be easily adapted to it, and also it can be equivalently used in variants of Semantic Web Services, such as WSMO and OWL-S.
SUA also features an iterative selection refinement process for finding not only the suitable web services, but also the best service offers for users' goals.

REFERENCES

Benatallah, B; Hacid, M; Rey, C and Toumani F. 2003. "Towards Semantic Reasoning for Web Services Discovery", In *Proc. of the International Semantic Web Conference* (ISWC'03), Springer Verlag, Sanibel Island, Florida, USA Oct 2003.

Kavantzas, N.; D. Burdett; G. Ritzinger; T. Fletcher and Y. Lafon (editors). 2004. "Web Services Choreography Description Language Version 1.0", *W3C Working Draft*

Keller, U.; R. Lara and A. Polleres (editors). 2004. "WSMO Web Service Discovery", *WSML Working Draft* 12 11 2004

Li, L. and Horrocks, I. 2003. "A software framework for matchmaking based on semantic web technology". In *Proceedings of the12th International Conference on the World Wide Web*, Budapest, Hungary.

Nern, H. J.; Pautzke, F. and Atanasova T. 2003. "New Approach for Online Case-Based Reasoning WEB-Services in Knowledge Management Applications", *Euromedia 2003*, Plymouth, United Kingdom, April 14-16, 2003

Nern, H. J.; Atanasova, T., Agre, G. and Saarela, J. 2003. "INFRAWEBS - Intelligent Framework for Networked Businesses and Governments Using Semantic Web Services and Multi-Agent Systems", *7th WSEAS International Conference on Computers* , WSEAS, Corfu, Greece, 2003

Roman, D. (editor). 2004. "Choreography in WSMO", *WSMO Working Draft* 12 November 2004

Sycara, K. P.; Paolucci, M.; Soudry, J.; Srinivasan, N. 2004. "Dynamic Discovery and Coordination of Agent-Based Semantic Web Services", *IEEE Internet Computing* 8(3): 66-73 (2004)

SEMANTIC INFORMATION ROUTING AS A PRE-PROCESS FOR SEMANTIC WEB SERVICE GENERATION - SIR & OM

T Westkaemper
Profium Ltd.
02600 Espoo, Finland
timo.westkamper@profium.com

J Saarela
Profium Ltd.
02600 Espoo, Finland
janne.saarela@profium.com

H Joachim Nern
big7.net GmbH & Aspasia Knowledge Systems
40210 Duesseldorf, Germany
info@aspasia-systems.de

KEYWORDS

Semantic Web Services, Semantic Information Routing, Web Services, Business Logics, Query Languages, SPARQL

ABSTRACT

In this paper the main semantic pre-processing module in the Infrawebs framework – the Semantic Information Routing module (SIR) – in connection to an organizational memory (OM) is described. This module is dedicated to the pre-treatment of semi-structured knowledge artefacts provided by the OM, which are semantically enriched and adequately processed for routing to Semantic Web Services designing and composing cycles (SWS – Designer, SWS-Composer). The paper presents the basic coherences with the framework modules and describes the basic products the modules are based on.

INTRODUCTION

This paper presents the Broker side architecture of the INFRAWEBS project [1], an EU founded specific research project with the aim to build an environment for the design, composition and execution of Semantic Web Services (SWS).

The main partners involved in the design of the Broker side of the INFRAWEBS project are Aspasia Knowledge Systems from Duesseldorf, Germany, a consulting company specializing in Knowledge Management and Portal design, and Profium Ltd., a software company located in Espoo, Finland specializing in Semantic Web and Content Management consulting and application design.

OBJECTIVES OF THE INFRAWEBS PROJECT

The main objective of the INFRAWEBS project is to develop an ICT framework (Information and Communication Technology) which enables the generation of an open and extensible platform for the generation and execution of Semantic Web-Service applications. The main users will be software providers which integrate their proprietary systems into the INFRAWEBS framework as well as service providers, which register their services into the system.

The INFRAWEBS framework provides a loosely-coupled set of adaptable system components and tools to analyse, design and maintain Semantic Web Services in the whole service life cycle: discovery, selection, execution and monitoring.

PRIMARY OBJECTIVE OF THE RESEARCH PROJECT

The challenges and main objectives for INFRAWEBS are to develop a framework to:

- provide a comfortable and "easy to use" knowledge brokering unit,
- used within a collaborative workflow environment for analysis, design and composition of SW services
- supplemented with security and privacy preserving run-time and maintenance tools

The architecture of the INFRAWEBS framework is based on a two-tier architecture, which defines the two tiers *Client & Provider* and *Broker*. The two tiers are further divided into sets of modules which are mostly customizations of already existing software products.

The *Client & Provider* acts a *Client* in relation to already existing fine-grained Web Service instances and as a *Provider* in relation to the clients of the Semantic Web Service instances. The Broker component is where the registration of non-semantic Web service instances takes place. Additionally further knowledge items related to the service descriptions are provided via the modules of this tier.

The component diagram in **Fig. 1** illustrates the set of modules of the two tiers of the INFRAWEBS framework. .

As the focus of this paper is on the Broker side of the INFRAWEBS system, the components of the Client & Provider side will only be presented briefly. The components of the modules WP4 and WP5 are design time components for the design, composition and storage of semantically enhanced service descriptions.

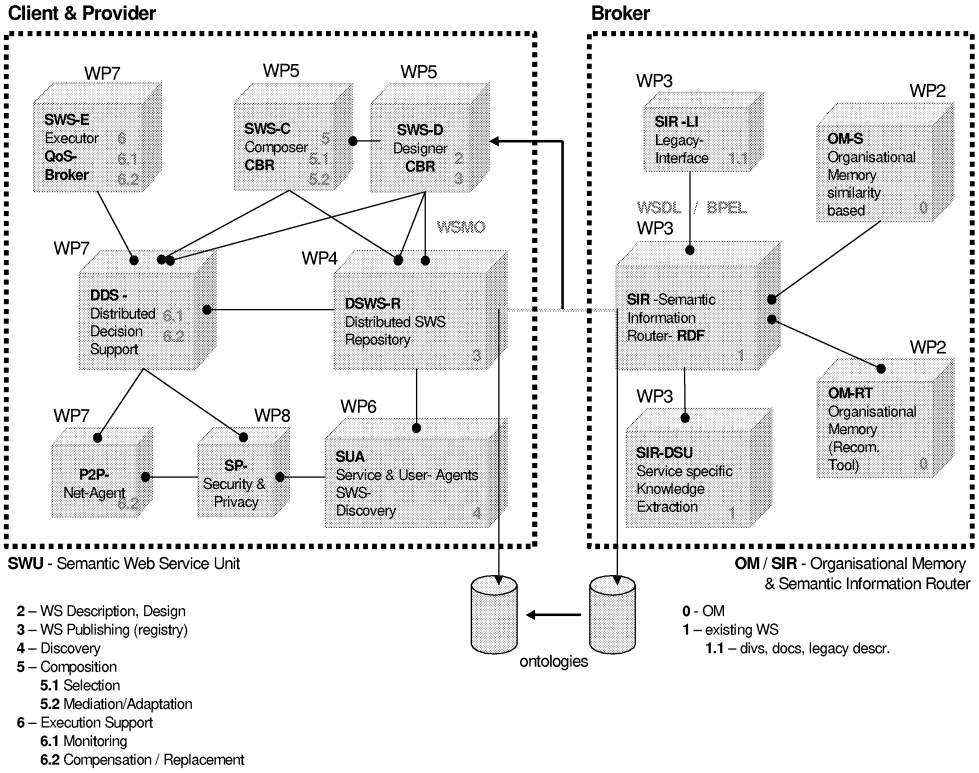


Fig. 1: Initial Design of the INFRAWEBS framework

The components of the modules WP7 and WP8 are related to the execution and monitoring aspects of the INFRA-WEBS framework.

SHORT INTRODUCTION TO WSMO

The Web Service Modeling Ontology (WSMO) [2] is used as the conceptual framework for the service and goal descriptions on the *Client & Provider* side of the INFRAWEBS framework.

WSMO is divided into three incremental conceptual layers. *WSMO-Lite* provides a basic subset of constructs for the design of Semantic Web services. *WSMO-Standard* provides a more mature set of constructs and finally *WSMO-Full* offers the full conceptual model and ontology of the WSMO framework.

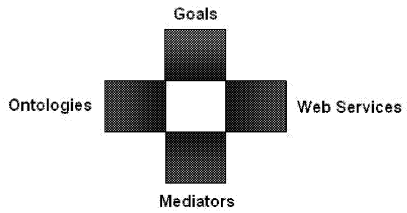


Fig. 2: Conceptual view on the WSMO framework

The concepts of the WSMO ontology are based on the Web Service Modeling Framework [3] by Christoph Bussler and Dieter Fensel. Both scientists are chairs of the WSMO working group. The concepts of the WSMF framework are *ontologies* which provide the terminology and related semantics used by the other elements, *web services* which describe the capabilities and interfaces of Semantic Web Services, *goals* which define the objectives of the service users and *mediators* which resolve interoperability problems between goals, ontologies and web services.

As the Broker side of the INFRAWEBS framework models Web services primarily via RDF data models [4], the WSMO framework won't be presented any further in this paper.

PRE-EXISTING SOFTWARE USED ON THE BROKER SIDE

The main products used on the Broker side of the INFRAWEBS framework are the SIR (Semantic Information Router) application by Profium Ltd. [5] and the FCS-OM (Fuzzy Concept Based Organizational Memory) application by Aspasia Knowledge Systems [6]. Both products will be presented in short in the following.

The Organizational Memory component by Aspasia Knowledge Systems is mainly used to structure and bundle atomic knowledge items, which are fed into the system as additional information to the formal service descriptions.

The Semantic Information Router (SIR) component on the other side is used to handle the formal descriptions that are submitted in the service registration process. The formalism used for data representation in the SIR component is based on the RDF framework [4] which describes knowledge structures as directed graphs where the nodes refer to instance references via URIs or ordinary data values via literals.

The sources from which documents and knowledge in form of RDF based metadata are fed into the SIR system

are legacy systems (ERP), which are mostly connected to SIR via SOAP interfaces, the service registration interface which will be implemented as a simple web-based GUI and the interface to OM, which is SOAP based as well.

SOAP is a W3C standard for XML based communication over various protocol stacks such as HTTP, SMTP etc. It is the de-facto standard for Web Service based interfaces.

API DESIGN OF THE INFRAWEBS BROKER SIDE

In the following the main components and interfaces of the Broker side will be presented. The components used are the Semantic Information Router (SIR), a set of legacy interfaces bound to it (SIR-LI instances), the Service-specific Knowledge unit and tools for Similarity based Knowledge Extraction combined with a Recommender Tool based on the Organizational Memory (OM) application.

SOAP interface to OM

The interface between the Semantic Information Router (SIR) and the Similarity based knowledge organization tool is based on the SOAP protocol. The interface is used to provide WSDL based service descriptions to the OM component and to query for content item references that are related to given service description references.

The message exchange patterns used in the communication between SIR and OM are the following :

- Retrieval of document references (of OM document space) via a set of keywords:

request from SIR : SOAP request with keywords

response from OM: SOAP response with document references (HTTP URIs) and metadata (mostly Dublin Core based).

- Retrieval of document references that are related to a given document URI within the SIR document space:

request from SIR : SOAP request with HTTP URI of document (which references a document within the SIR)

response from OM : SOAP response with document references (HTTP URIs) and metadata (mostly Dublin Core based) and a list of similar WSDL files.

- Retrieval of document references that are related to a given document URI within the OM document space:

request from SIR: SOAP request with HTTP URI of document (which references a document within the OM)

response from OM: SOAP response with document references (HTTP URIs) and metadata that are close to the document referenced by the given HTTP URI

SPARQL interface for querying service descriptions from SIR

The query interface of the SIR component is based on the SPARQL protocol. SPARQL is a RDF metadata query language designed by the W3C DAWG working group with protocol bindings defined for JDBC, HTTP and other protocol stacks [8]. Profium Ltd. is one of the companies present in the W3C DAWG working group and is going to provide a compliant implementation in the Semantic Information Router (SIR) product.

Technically the SPARQL interface is a metadata query language with syntax close to SQL which offers querying metadata as table and graph result sets. The protocol stack most suitable for J2EE environments is a JDBC type 4 (Java Database Connectivity) compliant driver [9]. Profium Ltd. is also involved in the implementation of a JDBC based protocol binding for SPARQL named SPARQL4j (http://sourceforge.net/projects/sparql4j).

Service registration in SIR

Registration of non-semantic atomic services is accomplished via a web-based GUI interface in the scope of the INFRAWEBS framework. The interface is mainly used to input WSDL and BPEL4WS based service descriptions and enter additional related non-functional properties. Additionally a UDDI registry interface might be provided for the SIR component. This could be established via SOAP as a UDDI Subscription Listener.

REFERENCES

[1] Infrawebs IST Consortium Website http://www.infrawebs.org/

[2] Bussler, C.: D17v0.2 WSMO Tutorial. WSMO Working Draft 17 January 2005 http://www.wsmo.org/2004/d17/v0.2/20050117/

[3] Fensel, D., Bussler, C.: *The Web Service Modeling Framework WSMF.* Electronic Commerce Research and Applications, 1(2), 2002.

[4] Manola, F., Miller, E. : RDF Primer. W3C Recommendation 10 February 2004 http://www.w3.org/TR/rdf-primer/

[5]Semantic Information Router (SIR) http://www.profium.com/products/sir.html

[6] H Joachim Nern, A Dziech, E Tacheva, "Fuzzy Concept Sets Applied to Semantic Organizational Memories within the Semantic Web Services Designing and Composing Cycle", Euromedia 2005, April, Toulouse, France

[7] Booth, D., Liu, C. K.: Web Services Description Language (WSDL) Version 2.0 Part 0: Primer http://www.w3.org/TR/wsdl20-primer/

[8] Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF. W3C Working Draft 12 October 2004 http://www.w3.org/TR/rdf-sparql-query/

[9] sparql4j Project Summary - SPARQL JDBC type 4 driver, http://sourceforge.net/projects/sparql4j/

FUZZY CONCEPT SETS (FCS) APPLIED TO SEMANTIC ORGANIZATIONAL MEMORIES WTHIN THE SEMANTIC WEB SERVICE DESIGNING AND COMPOSING CYCLE

H Joachim Nern, E Tacheva

Aspasia Knowledge Systems
D-40210 Duesseldorf, Germany
info@aspasia-systems.de

A Dziech

Group of Automatic Control
University of Wuppertal, Germany
dziech@uni-wuppertal.de

KEYWORDS

Fuzzy Concept Sets, Organisational Memory, Semantic Web, Semantic Web Services

ABSTRACT

As a repository for semi-structured knowledge artefacts a Fuzzy Set based organisational memory (FCM-OM) as bases for recommender issues is presented in this paper. It is an adaptation of existing system kernels and classification approaches already presented in [1], [2], [3], [4], [5], [6]. The OM management module is endowed with tools for knowledge acquisition and knowledge representation. Within the related Infrawebs project the discussed module is responsible for the collection, organisation, refinement, and distribution knowledge objects handled and managed by (Web) service providers.

INTRODUCTION

Adoption of the SOA, the service oriented architecture, is expected to enable entities, enterprises, public organisations as well as private institutions to "contract-out" their "non-critical" functions – there specific services. In the upcoming world of networked businesses, where processes transcend within entity specific departments (EAI – enterprise application integration) as well as organizational boundaries. This service oriented technique is expected to provide the ideal conditions to automate these different "advertised" processes.

The current technology based on SOAP, WSDL and UDDI addresses only syntactical aspects of a Web service. On this basis just "rigid" (static) services can be designed and invoked. Rigid and non-dynamically in the sense, that they cannot adapt and reflect changes - from automatic control point of view: **disturbances** - of the action environment without human intervention.

They are not "really" automatically driven – there action cycles depend massively on semi-automatic structures, due to missing closed loop features, as the characteristically feature of automation processes. Accordingly new streams and project initiatives are coming up to increase the degree of automation and accordingly the flexibility of service providing and service requesting capabilities.

The main objective of the INFRAWEBS project is to develop an ICT framework which enables the generation of an open and extensible platform for the generation and execution of Semantic Web-Service applications. The main users will be software providers which integrate their proprietary systems into the INFRAWEBS framework as well as service providers, which register their services into the system. The INFRAWEBS framework provides a loosely-coupled set of adaptable system components and tools to analyse, design and maintain Semantic Web Services in the whole service life cycle: discovery, selection, execution and monitoring. The Infrawebs approach follows strict the bottom up paradigm: it combines and couples knowledge management issues - representing the bases - with Semantic Web Services issues - representing the top of service advertising and offering environments. The specified organisational memory – presented in this paper – has the main task to handle and pre-process semi structured knowledge objects and finally "shift" this "tuned" objects to the (Web) service designing and composing modules within the Infrawebs framework.

CURRENT APPROACH

The current specification of the Infrawebs OM is based on the novel results of the research activities in the area of Fuzzy Set theory:

Considering the ambiguity, imprecision of concepts (electronic knowledge objects of an entity, including the objects handled and received in Internet related environments) an useful approach is the adaptation and application of FCM (Fuzzy Conceptional Matching) methods.

Within this approach a "concept" is defined (and represented) by a sequence (a set) of weighted keywords. Ambiguity in this concepts is defined by a set of imprecise concepts [7]. Whereas each imprecise concept is defined as a set of fuzzy concepts (using methods of implicit semantics), which is related to "a set of imprecise terms representing the context".

Involving and considering also formal semantics, this imprecise terms (words) are "translated" into precise terms (words) formalised as an ontology. [7].

In this sense two streams have been focused in realising this system component:

- one component of the OM (FCM-OM) is designed following the rules of **implicit** & **soft semantics** (using statistical based AI methods like clustering and classification)

- a second component (O-OM) reflects the Knowledge handling based on methods related to **formal semantics** (Ontologies) – hard semantics.

In this paper the first approach is considered.

RELATED WORK IN ORGANIZATIONAL MEMORIES

Since several years researchers from various fields like AI, CSCW, IR, SW and Knowledge Engineering addresses issues from Organisational Memories (OM).

Generally an Organisational Memory includes and embraces the complete knowledge of an organisation, which is mainly represented by their knowledge, experience, expertise – covered in document archives (electronic & paper based). The objective for handling an OM is to effectively enhance the individual's access to organisational information based on optimising structures, whereby the overall goal should be to improve the competiveness of an organisation by improving the way in which it manages its knowledge [8].

In the following section several approaches from different application and research fields will be shortly overviewed here specifically - CSCW - Computer Supported Co-operative Work.

The application of OMs in this area is often called: Group Memory, reflecting the underlying feature of supported user groups: example maybe the "Knowledge Depot" which is reported in "*From Group Memory to Project Awareness through the use of the Knowledge Depot*" Kantor & Zimmermann [9] and yields in archiving a.o. email documents. It is focused on formalising "project awareness" in the sense of comprising the awareness of discussions, decisions and changes during the project work. Classifying incoming emails by organising the group memory in a dynamically improved hierarchical section.

Also reported in [10] "*A Framework for Assessing Group Memory Approaches for Software Design Projects*" is an Project Compendium, which is aimed to capture formal documents (e.g. project reports). It installs a hypertext space consisting of documents of different formats and origins, which serves as a basis for "*Conversational Modelling*".

FCS-OM – FUZZY SET BASED ORGANISATIONAL MEMORY

Due to the arbitrary and varying types of knowledge objects and concepts handled within the environment of service provider (especially with respect to the given use case formulated in [11]) the decision have been made to realise in this WP2 an Organisational memory, which follows two (three) main development streams using:

- **implicit semantics** of concepts on the one hand and
- **formal semantics** of concepts on the other hand,

whereby the implicit one yields indirectly but finally in an

- **soft** (powerful) semantics [7], [12] approach.

The type of **Implicit Semantics** refers to the kind that is *implicit given in the patterns of data and that is not represented explicitly in any strict machine processable syntax* [7]. For instance, this patterns are given as a numeric similarity judgement between different knowledge objects in the corpus (the entity specific knowledge pool).The exploitation of this implicit semantics draws upon well know statistical techniques – clustering, classification, categorization etc. Basically all machine learning exploits implicit semantics, namely clustering, concept- and rule learning, Hidden Markov Models, ANN and others. In early steps towards the Semantic Web such techniques are used like for instance in the Vivisimo search engine approach, metadata extraction in the Web Foundation Technology [13], automatic classification and automatic metatdata extraction in Semagix Freedom [14].

It`s obvious to state that computers lack the ability to disambiguate and understand complex natural language. Statements or facts need to be expressed in a way that computers can process them. Semantics that are represented in some well styled syntactic form (governed by syntax rules) is referred to as **Formal Semantics**.
The features making a language formal are as follows:

- expressions in a formal language are interpreted in models
- "*principle of compositionality*" [7], [12]: the semantics of an expression is computed using the semantics of its parts, obtained using an interpretation function

From a less technical perspective, formal semantics means machine processable semantics. Recently DL (Description Logic) have been the dominant formalism for knowledge representation. **Formal semantics** of DL is based on a set theory. A principally drawback of using DL is the missing capability of representation of degrees of concept memberships and degrees of assignment and relationships between knowledge objects (concepts).

Soft semantics, fuzzy statistical analysis of data allows the exploration of relationships (determining the **degree** of relationship and set assignment as well as set membership) [7], [12]. Soft semantics combines the benefits of both worlds (implicit – formal semantics): hierarchical composition of knowledge and statistical analysis; reasoning on available information, but with the advantage that it can be formalised in a common language using statistical methods.

The used and partially investigated approach in this WP will try to follow this harmonisation efforts of the streams of formal and implicit semantics, which finally yields in a handling of "**SOFT SEMANTICS**".

In this section the knowledge object (concept) brokering, consisting of the filtering, classification and management processes are described. In the first stage as knowledge representation for the objects the modified (extended) version of the Dublin Core is used.

CONCEPT FEATURE VECTOR GENERATION

The filtering and classification process yields in a feature selection. The filtering and pre-cleaning process produces a structured feature term vector consisting of weighted terms. The preparation procedure is divided into 3 stages:

* a rough pre-cleaning,
* a (word) stem oriented filtering and
* a normalization resp. unification of the term vector based on a given set of keywords

The filtering and normalization is realized as a fuzzy matching procedure. Subsequently the generated feature vector undergoes a classification process, which is performed using statistical similarity approaches for determining relationship (and membership) functions

among the objects themselves resp. the pretended subject term vectors, to which the objects are allocated.

Meta Data Information Scheme

The given knowledge objects are represented by meta-data-information schemes and handled as meta data information objects. The well known Dublin Core is used as the knowledge representation layer. For considering specified DK dependent information (NFP – non functional properties) and characteristics this standard Dublin Core is enhanced with additional segments to the Extended Dublin Core (EDC). Special chosen vector components (like title, author, description, individually given keywords) are summarized and consolidated to one entire text object which is processed within the filtering and classification module.

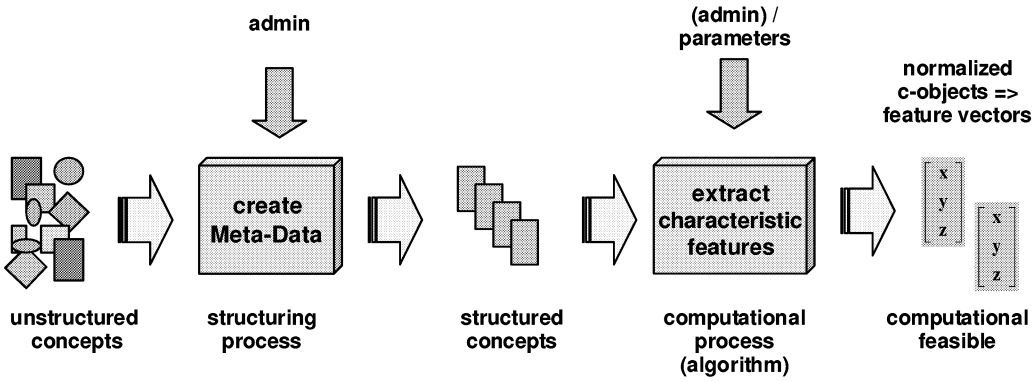


Fig. 1: Concept Feature Profile Generation

Term Filtering & Feature Vector Extraction

The first pre-cleaning process as a rough extraction process is applied to the text objects. In the following stage a further extraction procedure is started to check non-information terms and especially word stems and to generate a rough term vector.

A dynamically expanding dictionary provides non-info-terms and word stems which are fuzzy matched with terms in the term vector. This procedure affects like a unification/normalization based on word stems. A further applied fuzzy matching process using a predefined dictionary of standard keywords extracts and weights implicitly terms and normalize the term vector to the given set of standard keywords.

The dictionary of standard keywords is also maintained by the administrator. In such a way the keyword normalization is additionally combined with a weighting of terms, whereby this rating of terms is based on a statistical consideration of the distribution measure as well as pre defined preferences. This procedure yields in creating a final pre-cleaned and pre-filtered Concept Term Vector (CTV) consisting of weighted elements.

CONCEPT - CLASSIFICATION: CLUSTERER – TRAINER – CATEGORIZER

The suchlike generated CTVs representing the diverse concepts are processed in further different procedures according to there progress stage and properties:

* clusterer: a bottom up clusterer for generating concept classes

* trainer: a supervised trainer for generating predefined concept classes

* categorizer: a procedure for categorizing arbitrary concepts

Base condition for this different procedures is the determination of concept similarity. For calculation of overall similarity matrices a modified combination of the MinMax & Dice scheme is used:

Based on this similarity calculation Fuzzy Concept Sets are investigated and used either within the Clusterer stage or the trainer stage to generate new or predefined classes resp. (Super-) Fuzzy Concept Sets.

CLUSTERER: 3-STAGE BOTTOM UP CLUSTERING OF FUZZY CONCEPTS SETS

A three stage bottom up clustering is processed by using stepwise the Cluster Algorithm and processing the similarity matrix of the concepts given in the collection.

For each generated class a centroid concept vector (CCV) is calculated out of the CTVs referenced to the specified class. Accordingly the CTVs represent the nodes of a three-layered object/subject tree. The overall clustering stage yields in coupled subject classification trees, whereby their relations are expressed by relationship and/or membership functions.

As illustrated in **Fig. 2** the three stage bottom up clustering of a fuzzy set of concepts the

- **root fuzzy set**,

yields in a set of

- **fuzzy medion sets**
- **fuzzy bottom sets**
- **fuzzy element sets**,

whereby these fuzzy element sets finally yield in atomic

- **fuzzy word sets**

with related assignment resp. membership functions.

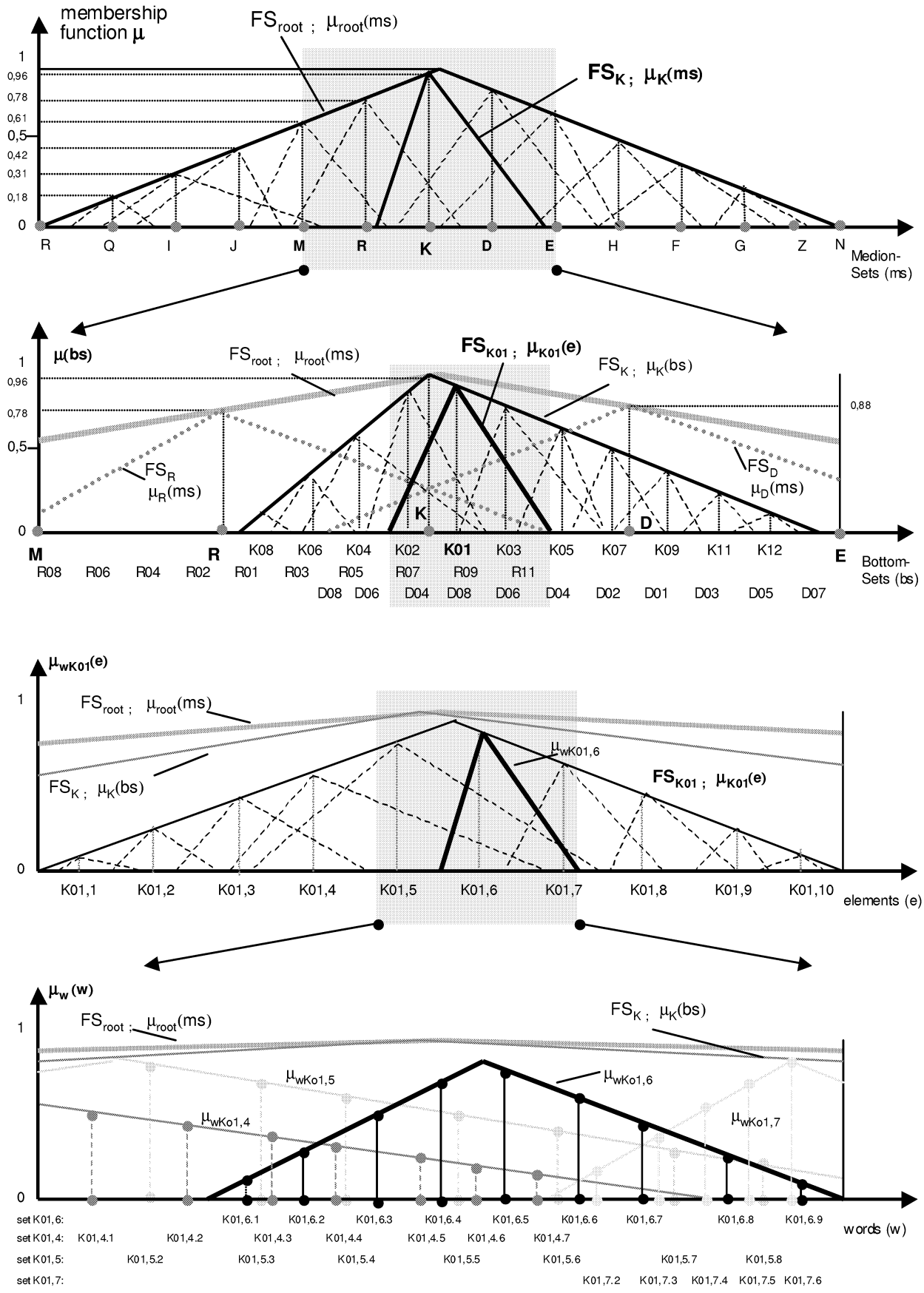


Fig. 2: Three stage bottom up clustering of concept sets yielding in: root set, medion sets, element sets, word sets

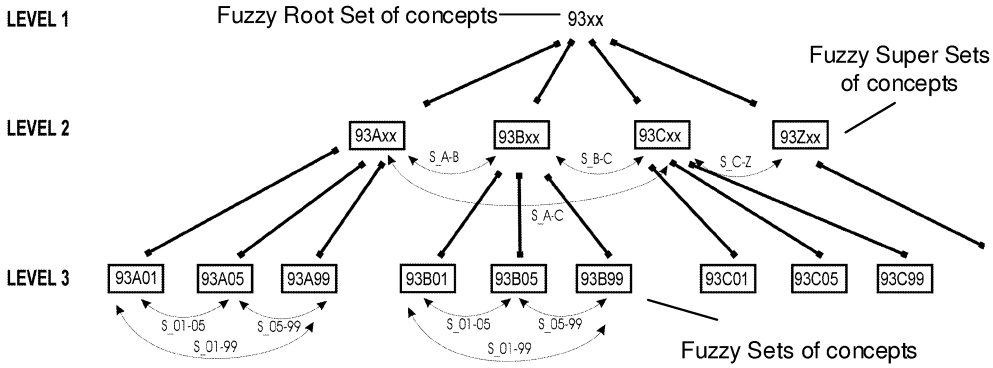


Fig. 3: Tree structure of a fuzzy root set of concepts, consisting of fuzzy super (medion) sets of concepts, consisting of fuzzy sets of concepts, consisting of fuzzy sets of elements, consisting of fuzzy sets of (atomic) words

The basic structure is still a tree structure as illustrated in (**Fig. 3**).

TRAINER: ESTABLISHING FUZZY SETS USING A-PRIORI KNOWLEDGE

In the case of processing concepts known to the administrator a trainer is realized, which generates and determines concept classes out of categories of predefined concepts. These classes are furthermore referenced to the concept set classification tree, whereby each node is again represented by a CTV.

CATEGORIZER: CLASSIFICATION OF FUZZY SETS

The intrinsic classification of new concepts consists in the first stage of an estimation of a assignment function (coefficient) between the CTV of the new concept and the existing normalized concept vector of the concept centroids. The resulting membership vector – consisting of the calculated assignment coefficients – is a measure for the degree of relationship between a group of new concepts and the concept centroids. A further three layered concept clustering of the concept vector leads finally to three (at least two) classes of concepts belonging to determined concept centroids representing the classes within the concept classification tree.

References

[1] Nern H.-J., Dziech, A., Rusc T., "Filtering and Classification of Meta-Data-Information Objects for Knowledge Sharing within an AI-Based Research Network" Euromedia 2001, 18-20th April 2001, Valencia, Spain Euromedia 2001

[2] Nern H.-J. et al "Realisation of a Task Oriented Recommender System applied to XPS-Bases Research Networking", Euromedia 2001, 18-20th April 2001, Valencia, Spain Euromedia 2001

[3] Nern H.-J., Atanasova T. et al "Decision Strategies for Rating Objects in Knowledge-Shared Research Networks", Proc.of 4th World Multiconference on Circuits, Systems, Communications & Computers (CSCC 2000), Athen, Greece, July 10-15, 2000,

[4] Grancharova A., Nern H.-J., et al; "Decision Support Unit as Part of a Global Knowledge-Shared Research Network", Euromedia 2000, May 8-10, 2000, Antwerp, Belgium

[5] Atanasova T., Nern H.-J., et al; "Multi-Agent Approach for Task Related Decision Supported Information Retrieval", Euromedia 2002, April 2002, Modena, Italy

[6] Nern H,-J. et al; "Final Report - Knowledge Shared XPS-Based Research Network Using Multi-Agent Systems", European project: ESPRIT HPCN : 977113 KNIXMAS; http://www.knixmas.de - http://www.cordis.lu/esprit/src/977113.htm , 2001

[7] Lotfi A. Zadeh. From Search Engines to Question-Answering Systems - The Need for New Tools. Proceedings of the 1st Atlantic Web Intelligence Conference, 2003

[8] G. van Heijst, R. van der Spek, E. Kruizinga. "Organising Corporate Memories", Proc. of 10th Knowledge Acquisition for Knowledge-Based Systems Workshop, 1996.

[9] M. Kantor, B. Zimmermann, D. Redmiles "From Group Memory to Project Awareness through use of the Knowledge Depot", Proc. of the California Software Symposium (CSS'97), California, 1997

[10] B. Zimmermann, A.M. Selvin. "A Framework for Assessing Group Memory Approaches for Software Design Projects", Proc. of DIS'97

[11] Deliverable 10.1.1, Atos Origin, Infrawebs project

[12] Lotfi A. Zadeh. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. In Journal of Statistical Planning and Inference 105 (2002) 233-264.

[13] Stephen Dill, Nadav Eiron, David Gibson, Daniel Gruhl, R. Guha, Anant Jhingran, Tapas Kanungo, Sridhar Rajagopalan, Andrew Tomkins, John A. Tomlin, and Jason Y. Zien. Semtag and seeker: bootstrapping the semantic web via automated semantic annotation. In Proceedings of the Twelfth International Conference on World Wide Web, pages 178--186. ACM Press, 2003

[14] V. Kashyap, C. Ramakrishnan, C. Thomas, D. Bassu, T. C. Rindflesch and A. Sheth TaxaMiner: An Experimentation Framework for Automated Taxonomy, lsdis.cs.uga.edu/~c-thomas/publications/taxaminer.pdf

APPLYING CLUSTERING METHODS TO DISTRIBUTED DECISION MAKING IN SEMANTIC WEB SERVICE DESIGNING AND COMPOSING CYCLES

H Joachim Nern
Aspasia Knowledge Systems
D-40210 Duesseldorf, Germany
info@aspasia-systems.de

A Dziech
Group of Automatic Control
University of Wuppertal, Germany
dziech@uni-wuppertal.de

Tatiana Atanasova
Institute of Information
Technologies
Acad. G. Bonchev 2, 1113 Sofia
Bulgaria
atanasova@iinf.bas.bg

KEYWORDS
Clustering, Optimization, Distributed Decision Support, Semantic Web Services, Semantic Web

ABSTRACT

Realization of the full potential of the Web services requires further technological advances in the areas of service interoperation, service discovery, service composition and orchestration. A possible solution to these problems can be provided by application of Semantic Web technologies. Within collaborative environments and platforms for SWS generation and handling like the one proposed and developed in the Infrawebs project an inherent capability of distributed decision making is given. This distributed decision making serves as an implicitly applied system optimization. In this paper an approach is presented for applying clustering techniques to fulfill decision making, tuning and optimization tasks in SWS designing and composing cycles.

INTRODUCTION

In [1] the partners of the Infrawebs project have described the main structure of the Infrawebs framework: it consists of an integrated approach of knowledge brokering units in conjunction with SW service (SWU) designing, composing and execution modules.

The semantic web service unit (SWU) is designed as a semantic and ontology based module handling and maintaining semantic web services as a workflow and knowledge sharing component - as a modern type of SW service based collaboration platform and interoperable middleware.

The networking and platform partners provide SW services, which are incorporated in their own services product range (composition). On the one hand, it is an expansion of the own services, on the other hand, the creation of new products and services is facilitated to develop value added products and services.

Applied distributed "Rule Bases" enable the capability of distributed decision support in collaborative SW service generation and maintenance cycles. In this sense the issues of collaborative decision making (Distributed Decision Support – DDS-Unit), Quality of Services monitoring and service execution is considered in the DDS and the SW service execution module. The SWS execution module provides a run-time environment for running SW services and monitoring the service execution process considering QoS metrics. Extracted QoS specific data patterns are feeded back to the designing and composing modules and guarantees in this way a closed loop optimisation cycle (**Fig. 1**).

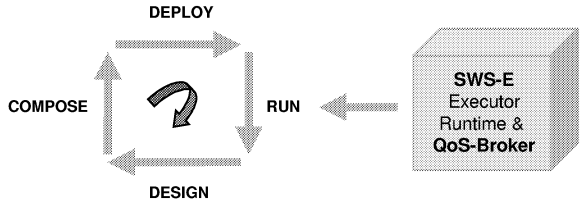


Fig. 1 The QoS-Broker within a closed loop process

In this paper the focus is set to apply the DDS approach to the designing and composing process in the SWS Designer and the SWS composer [2], [3], [4].

The behaviour patterns of the service designing and composing process are formalized and archived as design / composing rules and stored in local as well as global rule bases (Distributed Rule Bases for Distributed Decision Support - DDS).

In this paper the application of clustering techniques for enabling implicit optimization of the designing and composing process is proposed and presented.

EXISTING ALTERNATIVE APPROACHES WITHIN INFRAWEBS

In [2] further approaches to Distributed Decision Making in Semantic Web Service design and maintenance cycles are pointed out. This approaches are based on the fact that "Semantic Web Services are characterized according to their Quality of Service (QoS)", which includes their timeliness, quality of products delivered, cost of service, and reliability.

The DDS-Unit (Distribution Decision Support-Unit) [2] has the main task to monitor SWS execution and provide compensation mechanisms to repair "failed" services, where by the error handling of the service execution

process is based on a set of rules. The type of rules are distinguished into domain independent (so called global) resp. dependent (local) rules.

The (global) set of rules is mirrored into every Semantic Web Service Unit and concerns general planning and monitoring issues, whereas . The local rules are domain dependent and are specific for every particular set of semantic web services, connected to given ontology and business logic /05.At/. Distributed rule base contains experience about successful or failed execution of SWS.

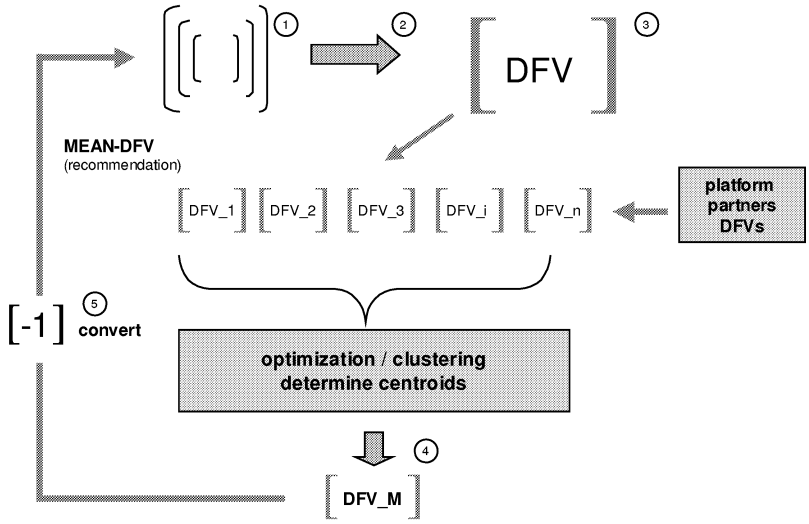


Fig. 2: Conceptual Design of the basic DDS cycle using SWS - designing and composing rules

Rule-Based Patterns [2] for DDS are constructed from the multi-level models of e-business templates for: Business logic, Integration, Composition, Application and Execution.

MAIN IDEA AND CONCEPTIONAL APPROACH OF THE INFRAWEBS DDS-UNIT BASED ON THE DESIGNING AND COMPOSING PROCESS OF SWS

In **Fig. 2** the DDS conceptual approach is presented as an information processing and closed loop cycle, which considers rules representing the SWS design and composing process, a transformation of linguistically formulated rules to numerical patterns, a clustering stage and a final conversion and recommendation phase.

The main idea of this approach is to represent the SWS designing and composing process by a set of rules (*If* … *Then* …), which are transformed to a numerical representation level and finally optimised by selecting centroids out of the resulting set of similar patterns processed within a clustering stage.

1 - The admin (expert, user) of the designer (resp composer) module defines according to his experience a set of rules, which describe the suitability and usefulness of the system.

2 - This set of rules (*If, Then* clauses) is transformed into a set of numerical patterns.

3 - This numerical patterns are formed to a computational feasible feature profile (vector DFV)

4 - The DFVs provided by different (distributed) platform partners (INFRAWEBS) are collected and constructed as a set of DFVs. This set of DFVs is multiple bottom up clustered to generate a unique set of centroids representing the DFV sub clusters. The finally chosen main centroid represents the optimised DFV: the mean DFV (DFV_M) – in this sense the result of a mean value building (optimization) process according to a low pass filtering process respectively collaborative filtering process).

5 – The extracted (numerical) DFV_M is used for recommendation purposes for the existing rule base, which is finally modified accordingly.

Due to the optimizing and recommendation feature, based on the analysis and evaluation of aggregated distributed knowledge patterns (DFVs from distributed platform partners) the whole procedure is characterized as an "implicit" distributed decision making process.

COMPONENTWISE GENERATION OF FEATURE VECTORS REFLECTING DIFFERENT DESIGN STAGES

The crucial point in this presented approach is to generate (and provide structures for the design of) a set of rules – a rule base – which comprehensively describes the SWS designing and composing process reflecting the experience of the expert (the user) of the system: a strong formalization problem.

Within the Infrawebs framework the SWS Designer [3] accesses and evaluates different types of knowledge artifacts provided by a preprocessing stage (**Fig. 3**) [5]:

1 - A set of "Domain Ontologies" (representing the domain specific knowledge), WSDL objects (containing the static information about the web service to be designed) and BPEL (BPEL4WS) objects (representing the business logic of the WS). In this sense three different information channels and accordingly design stages (phases) are applied.

2 - Each stage can be characterized by a set of rules.

3 - Each set of rules can be transformed to numerical patterns – and

4 - finally aggregated to a computational feasible feature profile (vector), representing the whole designing (resp composing) process

GENERATION OF FEATURE PATTERNS USING THE CASE BASE OF THE CBR - SWS DESIGNER / COMPOSER

As described in [2], [3] the SWS designer developed in the Infrawebs project operates as "a provider's tool for semi-automatic design-time composing of semantic web services", whereas the SWS design task and the SWS design solution are stored as "Case Pairs" in a case base.

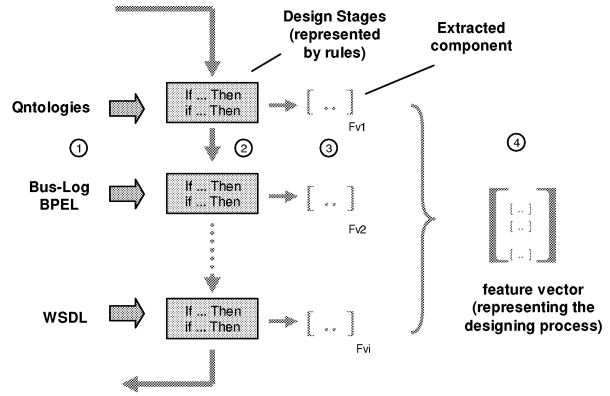


Fig. 3: Generating an overall feature profile representing different design stages

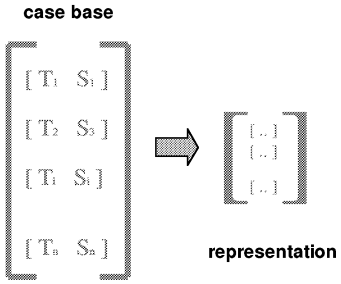


Fig. 4: Using the case pairs of the design resp composing case base for generation of characteristic feature profiles

In this section a further approach for generating an adequate and feasible feature profile (vector) is proposed and illustrated in **Fig.4**.

The task (goal) & solution pairs (linguistic and / or textual objects) are transformed to numerical patterns using adequate and normalized vocabularies. This vocabularies (and / or ontologies) consist of conceptually weights with respect to the given domain knowledge the SWS has to be designed for.

GENERATION OF NUMERICAL FEATURE PATTERNS BASED USER BEHAVIOR PATTERNS

The designer as well as the composer will be realized as programs and computational modules. Obviously this program consists of several and different program components like functions, sub functions, sub routines, classes, sub classes. Furthermore their exist dependencies and relations between this components as well as hierarchical weights and "impacts". This weights and "impacts" can be given numerically and represented accordingly by numerical values (metrics given by the expert realizing the system).

Using this weights and impacts in conjunction with time and frequency measures provides a computational bases for creating a system usage (behavior) pattern as well as a characteristically feature profile. This "numerical" profile can be used either itself solely or in combination with the rule base pattern within the basic DDS cycle in **Fig. 2**.

INFORMATION FLOW AND FUNCTIONS IN THE DDS (FUNCTION-MAP)

In the following the main functions and information flow of the DDS approach presented in this paper is described. According to the illustration in **Fig.5** the main idea is to create and compare design process describing rules, which are transformed to numerical profiles, clustered, implicitly optimized and finally used for recommendation and tuning of existing rule bases. The rules itself are stored, handled and maintained in two different kind of rule bases: a "local rule base" (with local indexing, LRB) installed within the environment of a individual platform partner and a "global rule base" (with global indexing, GRB), as central module accessible by all partners – a commonly used pool of rules.

1 - The designer user accesses predefined rules located in the global rule base (applied with local and global indexes), containing previously posted rules by other partners

2 - Furthermore he has the capability to post "new" rules, which he defines as useful.

3 - Posting a "new" rule triggers the generation of a global index, which is pushed to the local base

4 - Using his local rule base the user defines and determines according to his experience and assessment "rule weights" as well as "clause weights". The two indexes (global, local), the user defined rule weight and the two user defined clause weights are constructed as and merged to a 5 x m matrix (m – amount of rules in the local base) consisting of numerical values.

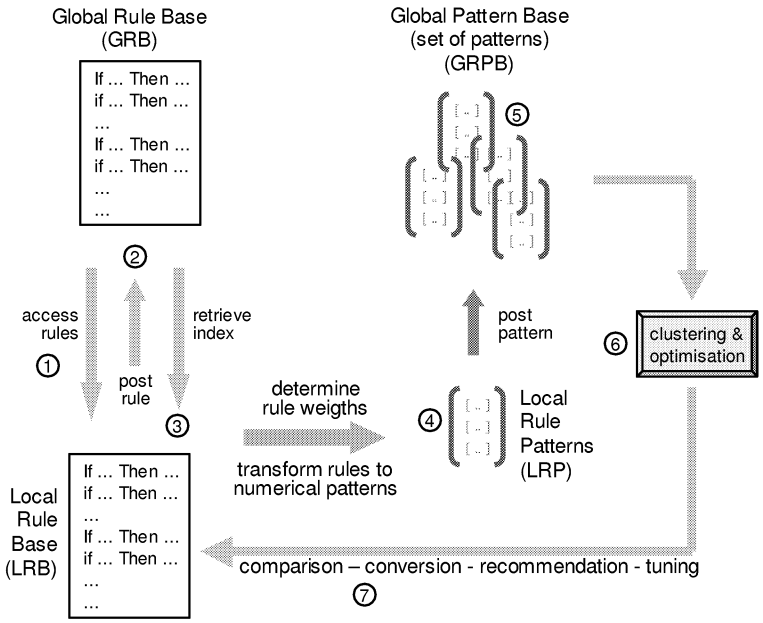


Fig. 5: Function-Map: Information Flow and Functions in the DDS cycle

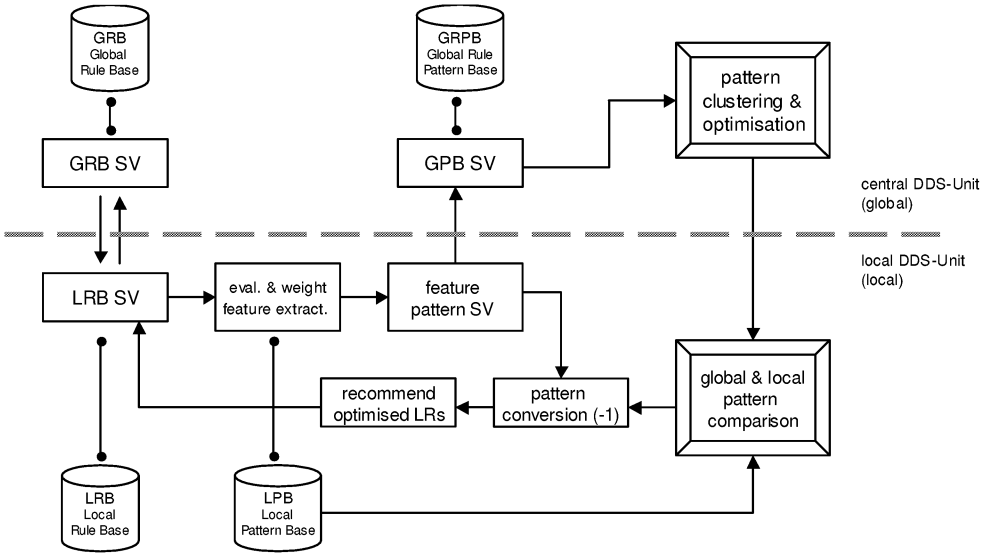


Fig. 6: Module-Map of the DDS

This matrix represents the rule base as a numerical pattern – the so called feature profile (Local Rule Patterns LRP).

5 – The local rule patterns (LRP) are posted to the central repository (Global Rule Pattern Base GRPB) consisting of the set of LRPs.

6 – The given (and dynamically growing) set of LRPs inf the GRPB is multi stage clustered with the result of an optimal LRP.

7- The optimal LRP is locally matched and compared with the local LRP. The extracted pattern is converted using the global and local indexes to a LRB (Local Rule Base), which is finally recommended to the existing LRB for tuning purposes of the LRB.

SW – MODULES AND ALGORITHMS OF THE DDS APPROACH (MODULE-MAP)

The module–map illustrated in **Fig. 6** gives an overview about the systems modules of the proposed DDS approach. The modules are installed on the one hand in the local environment of the platform partner and on the other hand in a central environment accessible by all platform partners.

Four repositories are installed: 1 local rule base (LRB), 1 local pattern base (LPB) in the local DDS unit and for the central DDS-Unit 1 global rule base and 1 global rule pattern base.

A local rule base server (GRB SV) accomplishes the handling of the LRB and fulfills communication tasks with global rule base server (GRB SV), which maintains the global rule base (GRB).

A feature extraction module is devoted to semi-automated weighting of rules and clauses as well as for evaluation purposes and connected to the local pattern base for exchanging purposes. The feature pattern server deals with the global pattern server (GPB SV) for handling the posted local rule patterns. Furthermore it triggers the main algorithm – the pattern clustering and (implicit) optimization process.

The resulting "optimal" pattern is posted to the local DDS-Unit to the comparison module, which extracts the main deviations between local and global rule pattern.

A pattern conversion module converts the numerical rule patterns to linguistic rule bases using the global as well as local indexes.

The resulting (optimised) rule base taken by the LRB SV for tuning and modification purposes of the local rule base.

ACKNOWLEDGEMENT

This work is carried out under EU Project INFRAWEBS - IST FP62003/IST/2.3.2.3 Research Project No. 511723.

REFERENCES

[1] Nern, H.-J., G. Agre, T. Atanasova, J. Saarela. 2004, "System Framework for Generating Open Development Platforms for Web-Service Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems - INFRAWEBS II. *WSEAS TRANS. on INFORMATION SCIENCE and APPLICATIONS*, 1, Vol. 1, 286-291.

[2] Atanasova, Tatiana; Nern, Joachim, 2005, "Distributed Decision Support in Multi-Agent Environments for Semantic Web Services", Euromedia 2005, Toulouse, April 11-13, 2005, IRT, Université Paul Sabatier, Toulouse, France

[3] Gennady Agre, Tatiana Atanasova, Joachim Nern, "Case-Based Semantic Web Service Designer and Composer", Euromedia 2005, Toulouse, April 1113, 2005, IRT, Université Paul Sabatier, Toulouse, France

[4] Atanasova, Tatiana; Gennady, Agre; "Specification & Architecture of SWU", Deliverable D5.1.1-v1 Infrawebs project, accessed 01.05

[5] Janne Saarela, Timo Westkaemper, Tero Hagstrom, "Specification SIR – Semantic Information Router", Deliverable D3.1.1, Infrawebs project, 2005, accessed 01.05

LATE PAPER

PARIS: A MPEG-7 GEO-REFERENCED PERSONAL PHOTOGRAPH DATABASE

Pei-Jeng Kuo, Terumasa Aoki and Hiroshi Yasuda
{peggykuo, aoki, yasuda}@mpeg.rcast.u-tokyo.ac.jp
The University of Tokyo, Research Center of Advanced Science and Technology,
4-6-1 Komaba, Meguro, Tokyo, Japan

KEYWORDS
Ontology, Digital Image Database, MPEG-7, Spatial-Temporal, image retrieval, metadata, Semantic Web, semi-automatic annotation

ABSTRACT

Geo-referenced digital photographs stamped with location information such as GPS data have become possible with the aid of GPS receiving devices and increasingly sophisticated digital cameras. In previous publications, we have proposed a MPEG-7 based annotation architecture, Dozen Dimensional Digital Content (DDDC), specially designed for facilitating the managing, browsing and retrieving of personal digital photograph collections. While our proposed DDDC annotation architecture supports spatial and temporal attributes, constructing a geo-referenced digital photograph library is still time-consuming and cumbersome. In this paper, we explain the concept and structure of our MPEG-7 based geo-referenced digital photograph library prototype system, named PARIS (Personal Archiving and Retrieving Image System). In addition to utilizing the temporal information recorded directly from image devices, our approach integrates spatial coordinates to realize novel retrieving methodology with geo-referenced photograph collections.

1. INTRODUCTION

People are taking an increasing amount of photographs while they migrate from traditional cameras to digital ones. The proliferation of image devices such as digital video camera, digital camera and mobile phone camera enable individuals to capture continuous shots for life events with almost no additional cost. Consequently, consumers began to accumulate hundreds of photographs when travel to new locations or during special events. In the past, most image retrieval researches focused on stock image collections or internet image repertoire where thousands of images are indexed and compared based on their signal features such as color histogram or color layout. Lately, growing research efforts are emphasizing on the semantic level image retrieval based on annotated metadata or surrounding information. We are now moving towards an era that any individual can easily archive thousands of photographs, thus fast and precise consumer level image retrieval implementations have become keenly essential. Unfortunately, satisfactory personal image retrieval system has yet appeared.

We argue that people make more photographs while they visit some new locations or during special events. And hence spatial and temporal attributes of personal digital photographs could contain the most relevant context information. Because of the very different characteristic compare with traditional commercial stock image collections, which normally used for previous image retrieval researches, organizing, archiving, and retrieving personal photograph collections requires also different approach. To tackle this problem, we utilize the MPEG-7 multimedia description standard, which describes not only the signal low-level audio-visual content features such as color, texture, motion, audio energy as well as high-level features of semantic objects, event, content management related information and so forth. In this paper, we focus on the part of spatial and temporal related semantic annotations.

Consumer photograph collections contain a burst structure. In other words, a group of photos may be taken for a semantic related event, but a few, if any, photos may be taken until another significant event started. [1, 7] In addition, the burst structure within collections of personal photographs tends to be recursive, as small bursts exist within big bursts. This recursive structure can be represented as a cluster tree, where photographs are stored only at the leaf nodes. [7]

While previous research [1, 7, 9] tried to enable semi-automatic event segmentation based on the recorded time tags made possible by most recent image devices, our research emphasis the importance of an integrated approach utilizing spatial information in addition to temporal information. We have designed metadata description architecture, DDDC (Dozen Dimensional Digital Content), extended from MPEG-7

multimedia description schema for annotating personal digital contents. And we also proposed the concept of "Spatial and Temporal Based Ontology", constructed based on the special pattern of personal photograph collections as we argue that time and location are two most important attributes in terms of personal photograph retrieval.

The integrated system constructed with the above proposals is named as PARIS, Personal Archiving and Retrieving Image System. In Sec 2, we discuss the importance of geographical information in terms of personal photograph collection retrieval. Sec 3 elaborates our proposed DDDC architecture and Sec 4 explains the concept of "Spatial and Temporal Based Ontology". The structure of our prototype system is illustrated in Sec 5 and Sec 6 describes the potential application based on our system and concludes this paper.

2. GEO-REFERENCED PHOTOGRAPH ARCHIVE

In [14], Rodden and Wood concluded that two of the most important features of an efficient, reliable and well-designed system for managing personal photographs are: automatically sorting photos in chronological order, and displaying a large number of thumbnails at once. While people are familiar with their own photographs, laborious and detailed keywords annotating are not specifically motivated for most people. Digital camera has become popular in the past few years, which means, a majority of people have only digital photograph collections accumulated within a limited time span of a few years. If we envision a continuous lifetime digital photograph archiving process, one might recalls, for example, his or her several trips to Paris a few years ago and one trip to Tokyo during Christmas season, but could not clearly memorize the exact year or dates.

While most Image retrieval researches focus on general image collections and a lot of prototype systems were developed with part of the COREL photo CDs , we would like to emphasize our research on consumer photograph collections where time and location count for important clues toward image contents. Fig. 1 shows an example of the raw data folder structure of our PARIS system. Those raw image data are continuously archived by one of our authors with a period of around three and half years and contain around 700 folders. Those folders were named with a brief line starting with the date when the groups of photographs were taken and following by a brief indication of the event, subject or location related to the content. The annotations convention of those folder names were done and followed constantly without previous knowledge of this



Figure 1. Original Folder Structure of PARIS System

study and hence are unbiased in terms of serving as a preliminary indication of the normal consumer naming tendency towards their digital photographs.

In [15], location has been argued as one of the strongest memory clues when people are recalling past events. In analyzing our prototype database raw data, we also find that most folder names contain words related to geographic information such as "Asakusa", "Yokohama", "Boston", NYC", "Fukoka", and "Paris", as well as temporal information such "France National Day", "Christmas", or " Winter". All of the folder names start with temporal information and 673 out of the 711 folders contain geographical related words, which is about 93% in percentage.

While our prototype system only contains images from a single person, we also conducted oral interviews with morn than 10 active digital image users to ensure an unbiased presumption towards common user naming convention. The interview results show that all of the interviewees store their digital photographs in different folders roughly named by location, event, and time period, for example, "A trip to Europe_June_2004", "Birthday Party with Nico_20040125", etc.

According to the above discussion, we see a strong association between the image context and its respective spatial and temporal clues for personal digital photograph collections. With the aid of recent GPS receiving devices, it is also possible to embed geographic information in GPS format onto digital image data. Although the recorded device GPS information does not guarantee the exact location of subjects appeared in the photographs as we had discussed in [4], utilize GPS information as a retrieving attribute greatly reduce the human effort compare to manual input. In Sec 4, we will further discuss the semi-automatic geo-information generation methodology conducted in PARIS system. Before that, we will briefly explain the Dozen Dimen-



Figure 2 Sample Image with free text annotation "2001_07_14_People and Eiffel Tower in Paris on the National Day of

sional Digital Content (DDDC) description architecture utilized in our annotation process in the following section.

```
<TextAnnotation id="Ann1">
    <FreeTextAnnotation xml:lang="en">
    2001_07_04_People and Eiffel Tower in Paris on the
    National Day of France
    </FreeTextAnnotation>
    <StructuredAnnotation>
        <Who>
            <Name xml:lang="en">People
            </Name>
        </Who>
        <What>
            <Name xml:lang="en">
                Eiffel Tower
            </Name>
        </What>
        <Where>
            <Name xml:lang="en">
                Paris; France
            </Name>
        </Where>
        <When>
            <Name xml:lang="en">
                2001-07-14T20:17+01:00; Na-
                tional Day of France
            </Name>
        </When>
    </StructuredAnnotation>
</TextAnnotation>
```

Figure 3 DDDC Annotation example of Fig. 2 with FreeTextAnnotation and StructuredAnnotation

3. DOZEN DIMENSIONAL DIGITAL CONTENT (DDDC)

Extended from the StructuredAnnotation Basic Tool of MPEG-7 Multimedia Description Schemes (MDS), we propose a semantic description tool of multimedia content. The proposed content description tool annotates multimedia data with twelve main attributes regarding its semantic representation. The twelve attributes include answers of who, what, when, where, why and how (5W1H) the digital content was produced as well as the respective direction, distance and duration (3D) information. We define digital multimedia contents including image, video and music embedded with the proposed semantic attributes as Dozen Dimensional Digital Content (DDDC). Fig. 2 is a sample image with free text annotation " 2001_07_14_People and Eiffel Tower in Paris on the National Day of France". And Fig. 3 shows part of the MPEG-7 based DDDC annotation of Fig. 2. Due to limited space, detailed explanation and example codes can be found in [4].

4. CONSTRUCT DOMAIN BASED ONTOLOGY BASED ON SPATIAL AND TEMPORAL ATTRIBUTES

The above DDDC architecture provide a structured methodology to annotate most significant, if not explicit, semantic answers of personal digital photograph collection contexts. However, some of the DDDC annotations such as the free text part of who, where and what attributes still require manual inputs. In [6], several difficulties have been pointed out in terms of the annotation process. First, different annotator might use different terms to annotate the same concept. Second, the users who do not have specific domain knowledge might not be able to input the right keywords or natural language query for semantic image retrieval. And third, the manual annotation of a large amount of personal digital photograph collections, if not impossible, is a laborious task.

We propose to build location specific "Domain Ontology" for popular tourist stops such as the city of Paris, Tokyo and New York based on their respective spatial and temporal attributes. Illustrations in Fig. 4 and Fig. 5 present example of spatial and temporal based hierarchical ontology for the city of Paris.

In building Spatial Ontology, we firstly separate Paris into several popular tourist districts such as "The Latin Quarter", "The Eiffel Tower Quarter", "Champs-Elysées" and "St-Germain des Prés". Under each district, we again separate it into sub-districts or point of interests such as "Café de Flore", "The Eiffel Tower" and "Café les Deux Magots". Each node of the sub layer inherits the properties of their upper layers; therefore, when we annotate a photograph with "Café de

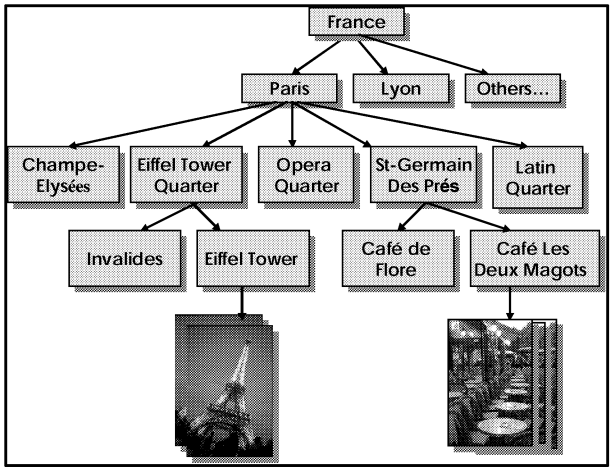


Figure 4 Concept of Proposed Spatial Ontology

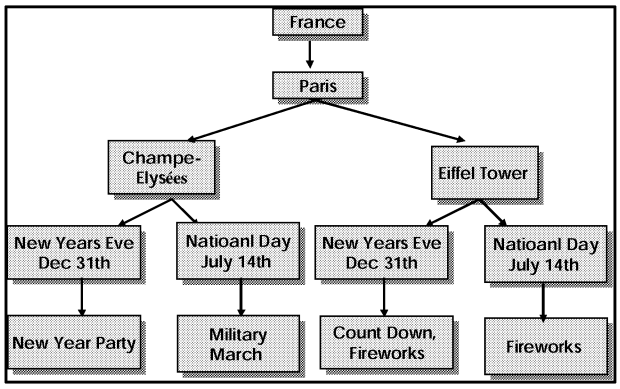


Figure 5 Concept of Proposed Spatial and Temporal Ontology

Flore" metadata, upper layer properties of "St-Germain des Prés" and "Paris", "France" would also be included.

The construction of Temporal Ontology requires more domain knowledge of the specific location. For example, the seasonal events periodically happen in the area, or special event occurs on specific date. As suggested in [12], there is no single correct class hierarchy for any given domain. And the ontology should not contain all the possible information about the domain but only specific enough for what you need in the application. We suggest building up the location specific "Spatial and Temporal Ontology" according to the photographer's personal interest and experience. In addition, we can also construct that with the aid of third party databases such as travel information portals or existing geographic metadata initiatives. In Fig. 5, we demonstrate event tags come from our Temporal Ontology such as "New Years Party", "Military March", "Fireworks" and "Count Down", which are associated with different image groups that were taken at the location of "Champs-Elysées" and "The Eiffel Tower" at special time such as "New Year's Eve" or "National Day".

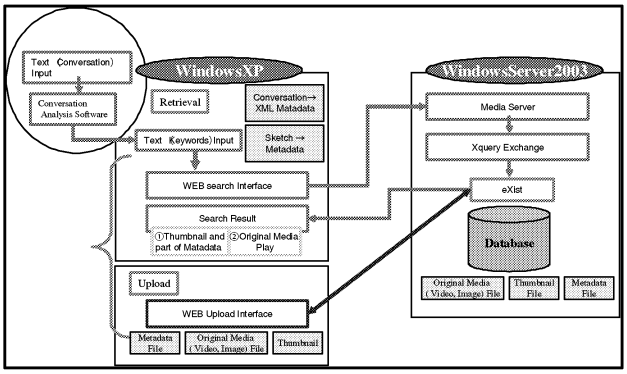


Figure 6 Testbed Structure used for constructing PARIS System

5. TESTBED STRUCTURE FOR PARIS SYSTEM

We are building up our personal photograph collection database based on the DDDC architecture described above, which annotates multimedia data with twelve main attributes regarding to its semantic representation. And we utilize third-party software in generating our proposed "Spatial and Temporal Based Ontology" for popular tourist cities such as Tokyo, New York and Paris. Different from previous researches, our "Spatial and Temporal Based Ontology" are designed based on the special pattern of personal photograph collections while we argue that time and location are two most important attributes in terms of personal photograph retrieval. The PARIS system are build in conjunction with a main project called AVR (Advanced Video Retrieval), funded by Japanese Government. Fig. 6 illustrates the structure of the AVR testbed. At current implementation, web interface is available for query sentence inputs. After the input, query sentence will be analysized by conversation analysis software called "Sen" and be converted into xQuery inputs. Unfortunately, the current input analysis is only available for Japanese query inputs; however, we will work on providing a more multilingual compatible solution in the future. The spatial and temporal associated MPEG-7 annotation files are stored in a XML database called eXist and query results and thumbnails will be returned to our web-based interface. While AVR project explores novel methodologies for multimedia content retrieval including video, audio and image; PARIS system targets on personal photograph collection retrieval and is built under the whole AVR testbed. Raw image files, thumbnails and XML based DDDC MPEG-7 annotations are stored in the database. The spatial attributes are firstly generated semi-automatically by third-party GPS program and parse to our DDDC annotation schema. We are currently investigating on collaborat-

ing with GPS receiving device and generate "true" geo-referenced photograph data. However, at this point, we utilize image data that were taken before the availability of GPS receiving device and semi-automatically annotate the location information with the aid of third-party software. Currently, the PARIS system contains around 80,000 photographs with a time span of around three years, and the number of images keeps increasing with a constant rate. Around 20,000 of the photographs were scanned from traditional films and hence to not contain exact temporal data. The other around 60,000 photographs was taken from contemporary digital cameras and all contain detailed temporal information. At current stage, around 10,000 of them are geo-referenced and hence related location annotations are generated based on our "Spatial and Temporal Ontology".

6. CONCLUSION AND FUTURE WORKS

PARIS, Personal Archiving and Retrieving Image System, is designed and implemented specially towards the emerging trend of continuous capture and storage for personal experience. While digital recording devices become pervasively available, consumers can easily accumulate thousands of digital photographs within a very short time period. Thus, image archiving, retrieving and managing methodology specially tailored for consumers are keenly needed. We have designed a MPEG-7 based DDDC multimedia description schema which annotates multimedia data with twelve main attributes regarding to its semantic representation. As we argue that location and time contains the most curial context information for personal photographs, we proposed a machine-understandable "Spatial and Temporal Based Ontology" representation for the above DDDC semantics description to enable semi-automatic annotation process particularly for enabling spatial and temporal based image retrieval for geo-referenced personal photographs.

Different from previous researches, the DDDC architecture complements the attributes specifically important for personal photograph collections. Besides, our "Spatial and Temporal Based Ontology" is a new concept designed based on the special pattern of personal photograph collections. The PARIS system is implemented is in conjunction with a three-year long project called AVR (Advanced Video Retrieval) funded by Japan Government and currently a text-based web retrieving interface has be realized. We anticipate a more sophisticated demonstration system to be presented by the end of year 2005.

As personal digital photograph collections have specific characteristics and are particularly spatial and temporal associated, we envision various novel browsing possibilities at semantic level can be developed based on the proposed DDDC architecture with the aid of our domain specific "Spatial and Temporal Based Ontology". While this paper explains the underly structure of PARIS system constructed based on our proposed methodology, we are also conducting experiments on our designed spatial and temporal image retrieval. We expect to publish the evaluation results on future publications.

7. REFERENCES

[1] J. C. Platt, et al., "PhotoTOC", Microfort TR, Feb 02.

[2] ISO/IEC 15938-5:2001, "Multimedia Content Description Interface – Part 5", version 1.

[3] P. J Kuo, T. Aoki and H Yasuda, "Semi-Automatic MPEG-7 Metadata Generation of Mobile Images with Spatial and Temporal Information in Content-Based Image Retrieval", Pro. SoftCOM 2003, Oct 03.

[4] P. J Kuo, T. Aoki and H Yasuda, "MPEG-7 Based Dozen Dimensional Digital Content Architecture for Semantic Image Retrieval Services", Proc. EEE-04, Mar 04.

[5] V. W. Soo et al., "Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-based Learning Techniques", Proc , JCDL'2003, May 03.

[6] A. T. Schreiber et al., "Ontology-Based Photo Annotation", IEEE Intelligent Systems, May 01.

[7] A. Graham et al., "Time as Essence for Photo Browsing Through Personal Digital Libraries",Proc.,JCDL'02, Jul 2002.

[8] E. Hyvönen, et al., "Ontology-Based Image Retrieval", Proc. of XML Finland 02.

[9] A. Stent, et al., "Using Event Segmentation to Improve Indexing of Consumer Photographs", Proc. ACM SIGIR'01.

[10] A. Jaimes et al., "Semi-Automatic, Data Driven Construction of Multimedia Ontologies", Proc.ICME 03.

[11] N. Noy et al., "Ontology Development 101", SMI TR, 01.

[12] J. Hunter, "Enhancing the Semantic Interoperability of Multimedia Through a Core Ontology", IEEE Transactions on Circuits and Systems for Video Technology,Jan 03.

[13] P. J Kuo, T. Aoki and H Yasuda, "Building Personal Digital Photograph Libraries: An Approach with Ontology-Based MPEG-7 Dozen Dimensional Digital Content", Proc. Computer Graphics International-2004, Jun 04.

[14] K. Rodden and K. Wood, "How do People Manage Their Digital Photographs?" ACM Conference on Human Factors in Computing Systems (ACM CHI 2003), Apr 2003.

[15] W. Wagenar, "My memory: Astudy of autobiographical memory over six year", Cognitive Psychology, Ch 18, 1986.

AUTHOR LISTING

AUTHOR LISTING

Abel B. **129**
Acquaviva A. **110**
Agre G. **216/221/226**
Akafuj T. **77**
Al Hammadi A. **42**
Alefragis P.S. **37**
Amoroso A. **62**
Aoki T. **255**
Atanasova T. **210/216/221/226/233/247**

Bahsoun J.P. **178**
Baldassarri M. **136**
Batatia H. **194**
Belhoul K. **42**
Bengherabi M. **169**
Benini L. **110**
Berhe G. **199**
Bourgeois J. **173**
Brunie L. **199**
Butsuen Y. **119**

Cacciaguerra S. **136**
Chaloub G. **164**
Charvillat V. **194**
Chbeir R. **164**

Dascalova H. **205/210**
Datcu D. **105**
Dayre P. **194**
De Cock J. **93**
de Haan G. **124**
De Wolf K. **93**
Dedu E. **190**
Dhondt Y. **93**
Di Pasquale M.P. **62**
Dziech A. **242/247**

Emmelkamp P.M.G. **129**

Farella E. **110**
Fitrianie S. **83**
Franzke M. **23**
Fujikawa K. **119**
Fülöp C. **236**
Furnell S.M. **67**

Grebenstein K. **146**
Grigoraş R. **183/194**
Grigorova V. **210**
Guy C.G. **31**

Haidar B. **178**
Harumoto K. **77**
Hasslinger G. **23**
Helinsk M. **5**
Henkemans O.B. **124**
Henriet A. **173**

Ikonen J. **47**
Ishikawa T. **119**

Joly P. **178**

Kajava J. **151**
Kandori K. **77**
Kassel S. **146**
Kobayashi M. **77**
Kosch H. **159**
Koskinen K. **47**
Kovacs L. **236**
Krijn M. **129**
Kuo P.-J. **255**
Kusuma I. **10**

Lacan J. **173**
Liang Z. **10**
Linck S. **190**

Masotti M. **62**
Mazurek C. **5**
Micsik A. **236**
Mirri S. **97/136**
Mostefaoui A. **169**

Nakaramura Y. **119**
Neerincx M.A. **124**
Nern J. **216/221/226/233/239/242/247**

Padiou G. **183**
Pérennou T. **173**
Phippen A.D. **67**
Pierson J.-M. **199**
Piramuthu S. **143**
Pleşca C. **183**
Polleres A. **230**

Qingping L. **10**
Queinnec P. **183**

Ricco B. **110**
Rothkrantz L.J.M. **83/105**

AUTHOR LISTING

Saad S. **164**
Saarela T. **239**
Salomoni P. **97/136**
Sato T. **119**
Schnitter S. **23**
Schuemie M.J. **129**
Schumann C.-A..... **146**
Scicluna J. **230**
Sehad A. **169**
Shimojo S. **77**
Spacil R. **47**
Spies F..... **190**
Sunahara H..... **119**

Tacheva J. **242**
Tittmann C. **146**
Triantafillou V.D. **37**
Trolin M..... **57**
Tsatsaronis K.E..... **37**

Van de Walle R..... **93**
van der Mast C.A.P.G..**124/129**
Varonen R. **151**
Voros N.S. **37**

Werla M. **5**
Westkaemper T. **239**

Yamamoto Y..... **77**
Yamazawa K. **119**
Yasuda H..... **255**
Yasukata M..... **77**
Yetongnon K..... **164**
Yokoya N. **119**

Zemerly M.J..... **42**
Zhang C..... **31**
Zincir I..... **67**