SCIENTIFIC PROGRAMME

KEYNOTE

WHERE DO WE WANT TO GO?
WOMEN AND IT IN THE 21st CENTURY

Caterina Rehm-Berbenni
CEO FUTUREtec GmbH
Hauptstrasse,188
D 51465 Bergisch Gladbach
Germany
E-mail: CRehm@futuretec-gmbh.de

KEYWORDS

ICT and gender, new methods of work, motherhood, profession

ABSTRACT

Industrialised countries are undergoing a process of transition from an 'industrial' to an 'information' society. One essential feature of the latter is, that everyone has access to information for political, economic, cultural and social development, so that there is a high level of information use among the general public. In the light of this, I would like to present a systematic approach to women's role in the information society and in modern professional life, highlighting the particularity of the feminine and maternal world from which we may discover new ways of managing civil, cultural and industrial-commercial life with a special emphasis on running new and small companies.

BACKGROUND

The way men and women use space in the home and at the workplace is different. Men and women also have different habits and these result from a complex mixture of culture, traditions, personalities, and in particular their different biological nature. This is also reflected in the different ways men and women earn their living. This difference has often been, and in some sectors is still viewed with scepticism. But I believe that the different way men and women live and feel offers 21st Century society an alternative way of analysing situations and doing things. Various cultures in the past have always felt such differences, but nowadays neuroscience confirms them more and more, and even amplifies them.

Therefore, entering the heart of the woman's universe, it is really worthwhile to concentrate on the fundamental meaning of motherhood, analysing three main characteristics of women's maternal world: vitality, flexibility and interacting through images. These characteristics are key aspects of beginning and running either a family or a small firm. They are also core elements of the information society we now work and live in.

Vitality

At the maternal level – Mothers have developed enormous vital energies at different levels:

- To cope with the challenging tasks of bringing up children and educating them
- To sense potential threats or conflict before they become manifest, which we also call 'sixth sense'
- To create and maintain a harmonious social environment keeping some space to realise their own ideas

This was always a guarantee for protecting human life, for creating a favourable and harmonious environment for bringing up children, and for meeting other people's needs as well as their own.

At the professional level – A woman has the capability:

- To communicate with others, absorbing and/or transmitting ideas and information
- To work hard and tenaciously
- To 'read between the lines', to feel existing or approaching conflicts, to 'smell' innovative products and markets
- To decrease tensions and mediate between conflicting parties
- To promote co-operative work

Flexibility

At the maternal level – The capability:

- To react very quickly to changing conditions in their environment -often superficially considered as a kind of moodiness- but in reality one of the most valuable talents for a mother
- To do several tasks at the same time
- To adapt to the persons she interact with: children, husband, relatives, neighbours

At the professional level – A woman has the capability:

- To react quickly to changing conditions in her professional environment (in the company and in the market), which means continuous review of

and, when necessary, adaptation of decisions and strategies
- To do different things and act at different levels simultaneously (multitasking)
- To network i.e. public relations and manage human resources

Interaction through images

At the maternal level – The capacity to think and communicate in images, which emerged from the necessity to make complex issues (traditions, ethical values, dangers) understandable to children and to people with different cultural backgrounds

At the professional level – The increasing complexity of commercial processes and the growing need for interdisciplinary skills (research, design, financing, technology transfer, production, marketing) requires the ability to simplify complex information without losing the essential features of it, in order to make it understandable to partners who are not being experts in the specific field.

The information society has an ever-growing need for this figurative way of explaining complex concepts 'to keep complexity manageable and make simplicity effective'.

All the particular capacities of women are reinforced by a high sense of responsibility and justice. All of this together, means that women's influence in enterprises, as well as in society introduces a tendency towards self-sustaining, durable development.

CONSEQUENCES

It is well known that one of the fundamental reasons why men and women are treated unequally in employment lies in the multiple roles women play in society, due to the fact that only women bear children.

For a woman, taking responsibility in management especially of an SME, often means accepting a double role and double work combined with the continuous desire to find a balanced harmony between family and profession, no matter how understanding and supportive her partner may be. However, it also means bringing a human perspective into a technology focused environment, in such a way that technology as such is not the focus, but rather a tool that supports all of us as we seek to improve the quality of our work and life. It also means utilising technology as a help towards long-term development and the wellbeing of future generations

One of the best opportunities for women, lie in transferring their ancestral skills to the typical processes of the information society e.g. internet, information filtering, knowledge management, software development and quality assurance and knowledge quality assessment, offers an essential contribution to the building of a sustainable community. In this way women, precisely as women, offer an important contribution for all forms of sustainable development and conflict resolution, as well as for the work towards social, economic and political justice.

It is therefore desirable that men and women are equally empowered economically and politically, in order to achieve the long-term sustainable and viable development of the information society (http://www.uia.org).

Information society and particularly the internet open up opportunities for new professions and services. These will benefit women enormously, because they will allow women to work at home and, with the support of their partners, fulfil the double role of being mother and professional.

The new and most interesting approaches to work in the information society are work from home, distance work, on-line employment, tele-working, tele-cooperation, e-commerce, e-teaching, e-learning, on-line services.

All this is in favour of a small, flexible and well networked enterprise becoming an ideal model for the 21st Century information technology enterprise. To achieve this, governments should give support developing the infrastructure for SMEs, promoting small scale production units, creating small offices and supporting small size businesses. In this way the state can play a central role in creating favourable conditions for women to make their special contribution to SME entrepreneurship.

Sometimes we feel helpless regarding the future of a globalising world and seem to be overwhelmed by the problems surrounding us, because we have the feeling that any contribution we could make would be too small to produce an effect.

The main concern today should not in the first place be global problems like resource depletion, over-population and pollution of the environment but rather the scale on which these problems are occurring. The larger the scale in which we view these problems, the less soluble they seems to be.

By contrast people feel empowered to have a real influence in smaller scale communities. When the information society enables these smaller scale communities' to work together, they will achieve global effectiveness. By focusing on the cell-like or family-based nature of society, we enhance the vital role that women can play.

CONCLUSION

Summarizing, we can say that there is a great need to create a model of society in which man and woman, 'united without confusion and distinct without separation', build up a solid civil project of democracy with more and better jobs for all.

And it may well be that the information society-based 21st Century will be the host of a transition from a defensive feminism to a mature feminine conscience, more and more valued and integrated into the family system, as well as in the socio-political, regional and global system.

REFERENCES

Moore, N., 1999. Partners in the information society. Library Association Record, Vol. 101, No. 12: 702-703.

Encyclopedia of World Problems and Human Potential, http://www.uia.org

Berbenni, G. "Hypertechnologies and Society in the XXI century", from an article for the periodical "DLR", Die Nationale Koordinationsstelle bei der Deutsohen Forschungsanstalt für Luft-und Raumfahrt, Projektträger des BMBF *f*ür In*f*ormationstechnik Abteilung EU-Programmbegleitung. Germany, May 1996

BIOGRAPHY

BERBENNI-REHM CATERINA, Born 03.08.52. PhD in Philology (1977) Bergamo (Italy).
International marketing at FERRERO (6 years); direction of a summer organisation with cultural exchange (Italy, England, France and Germany), directing 153 people (9 years). Since 1994, CEO of FUTUREtec. Expertise in *ICT:* E-business, e-learning, new methods of work, cooperative working; technology transfer; international marketing. Long years experience in interdisciplinary work.
International and national publications.

WEB TECHNOLOGY FOR BUSINESS APPLICATIONS

Distributed Decision Support Embedded in Semantic Web Service Development and Maintenance Environments for Intelligent e-Business Applications

Hans-Joachim Nern[1] and Janne Saarela[2]

[1]Aspasia Knowledge Systems,
D-40210 Duesseldorf, nern@aspasia-systems.de

[2] Profium SAE, FIN-02600 Espoo,
Janne.Saarela@profium.com

KEYWORDS

Decision Support, Distributed Decision Support, Closed Loop Approaches, Semantic Web, Web Services, Semantic Web Services, E-Business,

ABSTRACT

The paper introduces an approach for embedding distributed decision support mechanisms in Semantic Web Service development environments. To cover several non-functional aspects, like interoperability, adaptability, Service composability and scalability a set of AI-methodologies are applied. Interoperability is achieved by exchanging semantically enriched Services among platform partners within P2P networks. The machine processability of ontology based structures in conjunction with "closed loop" approaches in the Service operation, maintenance and feedback cycle ensures implicitly a high degree of stable adaptability and scalability. A Quality of Service Broker as part of the distributed decision support unit evaluates experience feedback patterns archived in distributed rule and case bases.

INTRODUCTION

The European research and RTD area related to "Software Architectures & Components" resp. the area of "Engineering of Service Functionality" is characterized by several European projects running or being completed since 1999/2000 [1]. As described and demonstrated in [2], [3] the formerly formulated vision of the Semantic Web (SW) is becoming reality, especially considering the aspect of merging Semantic Web technology and Web Service approaches [4], [5], [6], [7].

Apart from the existing European Research and Development activities also in the area of B2B Integration and Workflow Systems this new techniques based on the Semantic Web vision are comprehensively treated and intensified comprehensively.

The accessibility of Services is, however, only starting to take advantage of the latest industry standards such as XML, SOAP, WSDL, UDDI. The last-mentioned three technologies are often associated with such "Web Services" concept as Service Oriented Architecture (SOA), where Services implemented by various technologies should be able to interface with other systems using open standards independently of programming languages and architectures.

In 2002 the "Web Services Architecture Working Group" of the W3C [8] specified a set of requirements for the Web Services stack (W3C Working Draft Nov. 2002), which later (in 2003-08) has been extended to a description identifying the functional components and layers of a SOA architecture (Wire, Description, and Discovery; three fundamental components: Service Provider, Service Requester, Service Registry). Likewise OASIS members have formed recently (2003-09) a new WS-CAF Technical Committee [9] to "define a generic and open framework for applications that contain multiple Services used in combination (composite applications)."

The B2B area is especially influenced by the new streams and action lines, because the B2B integration process can be generally identified as an application-to-application (Service to Service) process, which makes, the Web Services approach very applicable to this field. As a result the global players are enhancing and modifying their B2B Integration Systems (e.g. Oracle9iAS Integration, MS Biztalk, IBM CrossWorlds, BEAs WebLogic) toward the involvement of Web Services techniques and methodologies.

Furthermore workflow systems provided by such global vendors like DominoWorkflow (Lotus/IBM), MQSeries (IBM), Changenine (HP), I-Flow (Fujitsu) or SAP R/3 Workflow are also influenced by this new Service oriented stream and a continuously adaptation of SOA approaches can be observed.

The representatives of the Java oriented world (SUN, JSSL) released in 2003 a framework for Web Service development - the Web Services Reference Architecture (WSRA) [10], in which the main Web Service related principles and structures are given (guidelines for infrastructure, design, interoperability, testing recommendations, reference implementations of core Services).

In 2003, the IBM / Microsoft Corporation represented world reacts with a comprehensive specification of the "Architec-

ture and Composition of Web Services" providing a broad overview on the SOA [11], which comprises Web Service specifications and functions (transport, message, WS-addressing and description, security, transaction). It emphasises on the aspect of "composability" and introduces the specification of the BPEL4WS (Business Process Execution Language for Web Services). BPEL4WS provides three constructs supporting such composition aspects as "structure, information and behaviour". In this sense BPEL4WS represents the more "dynamical" feature supplementing the rather "static" Web Services stack characterised by WSDL, XML/HTTP/SOAP, UDDI.

The SONIC ESB (Sonic Enterprise Service Bus) [12] should be mentioned as an example of fully market-oriented approaches related to the organisational (platform oriented) management and distributed use of Services. It supports the WSDL, UDDI, SOAP, HTTP, Xpath, XSLT, XQuery etc. and combines XML, enterprise communication services, and a Service oriented architecture based on "static" Web Service standards.

A state-of-the-art analysis shows the lack of systems for effective collaboration in reconfigurable networks as well as adaptable platforms for the design, deployment and maintenance of Web Services. The present software tools and systems are mostly incoherent and static, which is due to proprietary aspects (see Sonic ESB) or features preventing an effective adjustment and balancing of information pools. This is also a reason for missing reconfigurability, adaptability and self-organising features in existing development and networking platforms.

The market of static Web Service software is mainly characterized by existing environments using open exchange and communication standards, like WSDL, UDDI, XML, RDF [13]. The coupling and communication between proprietary business solutions are still retarded and impeded due to missing adjusted and standardized formalization (resp. normalization) features for domain specific knowledge pools.

Due to their machine - processability characteristics ontologies provide some kind of normalization of knowledge domains of the cooperating entities and platform (and/or networking/workflow) partners - a basic precondition for successful communication and collaboration in team-processed design, production and maintenance of Web Services enabled applications.

It should be also stated that there is a lack of semantic based systems using distributed decision supported techniques – tools and systems, which enable based on feedback evaluation an “easy to use” Web Service designing, composing and maintaining process.

The distributed decision support features introduced in this paper are integrated in one of the main system modules applied to the Web Service modelling framework treated in the EC IST 6th framework project INFRAWEBS [14], [15], [16] (System Framework for Generating Open (Adaptable) Development Platforms for Web-Service Enabled Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems).

WEB SERVICE DEVELOPMENT AND MAINTENANCE FRAMEWORK

The INFRAWEBS system design is to provide an intelligent framework**,** which enables software and Service providers to generate and establish open, extensible and reconfigurable development platforms for Web-Service applications. Established in such a way the open platforms consist of coupled and linked INFRAWEBS units, whereby each unit provides tools and adaptable system components to analyse, design, conjointly compose, and maintain Services (SW Services) within their complete life cycle.

The innovative feature of the INFRAWEBS approach is the integration and "closed loop" characteristic, which enables a distributed decision making process coupled to distributed Web Services registries.

The application of this "*closed loop*" approach within the Service operation and feedback cycle (QoS) ensures implicitly system stability and a high "*quality of Service*" handling, deployment and maintenance. Service *composability* is guaranteed by applying open standards like XML, RDF(S), BPEL4WS, WSFL, WSIL, DAML(s), OWL etc. Distributed decision making processes based on archiving experience feedback patterns in distributed rule and case bases (CBR) ensure the generation of *customisable modular* development environments

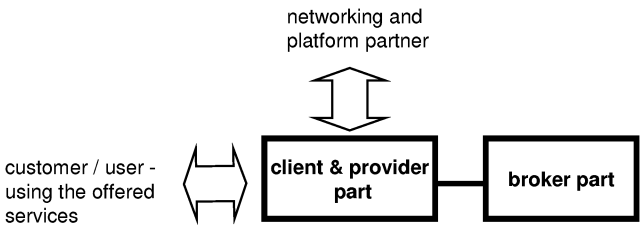


Fig. 1. The reduced scheme of the INFRAWEBS consisting of a broker and client & provider part

As illustrated in **Fig. 1** a knowledge broker is realised as a knowledge management unit (organisational memory). It is designated to endow the specific entity (business enterprise, governmental institution etc.) with comprehensive facilities to maintain and process the entity specific knowledge (knowledge artefacts) with respect to the SW Service design. Via this broker the user (developer) accesses the entity specific knowledge and extracts decision supported (via system recommendations) the SW Service relevant knowledge artefacts (see layer 1 in **Fig. 2**). These artefacts are transferred to the SW service unit, where the individual developer has capabilities and tools (designer, composer, executor) to “build” (design and compose) the knowledge artefacts to running SW Services.

As a base unit it enables the generation of open development platforms for SW Services. It interacts in two directions: to platform partner side and to the customer side.

The client & provider component is designed as a semantic and ontology based module handling and maintaining Semantic Web Services – the Semantic Web Service Unit as a workflow and knowledge sharing component - as a type of SW Service based collaboration platform and interoperable middleware. It acts in two directions: it provides Web Services to the platform and collaboration partner (layer 2) and secondly to the customer of the entity (layer 3). The main *B2C / B2B process* is conducted via this channel.

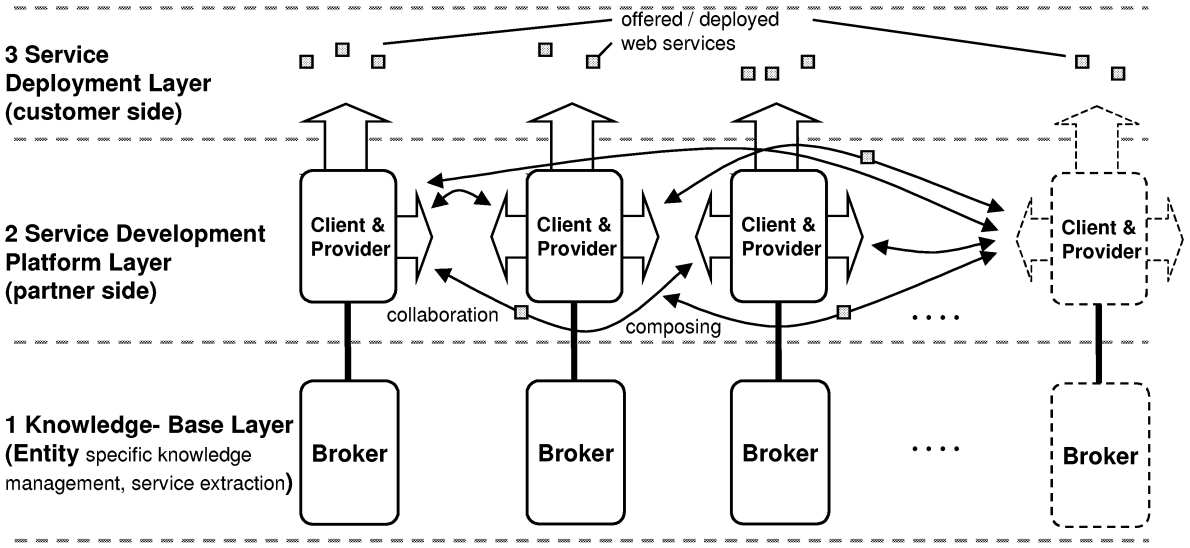


Fig. 2. The coupling of the INFRAWEBS units enables the establishing of reconfigurable self-organizing development platforms for the design, deployment and maintenance of Web Service enabled applications.

The networking and platform partners provide SW Services, which are incorporated in their own Services product range (composition). On the one hand, it is a conceptional expansion of "own" (entity specific) Services, on the other hand, the creation of new products and Services is facilitated to develop value added products and Services. This reflects an innovative type of semantic based platform oriented networking (semantic based interoperable middleware). The platform partners are endowed with facilities to optimise their partnership relations; furthermore, they are enabled to build faster and more effective partnerships with "new" business partners.

DECISION MAKING FEATURES IN DISTRIBUTED ENVIRONMENTS

In general, the concept of a UDDI is a centralised approach – by storing the web Services description in a central space. The existing trend to establish enterprise wide UDDI registries (using Microsoft Active Directory, Novell eDirectory) implies the disadvantage to be limited to central public registries (private registries) - and difficulties occur in Service discovering [6].

In opposition to this trend the present system approach applies distributed and replicable registries (**Fig. 3**) in connection to the distributed decision support modules. This is achieved by connecting private (locally in each unit) registries within the network.

The distributed repository supports decentralised architectures allowing multiple interconnected repositories within the generated and established platform. Furthermore, replication between the repositories is considered to enable accessibility and availability. The functionality of this repository module comprises the storage, retrieval and lookup of WSML (web service markup language) descriptions. It is endowed with administration, management and configuration capabilities.

The distributed registry supports decentralised architectures operating in p2p environment, allowing multiple interconnected repositories considering replication facilities between the local repositories. This enables a higher degree of accessibility and availability.

The local registry of the domain-specific SW Services makes them accessible to external SW clients. The registry meets the requirements of the provider & client architecture for both dynamic and static WS usage scenarios. The registry will be automatically populated by instances of SW Services composed and executed by the local provider & client modules.

The used kernel modules are mostly adopted out of the SWWS, OntoWeb projects [1].

DISTRIBUTED DECISION SUPPORT UNIT WITH EMBEDDED SERVICE EXECUTOR AND QoS-BROKER

The issues of collaborative decision making (Distributed Decision Support), Quality of Services monitoring and Service execution is considered within the INFRAWEBS approach in specific modules called Distributed Decision System and SW Service Executor (DDS; SWS-E). The SWS-E provides a run-time environment for running SW Services and monitoring the Service execution process considering QoS metrics. Extracted QoS specific data patterns are fed back to the designing and composing modules and guarantees in this way a closed loop optimisation cycle (**Fig. 4**).

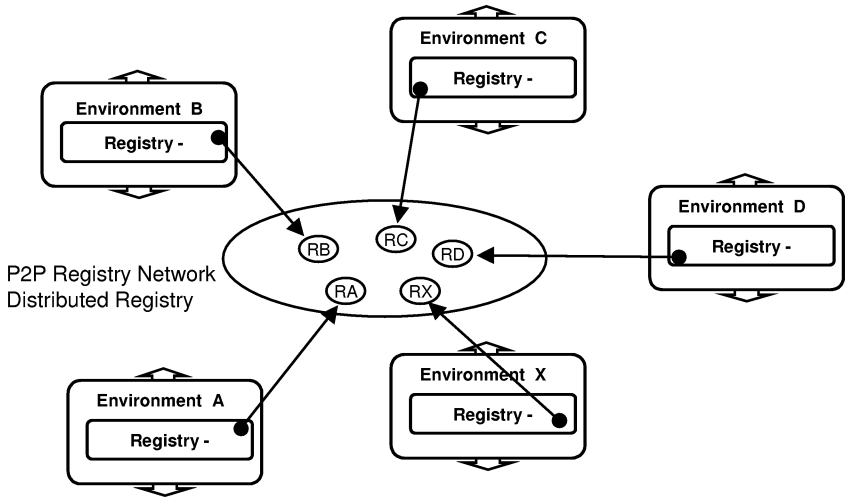


Fig. 3. Approach of a distributed SWS registry and repository with replication facilities

The behaviour patterns (profiles) of the Service designing and composing process are formalised and archived as design / composing rules and stored in local as well as global rule bases (Distributed Rule Bases for Distributed Decision Support). An agent based replication and adjustment procedure ensures a "global" consistency of the rule bases implemented in the different platform environments of the platform partners.
The function of the DDS are:

- Monitoring the SW Service execution process oriented on QoS metrics
- Extract QoS related statistical data patterns out of the monitoring process
- Feedback the QoS patterns to the designing and composition process
- Extraction of design / composing patterns of each coupled platform partner
- Formalisation, transformation and archiving of adequate rules
- Analysing the distributed local rule bases for adjustment of the global rule base (global rule base replication)

It is responsible for the distributed decision support structures. It is connected to the distributed Web Service repository (see previous section) via a Service Net agent (p2p net Agent) to the local rule bases of the platform partner modules. It summarises and investigates the decision relevant information tags for forming and formalising general influences from the platform partners.

The executor - realised as a run-time module - is responsible for executing and managing the constructed Service compositions. Through evaluation and interpretation of performance specific metrics QoS specific patterns are extracted and brokered.
The QoS Broker operates via the central unit (global rule bases) and influences the Service designing and composing process (feedback) in the sense of a closed loop optimisation procedure. The broker evaluates the running SW Services according defined major requirements (availability, accessibility, integrity, performance, reliability, regulatory, etc).

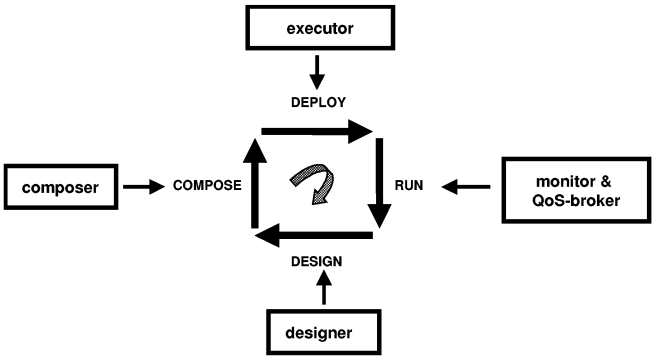


Fig. 4. The QoS-Broker within a closed loop process

CONCLUSION

Experience feedback patterns in distributed rule and case bases are archived, evaluated and utilized within distributed decision making processes. For user friendly handling of the collaboration and composing structures, several supporting tools are designed to assist the developer in execution of his problem, such as Web Service designers, composers and executors. A multi-agent-system provides user interface agents as well as agents to perform Web Service discovery and mediation.
The application and use of development platforms (like this presented one) will decrease and minimise the productivity gap (ITEA) [17] between the existing amount of software in Service products and the individual software productivity. This will be achieved by applying decision supported Service development, conjointly design, composing, and incorporating of Services and closed loop handling and treatment of the full life cycle of SW Services (Quality of Service Brokering, experience feedback, reusability, evaluation and optimisation).
Distributed decision support systems and tools open complex and impervious structures to the user and the customer. This aspect is especially related to the business sector, where the amount of materials, objects, coherences and relations is in permanent growth. This permanent growth provides - implicitly - a massive loss in straightforwardness – intelligent platforms and tools like the presented one could help to bring back what has been lost. Being directly connected to business partners and having direct access to their Services and products, advances the stimulation of innovation "through positive effects on the knowledge value chain".

REFERENCES

1. Lacroix, M. Bus, J. :Technologies & Engineering for Software, Systems & Services, IST IV.3, http://www.cordis.lu/ist/ka4/tesss/projects.htm.

2. Fensel, D.,: The Semantic Web and its Languages, IEEE Intelligent Systems, Trends and Controversies, pp.1, November/December, 2000.

3. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web, Scientific American, 2001

4. Bussler, C., Maedche, A., Fensel, D.: Web Services: Quo Vadis? In: IEEE Intelligent Systems, Trends and Controversies: Web Services: Been There, Done That? Vol. 18, No. 1, January/February 2003

5. Bussler, C., Fensel, D., Maedche, A.: A Conceptual Architecture for Semantic Web Enabled Web Services. In: SIGMOD Record, Vol. 31, No. 4, December 2002

6. Fensel, D., Bussler, C.: The Web Service Modeling Framework WSMF. In: Electronic Commerce Research and Applications, Vol. 1, Issue 2, Elsevier Science B.V., Summer 2002

7. Fensel, D. , Bussler, C.: Digital Enterprise Research Institute, http://deri.ie/projects.html , http://knowledgeweb.semanticweb.org/

8. W3C: Web Services Architecture Working Group, www.w3.org/2002/ws/arch/

9. OASIS: WS-CAF Technical Committee, xml.coverpages.org/WS-CAF-Announce.html

10. JSSL- Java Smart Services Laboratory: Web Service Reference Architecture, www.jssl.org/jssl/

11. Microsoft: Architecture and Composition of Web Services, http://msdn.microsoft.com/library/en-us/dnwebsrv/html/wsOverView.asp

12. SONIC: ESB (Sonic Enterprise Service Bus) http://www.sonicsoftware.com/products/sonic_esb/index.ssp

13. Lassila, O., Swick, R.: Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation, World Wide Web Consortium, Cambridge (MA), February 1999; available on-line as http://www.w3.org/TR/REC-rdf-syntax/

14. Nern, Hans-Joachim, Atanasova, T., Agre, G., Saarela, J.: System Framework for Generating Open (Adaptable) Development Platforms for Web-Service Enabled Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems – INFRAWEBS II, submitted paper to 8th WSEAS International Conference on COMPUTERS (part of the 8th CSCC Multiconference) , WSEAS, Vouliagmeni, Athens, Greece, July 12-15, 2004

15. Atanasova, T., Nern, Hans-Joachim, Agre, G., Saarela, J.: 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

16. Nern, H.-J.: Final Report KNIXMAS: Knowledge Shared XPS Based Research Network Using Multi Agent Systems", EU-FP4, Esprit 977113

17. ITEA - Information Technology for European Advancement, http://www.itea-office.org/

DESIGN OF A WEB SERVICE ARCHITECTURE TO MODEL BUSINESS LOGIC AS A DECISION SERVICE

Tim De Troch, Christophe Mues and Jan Vanthienen
Katholieke Universiteit Leuven, Department of Applied Economic Sciences,
Naamsestraat 69, 3000 Leuven, Belgium
Email: Tim.De.Troch@arinso.com, {Christophe.Mues; Jan.Vanthienen}@econ.kuleuven.ac.be

ABSTRACT

In this paper, a decision table consultation web service is discussed, which allows business users and managers to separately manage and update business rules and policies, while enforcing their consistent use throughout various types of application settings. Thus, changes in the underlying business logic will have immediate consequences in all client applications using the web service. Two possible types of consultation services are implemented: either all available information is immediately supplied to the web service and the latter will return the action(s) resulting from this information, or the consultation takes on the form of a question/answer sequence. This allows for a flexible architecture, in which inferencing logic and user/database interfacing are very loosely coupled and which enables to model and maintain the business logic by end users domain experts.

Keywords: decision support, web services, business rules

INTRODUCTION

There is a high demand for solutions to separate the business logic of a system from the execution of that logic in a number of applications. A business user should be able to easily define and modify the business rules without having to update multiple applications. The Web Service, presented in this paper, illustrates how this can be organized.

This paper will present the use of a web service that allows to make a clean separation between the business logic and other parts of an application. Because web services can be called by applications written in various languages, this service allows a consistent use of the business rules throughout applications and enables to model and maintain the business logic by end users domain experts.

THE CONCEPT: BUSINESS LOGIC AS A SERVICE

There is a high demand for solutions to separate the business logic of a system from the execution of that logic in a number of applications. A business user should be able to easily define and modify the business rules without having to update multiple applications. The Web Service, presented in this paper, illustrates how this can be organized.

As an example, consider a bookstore that wants to give a certain discount, depending on some conditions like order quantity, location, whether the book ordered is a paperback, and whether the customer is a school. The decision on this discount needs to be taken by the business expert and not by someone from the IT department. If this business user would like to change the rules, he should be able to do so without having to bother the IT department.

There are two possible kinds of consultation of this business logic. Either all available information is given to the web service and it will return the action(s) that result from this information, or the web service asks relevant questions only and depending on the answer(s), it returns the resulting action(s) or it proceeds with asking new questions until a resulting action was found. The web service runs through the business rules the business expert entered and asks questions about quantity, location, etc. in order to find out which discount rate should be offered.

Applications where this web service can be used are numerous. The service allows a strict separation between the business rules and the application, which may be expressed in any language with Web Services support. Since also most server side scripting languages have SOAP support (e.g. PHP and ASP.NET), even the building of dynamic web pages is possible. An online bookstore can for instance also develop some business rules for the costs and delivery dates for every possible option of delivery (standard mail, air mail, courier service). These business rules can be used for internal use by the ERP application, used by an employee to enter orders coming in by phone and on the web site to allow customers to make a choice between the different delivery schemes.

Web Services

The decision service is offered to a wide variety of development tools by using Web Services. Web Services are services that offer some functionality, in this case the consultation of business rules, using standard internet techniques (see Peterson & Davie, 2000). The most well known protocol of Web Services is SOAP (Simple Object

Access Protocol). SOAP is a way of calling remote functions and answering remote function calls using simple messages in text format. These messages use XML (eXtensible Markup Language) in order to pass the name of the function to be called, the parameters and their values, and the response to the method call (see Snell, Tidwell & Kulchenko, 2002).

Because messages are only passed in text format, a web service can use http (HyperText Transfer Protocol) to communicate. This offers the advantage that a Web Service that is on a normally configured and secured web server does not need extra firewall configuration. The Web Service can use the same port as a web server uses to pass simple html pages.

Consulting the decision service

There are two possible kinds of consultation of this business logic. Either all available information is given to the web service and it will return the action(s) that result from this information, or the web service asks relevant questions only and depending on the answer(s) it returns the resulting action(s) or it proceeds with asking new questions until a resulting action was found. Questions can be answered by asking the user or by sourcing the information from databases.

This Web Service runs through the decision rules and returns a relevant question when it does not have enough information to make a decision amongst the possible actions. This means that if it turns out that a condition is not relevant for decision making in this particular case, based on the information acquired from previous questions, this condition will not be sent to the client.

IMPLEMENTING THE DECISION WEB SERVICE

Interface

In order to offer the decision functionality, the Web Service will offer two functions. One returns the first question, the other processes the client's answer and asks another question if necessary, or returns the resulting action. The interface looks like this in Borland Delphi syntax (for information about Web Services and Delphi, see Gunzer, 2002, and Darakhvelidze & Markov, 2002):

```
IDecisiontableWS = interface(IInvokable)
  ['{1842C7AA-A656-4489-B14B-2694308C9D3B}']
  function GiveFirstCondition(PFilename:
       string): TConsultationMessage; stdcall;
  function AnswerQuestion(PFileName: string;
    PMessage: TConsultationMessage; PStatenr:
    integer): TConsultationMessage; stdcall;
end;
```

GiveFirstCondition

The first function examines the rule set and looks for a value that needs to be filled in, in order to obtain the result of the decision. It just returns the first condition with its different states. The only parameter it needs, is the name of the file that contains the rule set. The result is a Consultation Message. This message contains all the information the client needs for asking questions and returning actions and all information the server needs to know which answers have been supplied before. Such a consultation message looks like this (in Borland Delphi):

```
type
  TConsultationMessage = class(TRemotable)
  private
    FQCondition: string;
    FQStates: TStringDynArray;
    FQRelevant: TIntegerDynArray;
    FQCondNr: integer;
    FQEnd: boolean;
    FQAction: TstringDynArray;
    FQStateNrs: TIntegerDynArray;
  published
    property QCondition: string
     read FQCondition write FQCondition;
    property QStates: TStringDynArray
     read FQStates write FQStates;
    property QRelevant: TIntegerDynArray
     read FQRelevant write FQRelevant;
    property QCondNr: integer
     read FQCondNr write FQCondNr;
    property QEnd: Boolean
     read FQEnd write FQEnd;
    property QAction: TStringDynArray
     read FQAction write FQAction;
    property QStateNrs: TIntegerDynArray
     read FQStateNrs write FQStateNrs;
  end;
```

The meaning of the properties:

QRelevant: This property is used to remember which columns of the table are still possible. The possible states are those states that are not excluded yet based on the previous answers the user gave.

QCondition: This is a string with the condition name.

QStates: This is a dynamic array with the possible states of condition QCondition.

QAction: This is an array with resulting action(s). This will of course only be filled in when there are no more conditions that need to be answered.

QEnd: This is a boolean that tells whether the result has been found. If this is false, the client will need to process QCondition and QStates. If this is true, thee client will process QAction.

QCondNr: This value tells which condition was reached in the web service.

QStateNrs: This arrays maps an item in QStates to a state number. This is necessary because when concatenated conditions occur (a or b) the location in the array QStates does not map the number of the state. State 3 might e.g. be the second element in QStates, after 'a or b', but it is in

fact the third state. In our small example QStatenrs would be [1, 3].

AnswerQuestion

The second function, that will be called multiple times, first processes the answer on the previous question that was returned by the client and then looks for another value that is needed to make a decision. When the next unknown variable is found, a new question will be returned. If there are no more questions to call, the Web Service will return all results that match with the values returned by the client.

Building a client

Calling the web service from a client application is quite easy. To explain this, the development of a client will be explained step by step.

Importing the web service interface

The first thing that needs to be done is importing the service. Thanks to WSDL (Web Service Description Language), a Web Service can declare its own functionality. Most programming tools have a parser that understands WSDL and generates an interface to the web service as if it would be a local class.

The User Interface

You can build the entire GUI but instead of hard coding the questions that are behind this GUI, calls to the Web Service will be used to determine what should happen.

Connecting the GUI to the Web Service

The first user action calls the GiveFirstCondition function:

```
result :=
GetIDecisionTableWS.GiveFirstCondition(PFileName);
```

All subsequent actions send a request using AnswerQuestion:

```
  result :=
GetIDecisionTableWS.AnswerQuestion(PFileName, Msg,

Msg.QStateNrs[Form1.RadioGroup1.ItemIndex])
```

The results of these calls are always of type TConsultationMessage. If the attribute QEnd of the result is false, this means no action was found yet, so the next relevant condition and its states are shown.

When QEnd becomes true, a message with the resulting action(s) appears.

Modeling the business logic

The way the Web Service represents business rules is by means of decision tables. A decision table is a tabular representation used to describe and analyze decision situations, where the state of a number of conditions determines the execution of a set of actions. Not just any representation, however, but one in which all distinct situations are shown as columns in a table, such that every possible case is included in one and only one column (completeness and exclusivity).

The decision tables can be built manually, or using a software tool. The decision tables used in this paper are built using the Prologa (PROcedural LOGic Analyzer) tool (see Vanthienen & Dries, 1994) or the Prologa ActiveX component to build decision tables (see Figure 1).

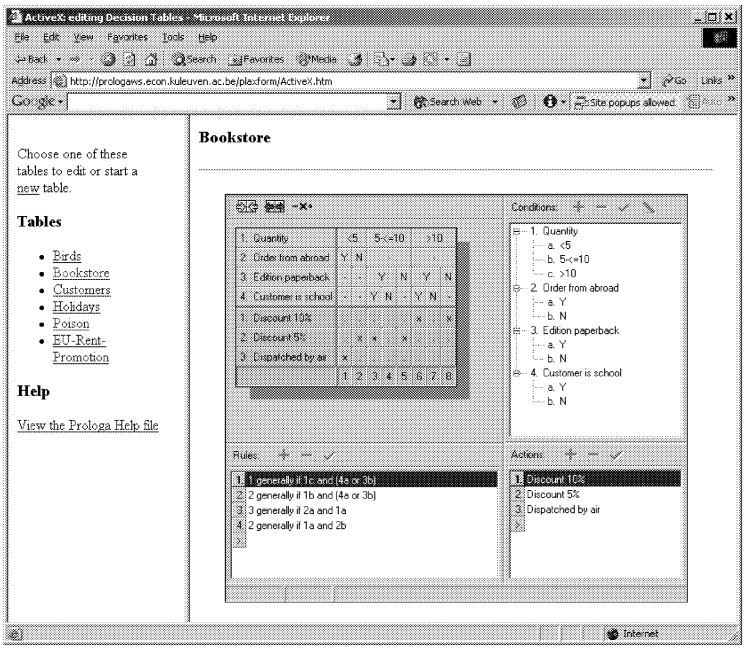


Figure 1: Decision table editing component

BOOKSTORE								
1. Quantity	<5		5-<=10			>10		
2. Order from abroad	Y	N	-			-		
3. Edition paperback	-	-	Y		N	Y		N
4. Customer is school	-	-	Y	N	-	Y	N	-
1. Discount 10%	.	.	.	.	.	x	.	x
2. Discount 5%	.	x	x	.	x	.	.	.
3. Dispatched by air	x	.	.	.	.	.	.	.
	1	2	3	4	5	6	7	8

Figure 2: Bookstore discount table

When modeling and representing the complex business logic to be dealt with in real business situations, we want to ensure the quality of the set of business rules from the start. Also, because maintaining the business rules, e.g. by end user domain experts, is not a trivial task and often introduces unnoticed anomalies, it is important that this quality is maintained and that the set of rules remains correct, complete, consistent and simple.

Decision tables are a powerful technique to represent and validate a set of related business rules in the form of tables (see Ross, 2003). Decision tables have proven a useful aid in modeling complex business situations of various sorts in a simple manner, easy to check by business experts for consistency, completeness and correctness. Experiences with decision tables and tools indicate that the representational and verification & validation advantages of decision tables fit very well with the world of business users (see Vanthienen, 2004).

AN INTERACTIVE EXAMPLE

In order to illustrate what the web service exactly does, an example table BOOKSTORE will be used. Suppose a bookstore wants to give a certain discount, depending on some conditions like order quantity, location, edition, customer type. The decision on this discount needs to be taken by the business expert and not by someone from the IT department. The business user might have defined the following decision table (Figure 2).

In order to apply the decision logic, the Web Service runs through the decision table and asks questions in order to find out which discount rate should be offered. Sometimes extra information is necessary. In an interactive application, the user might be asked to supply this information. A first question might be about the quantity of the order (see Figure 3).

Based on the answer given by the user, the Web Service continues to ask questions until a result is discovered. A possible outcome is shown in Figure 4.

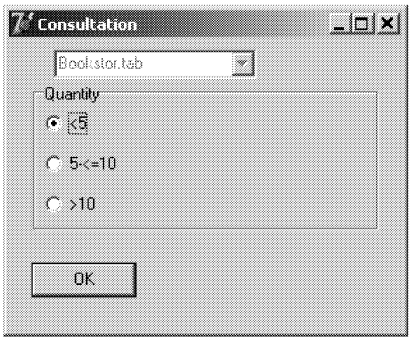


Figure 3: Question asked by the Web Service

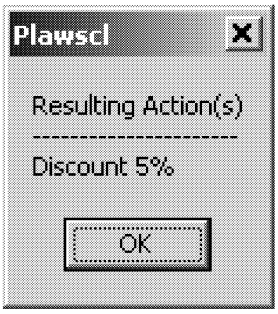


Figure 4: Resulting action

When we start the consultation of this table, we first call the function GetFirstCondition. The result of this call will be a Consultation Message that looks like this:

- QRelevant = (1,2,3,4,5,6,7,8), all columns are still relevant
- QCondition = 'Quantity'
- QStates = ('<5', '5-<=10', '>10')
- QAction = ''
- QEnd = false
- Qcondnr = 1
- QStateNrs = (1, 2, 3), because there are no concatenated states in QStates

Suppose the user answers Quantity '>10'. This reply is supplied to the function AnswerQuestion. After analyzing the decision columns, only columns 6, 7 and 8 remain valid. For all of these columns, the next condition (Order from abroad) appears to be irrelevant. The Web Service now

looks for the next condition. This appears to be a relevant condition, so this question will be asked to the user.

If the user chooses paperback (column 6 and 7 relevant) and indicates that the customer is a school, only column 6 will remain. This was also the last question, so the corresponding actions for columns 6 are retrieved. The function therefore returns the final action ('discount 10%') and sets the QEnd parameter to true.

If the business expert decides to change the discount policy, he can simply change the decision tables and all applications using the decision logic as a service can remain unchanged.

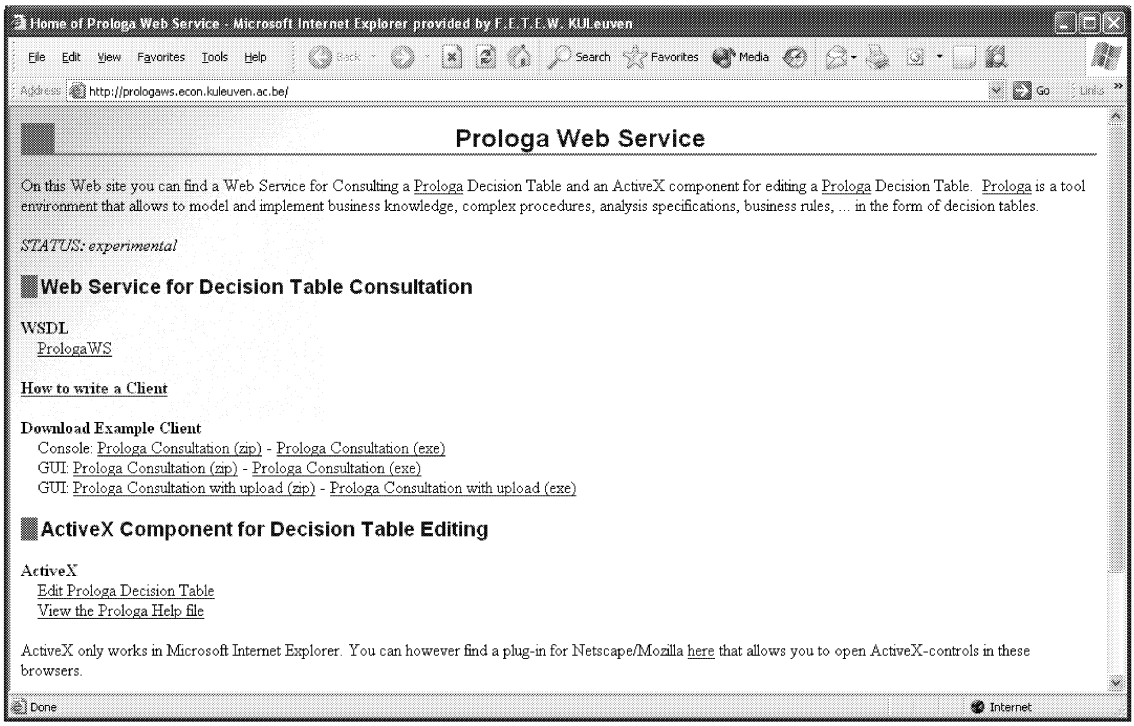


Figure 5: Web Service page

CONCLUSION

In this paper, a Web Service for separation of business rules logic and decision making was introduced. The major advantage is that it allows an easy and consistent application of business rules throughout the enterprise and that a business expert can easily update the business rules. Changes to the business rules will have immediate consequences in all applications using the web service.

This allows for a flexible architecture, in which inferencing logic and user/database interfacing are very loosely coupled and which enables to model and maintain the business logic by end users domain experts.

More information about the web service is available here: http://prologaws.econ.kuleuven.ac.be (Figure 5).

REFERENCES

Darakhvelidze P. & Markov E. (2002), *Web Services Development with Delphi.* A-List, LLC Wayne, 694 pp.

Gunzer H. (2002), *Introduction to Web Services,* White Paper, Borland.

Peterson L.L. & Davie B.S. (2000), *Computer Networks : A Systems Approach (second edition),* Morgan Kaufman Publishers, San Francisco, 748 pp.

Ross R. (2003), *Principles of the Business Rule Approach*, Addison-Wesley, 372 pp.

Snell J., Tidwell D. & Kulchenko P. (2002), *Programming Web Services with SOAP.* O'Reilly & Associates, Inc Sebastopol, 244 pp.

Vanthienen J., Quality by Design: Using Decision Tables in Business Rules, *Business Rules Journal*, Vol. 5, Issue 2 (Feb. 2004), 7 pp.

Vanthienen J. & Dries E., Illustration of a Decision Table Tool for Specifying and Implementing Knowledge Based Systems, *International Journal on Artificial Intelligence Tools*, 3(2), 267-288.

E-LEARNING STANDARDS

STANDARD AND ONTOLOGIES
PERSPECTIVES IN BUILDING AND PLANNING LEARNING FUTURE

Giorgio Valle
Raffaella Folgieri
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39
20135 Milano - Italy
Email: giorgio.valle@unimi.it
Email: folgieri@dico.unimi.it

KEYWORDS
E-learning, SCORM, Semantic Web, Ontologies, WebCen

ABSTRACT

Recent developments in ICT offer more opportunties to improve and widen communications with an obvious impact on E-learning methods and the distribution of educational content. E-learning represents a new concept of education and leads to a *common standard* to implement interoperability and to share resources. Teachers and students do not have to learn the use of different platforms, thus saving time, money and hard work. Moreover, it is more complex to manage different platforms than a single one. Finally, the adoption of such a standard allows economy of scale by allowing teachers to share contents created with different authoring systems. ADL Scorm is the more credited standard allowing the teacher to interact with other material from an existing course.
Along this line of integrating services, it emerges the innovative role of the *Semantic Web,* which aims to make web-based information services universally understandable and reusable. E-learning *standards* and the *semantic web* are strictly related, because they both aim to share and reuse information through the web. Moreover, since ontologies are a significative component of the Semantic Web, both of them will influence the development of new generation e-learning systems.
At the University of Milano are experimenting both concepts while accomplishing the new version of WEBCEN, www.webcen.unimi.it, an in-house educational environment that manages the four Undergraduate Curricula offered in the Information Technology Area.

INTRODUCTION

ICT Technologies involve a remarkable enhancement of the expressive possibilities with obvious impacts on learning methods and the distribution of educational contents. E-learning represents a new concept of education – it combines the advantages of distance education (accessible to all) with the total communication supplied by Internet and the new multimedial instruments. Its innovation potential grows over time while the medium itself becomes better understood, as it happened with televison technology in the past.
By comparison with the traditional education, e-learning allows a more effective student involvement. It enhances the teaching environment by enabling students to learn, to experiment, and to take a more interactive role in the educational process. Web sites and in general all the instruments of e-learning (including "text books") allow for the interaction with a great abundance of contents. In particular, the static contents of handbooks are continually developed (i.e. educational evolution), while web sites are upgraded – all to match the changing needs of students and tutors.
The dynamics of this approach represents a change in daily communication too: from the bi-directional communication teacher/student to a total interpersonal communication: face-to-face (i.e. as with conventional teaching), at a distance (i.e. without direct face to face contact), or on-demand from a central repository.
The information flow comes from a variety of sources, going beyond the boundaries of each classroom, resulting in a total communication that allows new approachs to training and learning as well as student/tutor contacts.
Due to the nature of Internet, the forecast is for remarkable increase in:

- the number of information sources
- the variety of information sources
- the possibility of documentation upgrades.

Obviously, this process requires us to carefully choose the correct approach to e-learning, such as the definition of ontologies and standards that enable: -

- teachers to easily prepare contents,
- the easy upgrade of the resources and exchange of information among teachers, students, repository of knowledge and information.

A COMMON E-LEARNING STANDARD: SCORM

E-learning needs interoperability thruogh a common standard that allows instruments and methods to create and share resources. Deciding a common standard actually means to be able: -

- to transfer the contents from an architecture to the others,
- to integrate them,
- to choose them for their unique characteristics and classifications,
- to certify them with a certificate that indicates the acquired competences.

The definition of standards that render compatible courses and platforms developed with various systems introduces multiple advantages. (GIORGIO – WHERE IS THE SECOND ETC – IF NONE THEN DELETE THE WORD "Firstly") Teachers and students do not have to learn the use of different platforms – thus saving time, money and hard work. As for technical aspects, it is more complex to care for the maintenance of different platforms than a single one. The adoption of development standards allows the teacher to integrate an existing course, by adding didactic material created with a various authoring systems. Among different models, ADL Scorm is the more credited and adopted standard.

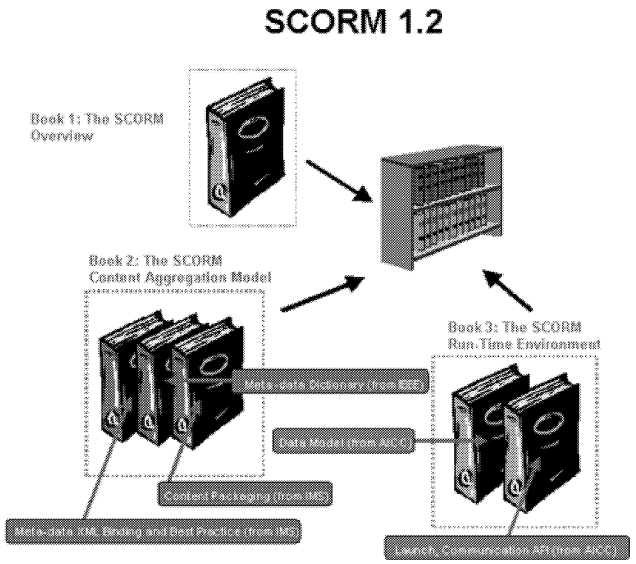
SCORM (Sharable Content Object Reference Model) defines model that consists in:

- a Web-based learning "Content Aggregation Model"
- a "Run-Time Environment" for learning objects.

A learning object is a small, or "atomic" component of a course being based on a single topic or a single lesson that is shareable as a resource to compose other teaching activities and courses.

SCORM consists in a collection of specifications that provide some guidelines to enable interoperability, accessibility and reusability of web-based learning content, following the RAID model (Reusability, Accessibility, Interoperability, Durable).

SCORM originates from the ADL (Advanced Distributed Learning) an initiative to join various groups and interests, such as IMS (IMS Global Learning Consortium) , AICC (Aviation Industry Computer-based Training Committee), IEEE (Eye-triple-E, Institute of Electrical and Electronics Engineers), ARIADNE (the European group).

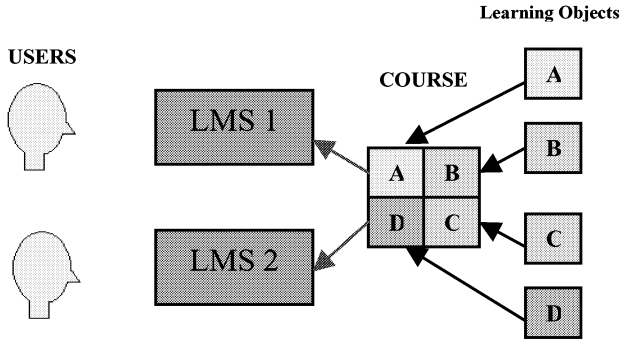


Figures 1: SCORM structure from ADL

SCORM includes guidelines on learning management systems, content authoring tools, course design and content developing.

SCORM architecture is composed by four essential elements: -

1. A Learning Object: the "atomic" component of a course. If it is SCORM-compliant, it can be shared it with other courses
2. LMS (Learning Management System): - the course management distribution system
3. CSF (Course Structure Format): - the exchange file to transfer the course in another LMS
4. Runtime: - the system that runs the course, following requests and monitoring user progress



Figures 2: Interoperability

SCORM AND ONTOLOGIES: THE SEMANTIC WEB

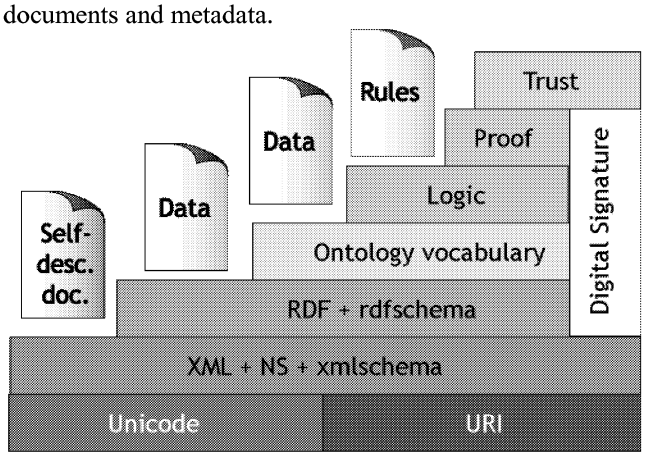
Talking about standards introduces the theme of ontologies. While Internet and E-learning have determinated the need for a common standard, such as SCORM, another emerging topic is the Semantic Web. As noted earlier this is a new web technology key to make web-based information and services universally understandable and reusable (both by human and machines).

There is considerable synergy between E-learning standards, such as SCORM, and the semantic web - both of them aim to share and reuse information through the web.

Ontologies, being a representation of a shared conceptualization of a particular domain, are a significative component of the Semantic Web. Many people think that ontologies and semantic web technologies will influence the new E-learning systems. In fact, while standards define the syntax, to exchange information and learning contents, ontologies can define semantic rules to guarantee the interoperability between different LMS or educational applications through the web.

Standards, ontologies and semantic web are closely linked because they deal with the ability to find and use information present on the web. The semantic web consists of "information about information", that look like e-learning SCORM metadata. Tim-Berners-Lee define the semantic web as "... an extension of the current web in which information is given well-defined meaning, better

enabling computers and people to work in cooperation." The link between SCORM objectives and the ontologies of Semantic Web is evident. The semantic web is mainly based on XML (eXtensible Markup Language) and RDF (Resource Description Framework). The XML permits the creation of tag to define a semantic structure of the document. XML tags are defined in specific Data Type Definitions (DTDs) or in XML-Schemas. RDF and RDF-Schemas are used to define the relationship between documents and metadata.



Figures 3: The vision of Semantic Web by Tim-Berners-Lee

The definition of a Learning Object reflects these concepts. They are provided of a complete classification system, thanks to metadata (LOM: Learning Object Metadata) written in XML with the consequent advantages as regards flexibility and content management. We have just seen that RDF adds the relational rules allowing us to identify an object as unique on the web, through its URL or URI (address that exactly localizes a resource in the web).

An important target for developers of web-based applications and E-learning developers, is to provide information and knowledge elements that are easily accessible and shareable with others.

Ontologies and web standards such as XML (used in SCORM), RDF, XTM, OWL, DAML-S, RuleML, allow specification of elements in a standard mode. In this way such elements become mobile and accessible within web information and applications, including the e-learning ones. E-learning and web-based educational systems combine standards and recently ontologies in order to define new methods to create courseware. Defining E-learning standards such as SCORM and ontologies such as those of the Semantic Web concur in satisfying the education community, by creating flexible and shareable resources which are useful to teachers, instructors, authors and learners.

Ontologies represent an additional level where concepts and logical rules are associated. An ontology can be seen as a vocabulary of terms with exact definitions and a collection of formal axioms related to their use. An ontology allows the performance of inferencies and, if shared, will improve the capability of component exchange and reuse.

CONCLUSION: WHAT'S THE FUTURE?

It is evident that the Semantic Web could be the universal standard formalization to represent knowledge in Internet. Also the Semantic Web has strong influence on the new E-learning approach. Moreover, ontologies represent an additional level to combine standards such as SCORM in e-learning with techniques and language of the Semantic Web. The best ontologies will probably emerge from those used by the largest number of developers and authors. We must appreciate that this could be a symptom of rigidity, yet it offers more opportunties of exchanging and sharing resources and information, throughout E-learning and the web.

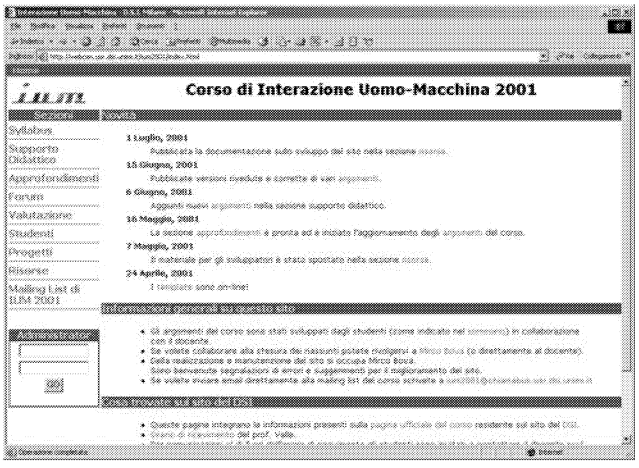
In E-learning the definition of ontologies will allow the development of a more flexible Learning Management System capable of satisfying users educational needs by means of the customization of the learning process [Cramp et al., 2000]. We are able to navigate through different learning object repositories and throughout the web, finding resources, concepts as well as their relationship.

In E-learning, ontologies could provide also a thesaurus of specific terms or specific didactic methods.

A SAMPLE APPLICATION OF STANDARDS: EVOLUTION OF WEBCEN

During the past academic years, particularly from 2000-2001 up to now, a significant project in web-centric didactics (WebCen, at URL: http://webcen.usr.dsi.unimi.it) has been developed according to:

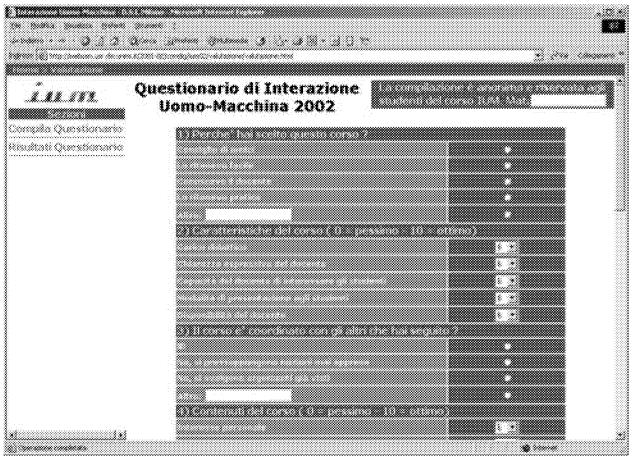
- teachers' and learners' needs
- progress in multimedia technologies
- standards evolution.



Figures 4: A Web page of WebCen

During the initial development of the project, WebCen offered the teachers a rich environment to manage education and useful instruments, whilst the learners the possibility to explore and use information. Since 2000 teachers can store and manage information, course material, a syllabus, news and every resource useful to communicate

interactively with students – all thanks to this web-based application.
Learners can access materials, interact with teachers by mail and discussion for a and actively participating in the creation of the cognitive process. Moreover, students can express preferences and suggestions through a valuation form that allows the establishment of a virtuous cycle - one that makes it possible to address the educational offer to specific needs – i.e. Quality Assesment, Assurance and Feedback in Education



Figures 5: The course evaluation form

Since 2003 the study of SCORM standard and semantic web themes has involved the experimentation of new features, such as:
- lesson video recording
- formative recovery through specific actions, such as thematic sites and personalized sessions for the students
- distance e-learning environment, such as LearnLinc (http://www.edtlearning.com/)

In fact, thanks to the student profiling and intermediate tests during a course, each student can monitor their personal learning level and have the possibility to recover gaps "in bite sized pieces" in their knowledge, i.e. educational actions only aiming to recover knowledge gaps emerging from the tests.
It is important to underline that the formalization of the learning objects is used not only to improve the knowledge base at the teachers' disposal, but especially for educational "catch up"actions. We have found, by involving students themselves, that the following are the most favoured:
- didactic methods
- SCORM standard
- emerging topics like the semantic web and XML.

Moreover, to recover educational gaps; documents, videos and samples, written by the students themselves (under the supervision of the teacher), can reflect their most favoured learning methods, enabling the development of the learning process, didactic practical sessions - all leading to an enhancement of student motivation and the learning experience – all resulting from the effective use of these new E-learning technologies.

The experimentation of LearnLinc represents a significant first experience of distance lessons and highlights the real possibility of strong interaction, even in distance learning.
The next steps in the WebCen project are:
1. the strong integration of the SCORM and Learning Object methods to better use all didactic instruments, distance learning and the power of the semantic web
2. getting teachers and students familiar with a specific cycle of lessons on SCORM standard, semantic web, and onthologies in order to create:
   - a web site that really supports students and teachers, thanks to the adoption of the SCORM standard,
   - a repository of shareable contents and experiences.

REFERENCES

Berners-Lee, Tim, Hendler, James and Lassila, Ora, 2001, "The Semantic Web on XML", Scientific American, May.
Dave Feasey, "More on Ontologies and eLearning", 29/01/2002
Crampes, M., Ranwez, S., "Ontology-supported and ontology-driven conceptual navigation on the World Wide Web, Hypertext 2000.
Vladan Devedžic, "The Semantic Web - Implications for e-learning", University of Belgrade, Serbia and Montenegro, tutorial presentation for iiWAS2003 Iakarta (Indonesia)
IMS (IMS Global Learning Consortium) http://www.imsproject.org
ADL (Advanced Distributed Learning) http://www.adlnet.org
AICC (Aviation Industry Computer-based Training Committee) http://www.aicc.org/
IEEE (Eye-triple-E, Institute of Electrical and Electronics Engineers) http://www.ieee.org
ARIADNE (europeo) http://www.ariadne-eu.org/

AUTHORS BIOGRAPHY

GIORGIO VALLE, born in Genova, Italy, on July 26, 1942, Giorgio Valle is Professor of Applied Informatics at the School of Mathematics, Physics, and Natural Sciences of the University of Milano, where he teaches courses on Human-Machine Interaction, CAD, CAM, and Information Management & Databases.
A graduate in Electronic Engineering at Polytechnic of Milano, his initial research activity earned recognition as ACM Siggraph Computer Graphics Pioneer for making the first computer graphics system in Italy in 1967. He has been active ever since in the area of computer-aided design and manufacturing systems both in research and education.
He lectured extensively: at the University of Bologna from 1970 to 1980, at the University of Calabria in 1980-83, and presently at the University of Milano since 1983.

RAFFAELLA FOLGIERI, born in Maratea, Italy, on June 23, 1967, Raffaella Folgieri has a long experience as high consultant in Information Technology, spent in Finance and Industry fields. A graduate in Computer Science at University of Milano, she is now involved in WEBCEN activities and in teaching support at the University of Milano and at the University of Milano Bicocca

SCORM as the Standard for E-Learning Framework

Mieczyslaw L. Owoc, Krzysztof Hauke
Department of Artificial Intelligence Systems – Wroclaw University of Economics, Poland
ul. Komandorska 118/120 53-345 Wrocław, Poland
Phone: ++4871-3680513; Fax: ++4871-3679611; E-mail: {mieczyslaw.owoc, krzysztof.hauke}@ae.wroc.pl

KEYWORDS

SCORM, SCO, distance learning, education via WEB

ABSTRACT

The importance of standards to the continued growth, expansion, and evolution of e-learning cannot be overemphasized. Standards foster efficiencies and synergies that enable markets to grow and the promise of e-learning to be realized [1]. SCORM is probably the most important and widely emerging e-learning standard today, whose ultimate goal is ensuring ubiquitous access to the highest quality education and training, tailored to individual needs, and delivered cost effectively anywhere in the world at anytime [6].

WHAT IS SCORM?

The Sharable Content Object Reference Model (SCORM) [7] was first developed by the U.S. Department of Defense (DOD) to address training development and delivery inefficiencies across its service branches. E-learning content was being developed on different platforms, using different standards and specifications, and delivered on different, incompatible systems. To address these costly inefficiencies, the DOD knit together the best emerging e-learning specifications with those developed in the prior decade by the Aviation Industry CBT [13] Committee (AICC). The result is a field-tested common reference model published by the Advanced Distributed Learning (ADL) [15] Initiative, a collaborative effort between government, industry, and academia sponsored by the Office of the Secretary of Defense. The SCORM standard is focused on enabling the plug-and-play interoperability, accessibility, and reusability of Web-based learning content, with the ultimate goal of ensuring ubiquitous access to the highest quality education and training, tailored to individual needs, and delivered cost-effectively anywhere and anytime. Based on accepted technology standards including XML and JavaScript, SCORM [8] is fast-becoming the de facto e-learning technology standard widely embraced and supported today by world-leading corporations, universities, system providers, and content vendors.

HISTORY SCORM

The federal government spends millions of dollars each year to develop e-learning content, including online courses, courses distributed on CD's and intranets. In the 1990's the government recognized that it was difficult to reuse this content. The Department of Defense, for example, found that the various branches of the military had developed e-learning content on similar topics, such as management and acquisition rules. Even though those courses essentially covered the same content, it was nearly impossible to share e-content between military branches because they were developed without a common standard, and they were not designed for reuse in other courses. The government also realized the benefits of an international standard for e-content on the training industry. A common international standard for sharing learning content would stimulate an international learning economy, similar to the economy that is developing around the Internet. If standards allow for reusing learning content developed for one course, then learning content will become a commodity.

As a result, in 1997, the Department of Defense and the White House Office of Science and Technology Policy launched the Advanced Distributed Learning (ADL) [9] initiative. Its primary goal is to develop a learning economy by providing access to high-quality education and training material, easily tailored to individual learner needs and available whenever and wherever needed. To accomplish this goal, the ADL consolidated emerging e-learning specifications from the major international standards groups into a single specification, referred to as SCORM.

Simply stated, SCORM is a set of specifications for developing, packaging and delivering high quality education and training materials whenever and wherever they are needed. SCORM- compliant courses leverage course development by ensuring that compliant courses are ***RAID [5]*:**

- **Reusable:** easily modified and used by different development tools,
- **Accessible:** can be searched and made available as needed by both learners and content developers,
- **Interoperable:** operates across a wide variety of hardware, operating systems and web browsers, and
- **Durable:** does not require significant modifications with new versions of system software.

Although this is a government initiative, it is be wrong to think that this cannot be used in industry and academia. In fact the SCORM specifications are a composite of several specifications developed by international standards organizations, including the IEEE [12], IMS, AICC and

ARIADNE [10]. New versions of SCORM are now released every 3-6 months by the ADL. Each new release incorporates recent changes and expansions of existing international specification. This process is likely to continue for years to come.

COMPONENTS SCORM

SCORMcan be described in many ways. The Advanced Distributed Learning Co-Laboratories refer to this as the Sharable Content Object Reference Model (SCORM). SCORM [3] is described in terms of the following three components:

- Content packaging,
- Runtime communications,
- Course metadata.

Content packaging refers to the packaging of all resources needed to deliver a course into a single zip file. The format for this file is described by the SCORMaggregate model, which is based upon the IMS Content Packaging Specification, version 1.1.2. The zip file contains not only the course files, it also contains an XML file, referred to as the imsmanifest file, describing the course contents and content sequencing. The runtime communications in a SCORM-conformant course are conducted using two elements:

- Runtime commands for communicating student information to and from the LMS, and
- Student metadata for storing information on individual students.

Course metadata are data packaged with a course when it is archived in a SCORMrepository. These data allow a course author, or student, search a learning repository containing hundreds of lessons and courses and to identify the learning content they want to use or view. For example, the course title, description, keywords, etc. are all considered course metadata.

a. SCORMcourse packaging [3] - if a course is going to be shared between learning management systems or archived in learning repositories, then the organization and learning assets in the course need to be included with the course. In SCORM, this description must be included with the course and placed in an XML file with the name *imsmanifest.xml*. The structure required for this file is detailed in the SCORM . □□□content aggregation specification.

A SCORM imsmanifest file consists of four sections:

- a preamble section containing XML pointers to the schemas required for validating this file,
- a metadata section contain global course information, such as its title,
- an organizations section describing course sequencing,
- a resources section listing all the files used in the course.

The preamble section in this file is not listed, since its syntax is fixed for all courses. Details for the preamble section are found in the SCORMContent Aggregation Model specification2. The metadata section can be empty. However, in this example, the metadata section indicates that the schema for this file is the SCORMversion 1.2 schema, the specification introduced by the ADL in October, 2001.

The last two sections are generally the largest sections. The organizations section contains a list of lessons, or Sharable Content Objects, contained in this course. This example consists of a single SCO, identified as "OneSCO". That identifier appears both in the organizations section and the resources section. The resources section identifies the assets associated with this SCO. In this case there are two web pages, identified as "index.html" and "end.html," in this SCO [2].

A SCO is the smallest unit in a course that:

- contains meaning learning content by itself,
- might be extracted and reused in another course.

Frequently, it makes sense to define a lesson with 1-3 major learning objectives as a SCO. However, it would not be a violation of the SCORM specification to refer to each web page as a SCO. In fact, this is the approach used by the Macromedia SCORM extension. Each web page is treated as a separate SCO. In general, however, each SCO will consist of more than a singe web page.

b. SCORMRuntime Communications [3] - not all courses require runtime communications with the learning management system (LMS). However, many courses contain content that adapt to the learners actions in the course, including scores on assessments and reviewed content. This requires tracking of scores and progress of individual students. This is a major service provided by a LMS. Today, non-SCORM [3]learning management systems use proprietary methods for obtaining and tracking runtime information. Generally, this is restricted to assessment scores and simple indications of whether certain learning content has been reviewed. For SCORM-compliant learning management systems, on the other had are required to provide commands for reading and writing student information to its database. Currently there are 8 commands available in SCORMfor communicating 49 different student metadata elements.

These student metadata include:

- 15 elements for capturing the learning state of the SCO,
- 8 elements for describing and tracking learning objectives associated with an individual SCO,
- 5 elements for student language, audio and video preferences,
- 4 elements for tracking a student's progress and time limits for individual SCO's,
- 13 elements for describing and tracking a student's responses and performance on quizzes,
- 4 elements for communicating data between SCO's and the LMS.

c. SCORMCourse MetadataSCORM [3] - contains a rich dictionary of metadata terms that can be used for describing course content. These data are not needed, if a course is never going to be archived in a learning repository or shared with other authors. However, the vision of the Advanced Distributed Learning Co-Labs is to create a learning economy in which authors and students

will be able to search the Internet for learning resources. This type of searching and discovery requires that courses archived in a repository include not just its content, but also a readable description of that content.

The SCORMmetadata specification is essentially the IMS [14] Learning Resource Metadata specification, which itself is based upon the IEEE Learning Technology Standards Committee9 and the Alliance of Remote Instructional Authoring and Distributions Networks for Europe [17].

BUSINESS BENEFITS OF SCORM

From a business perspective, standards are beneficial because they are essential to the growth and expansion of any technology-based industry. Whether it's 802.11 for wireless networking, HTML for the Web [4], or standardized track gauges for railroad transportation, standards foster efficiencies and synergies that enable markets to grow. These benefits certainly apply to e-learning standards, allowing organizations that adopt SCORM [17] to create efficiencies, lower costs, reduce risk, and increase overall learning effectiveness and Return On Investment (ROI) [4].

a. greater efficiencies & lower costs - SCORM [17] provides opportunities for significant improvements in business and development efficiency and cost-effectiveness.

- Reuse content for faster development—Developing content once, then reusing it for multiple audiences and contexts, reduces development time.
- Share content between systems—Moving to SCORM makes integration easier between existing and future systems, protecting your infrastructure investments and lowering your cost of ownership. SCORM learning content can be integrated with, and delivered on, past and future SCORM compliant systems.
- Reduce cost of content maintenance—By enabling your organization to maintain content in-house using any tool you choose regardless of system or content vendor, SCORM lowers your overall cost of content maintenance.
- Maximize technology investments—SCORM content can be launched, operated, and tracked by any SCORM-compliant content delivery system, whether content was developed in-house or by a third party, enabling you to get the most mileage from your technology investments.
- Avoid proprietary authoring tools—The same tools that your development team is using to create Web content can be used to develop SCORM-compliant content, eliminating the need to use or develop proprietary tools.
- Train developers faster—As SCORM adoption continues to grow, the talent pool of content developers experienced in SCORM techniques and technologies grows too. A wide selection of SCORM training materials is available to get less experienced developers up-to-speed fast.
- Leverage best practices—Take advantage of the collective knowledge and expertise of the growing SCORM community that is continually evolving and enhancing SCORM functionality, tools, methodologies, and best practices with proven results.

b. reduced risk [17] - by design, SCORM reduces business and development risks because it enables content portability, durability, and interoperability.

- Future-proof courseware investments—A key advantage of SCORM-conformant courseware is that no matter who developed it, when they developed it, or for what learning platform, it can be "played" seamlessly on any existing or future SCORM-based delivery system. SCORM content is portable and durable, so you reduce the risk of not being able to play "old" content on "new" systems.
- Decrease reliance on proprietary tools and technology—Given that the future is inherently uncertain, the more your e-learning tools and technologies are based on standards, the more likely they will fit into tomorrow's e-learning environments.
- Reduce switching cost risk—Investing in standards today helps ensure that any costs associated with switching your e-learning initiatives from one platform to another in the future will be minimized.
- Lower obsolescence risk—Embracing standards also reduces the risk that the tools you are using, the content you are creating and deploying, and the knowledge and skills your people are acquiring will become obsolete. Standards help protect these investments, keeping reengineering, retooling, and retraining costs to a minimum.

c. improved learner experiences [17] - SCORM content and delivery systems enable organizations to create more compelling and effective learning experiences through dynamic sequencing, rich metadata, object-oriented design, and more.

- Dynamically configure personalized courses—Dynamically sequenced courses—those that are essentially built "on the fly" based on learner needs, roles, and knowledge levels—enable more targeted and personalized learning paths and experiences, which translates to more effective learning. A learner, for example, can "test out" of material they already know. The possibilities that dynamic sequencing opens up may prompt your instructional designers to create more effective solutions.
- Empower learners with more control—SCORM supports the building of content as discrete learning objects. When combined with the rich metadata standards in SCORM, object-based content allows you to provide learners with greater control over their individual learning experiences, thereby decreasing time-to-competency.
- Use performance data to motivate learners—SCORM enables the building of granular courses that provide rich and detailed tracking and reporting information about learner performance. Providing learners with detailed progress and performance information keeps them informed, motivated, and engaged throughout

the entire learning process, for a better overall experience and ROI.

THE FUTURE OF SCORM

SCORM [16] continues to evolve and develop functionality to meet the needs of e-learning developers, learners, and administrators. For example, the newly released SCORM version 1.3 includes sequencing functionality that greatly enhances support for varied course architectures and instructional designs. Future areas of SCORM functionality being explored include:

- Enhanced capabilities for delivering assessments, including support for randomization, question pools, item-level storing of response data, and robust assessment reusability,
- Better system-to-system integration,
- Support for simulations, dynamic presentation control, and searchable SCO repositories,
- A common format for describing skills and competencies that can be shared across learning management and delivery systems,
- Support for the authoring of learning experiences that include content (IMS or SCORM packages) plus collaboration with peers, tutors, and administrators
- Standard methods for accessing remote learning content repositories across a network

By making the move to SCORM now, your organization will be aligned to take advantage of these and many other measurable benefits as ADL integrates significant functionality into the overall standards road map. It is also important to note that, in its work, ADL is highly sensitive to backward compatibility, with great effort being expended to ensure that future and existing SCORM functionality harmonize well as an integrated whole that can be validated with real-world interoperability results. To keep abreast of the latest SCORM developments, visit SCORM is central to DigitalThink's [2] product strategy. Our own experience in making the move to SCORM has been that after start-up costs, SCORM is not only much less expensive from a development and maintenance perspective, but it provides much more flexibility in creating and delivering effective learning solutions aligned with the needs of our customers. Because DigitalThink supports hundreds of SCORM content developers internally, we have extensive practical experience developing, delivering, and supporting SCORM content and learning delivery systems. We leverage that experience to provide our customers with products and services that take full advantage of the benefits of the SCORM standard:

- SCORM-Compliant Custom Courseware [11] - If you are exploring outsourcing some or all of your content development, DigitalThink [2] is a proven partner. With thousands of hours of courseware developed, DigitalThink [2] is the most experienced custom e-learning company in the industry. Our seasoned e-learning strategists have decades of combined experience creating highly effective e-learning, and they understand that learning approaches that work well in the classroom must be adapted to work well online. A reflection of our SCORM-focused product strategy, all DigitalThink custom courseware [11] is SCORM compliant, and optimized to take full advantage of the benefits of the standard. To ensure the best results for our customers, we offer a full range of custom courseware capabilities and services—from planning, design, and development to delivery, maintenance, and support.
- SCORM-Native Learning Delivery System - DigitalThink's [2] L5 Learning Delivery System is the industry's only SCORM-native learning delivery system. "SCORMnative" means that L5 was designed from the ground up with SCORM as its native content format. L5 separates the learning interface layer—called the L5 Learning Environment—from learning content. Navigational tools, learning tools and services, and "look and feel" interface branding are all part of the L5 Learning Environment, and all are managed by the L5 Learning Delivery System rather than being hard-coded into the learning content itself. This approach takes maximum advantage of the SCORM standard, yielding benefits such as lower content development and maintenance costs, made possible by more effective content reuse. L5 offers configurable learning environments, online/offline delivery, developer tools, e-learning administration, and enterprise integration capabilities. To date, DigitalThink [2] has invested over $50 million in L5 to make it not only the most reliable delivery system available, but one that ensures a superior experience to every single learner, every single time. L5 has already delivered over 8,000,000 course hours—all while maintaining the industry's highest availability rates with uptime exceeding 99.7%.
- SCORM Developer Support Program - The DigitalThink L5 Developer Program provides tools and resources to help your organization's internal development team create and maintain courses in-house. These tools and resources enable your team to transition to SCORM as quickly as possible, by giving your developers a standard level of SCORM proficiency.

REFERENCES

1. Black, J. 1997, Online students fare better, available from C/Net News at http://news.com.com/2100-1023-263035.html.
2. DigitalThink 601 Brannan Street San Francisco, CA 94107 - www.digitalthink.com ©2003 DigitalThink, Inc
3. Jones, E. R. and Martinez, M., 2001, Learning Orientations in University Web-Based Courses. Proceedings of WebNet 2001, Oct 23-27, Orlando, FL., available online at http://www.tamucc.edu/~ejones/papers/webnet01.pdf. http://www.scorm.tamucc.edu.
4. Jones, E. R., 1999, A comparison of an all web-based

class to a traditional undergraduate statistics class. Proceedings of the Society for Information Technology and Teacher Education – SITE 99 San Antonio, Texas, Feb 28- Mar 4. Available online at http://www.tamucc.edu/~ejones/papers/site99.pdf .
5. Jones, E. R., 2000, Student behavior and retention in web-based and web-enhanced classes. J. of Computing in Small Colleges 15, 3, pp 147-155, available online at http://www.tamucc.edu/~ejones/papers/ccsc.pdf .
6. Schutte, J. G., 1997, Virtual teaching in higher education: The new intellectual superhighway or just another traffic jam? Online report, Calif. State Univ. – Northridge, available at http://www.csun.edu/sociology/virexp.htm .
7. The ADL Co-Laboratory, 2001, The SCORM Content Aggregation Model, available at http://www.adlnet.org.
8. The ADL Co-Laboratory, 2001, The SCORM Overview, available at http://www.adlnet.org.
9. The ADL Co-Laboratory, 2001, The SCORM Runtime Environment, available at http://www.adlnet.org.
10. URL: ARIADNE – http://www.ariadne-eu.org .
11. URL: Final Report of the Web Courseware Research Committee, Texas A&M Univ. Corpus Christi – http://www.tamucc.edu/~ejones/wcrc/wcrcreport.pdf .
12. URL: IEEE Learning Standards Technology Committee (LTSC) - http://ltsc.ieee.org/ .
13. URL: The Aircraft Industry CBT Committee – http://www.aicc.org .
14. URL: The IMS Global Learning Consortium, Inc. – http://www.imsproject.org .
15. URL: The Joint ADL Co-Lab - http://www.jointadlcolab.edu .
16. URL: The online SCORMcourse - http://www.scorm.tamucc.edu.
17. Edward R. Jones, Ph.D.: Implications of SCORM and Emerging E-learning Standards On Engineering Education, Computing & Mathematical Sciences Dept. Texas A&M Univ. Corpus Christi

BIOGRAPHIES

Hauke Krzysztof, Ph.D. is a lecturer in the Faculty of Management and Computer Science Wroclaw University of Economics, Poland. Dr Krzysztof Hauke has authored almost 40 publications mostly oriented on databases and intelligent systems topics. His scientific interests concentrate on expert systems, knowledge discovery problems, computer supported learning, databases, temporal databases, object databases. In recent years he has been engaged in the international projects: *"Knowledge Acquisition and Intelligent Distributed Learning in Resolving Managerial Issues"* with Belgium

Owoc Mieczyslaw L., Ph.D. Eng. is a lecturer in the Faculty of Management and Computer Science Wroclaw University of Economics, Poland. Dr Owoc has authored almost 100 publications mostly oriented on databases and intelligent systems topics. In recent years he has been engaged in the international projects: *"Knowledge Acquisition and Intelligent Distributed Learning in Resolving Managerial Issues"* with Belgium and *"Machine Learning 2000"* with Sweden and Latvia. His current research is in modern information technologies including distance learning and knowledge management with focus on knowledge acquisition and validation.

BLENDED LEARNING

BLENDED LEARNING WITHIN A UNIVERSITY CONTEXT: AN EVALUATION AND COMPARISON OF A MODULE IN ONLINE MODE AND TRADITIONAL MODE

Adriana Gnudi
Agostino Lorenzi
Lucia Malvisi
Università di Bergamo
Facoltà di Economia
Via dei Caniana 2
24127 Bergamo

Email: {adriana.gnudi, agostino.lorenzi, lucia.malvisi}@unibg.it

KEY WORDS
Blended learning, evaluation.

ABSTRACT

The changing face of students attending universities and ever increasing class numbers call for different approaches to teaching and the advent of distance learning creates an effective and stimulating learning environment for students that allows them to develop a collaborative mentality towards their peers. At University of Bergamo, Italy, courses have been developed where traditional classroom delivery sits side by side with courses that are delivered online. Such courses are termed "blended learning". A first year Informatics course for students within the Faculty of Economics offers two possibilities: the first a traditional classroom mode with some online activities and the second a more blended learning mode where a higher percentage of the course is delivered online. This paper sets out an evaluation and comparison of the two modes of delivery by comparing both qualitative and quantitative data in the shape of questionnaires, access to online materials, tests and exam results.

INTRODUCTION

New technologies allow student to access learning materials and tutoring services online without being constrained by time and place and thus new modes of learning are developed. The acquisition of basic Information and Communication Technology (ICT) skills is considered a vital starting point for students in the first year of any University course. At the Economics Faculty, Bergamo University, Italy, a level one Informatics course is offered to all students newly enrolled on a degree course within the faculty, in order to achieve a uniform standard in first year students. The module offered also sets the standard for the minimum ICT skill set required by a student embarking on an undergraduate course. Whilst planning the course, lecturers decided that the use of an online environment should also become a prerequisite skill for new students. These skills must be acquired in the first year of study so as to allow students to gain most advantage from them as they progress through their University career. The platform Lotus LearningSpace Forum was chosen as the VLE (Virtual Learning Environment), as it allows the creation of certain key elements fundamental to any online learning environment:

- the ability to structure the course into distinctive modules or learning units
- the facility to undertake tests and assignments on-line
- the possibility of interacting with lecturers and students via the web
- the ability for students to participate actively in the course by posting questions and replying to discussion threads, as well as reading or downloading files.

One difficulty facing all modern Universities is large cohort numbers. By allowing students to participate in courses through VLEs, many of the problems associated with large classes are resolved. Student support becomes more immediate within an online environment, where students have greater access to interaction with lecturers, which in traditional classroom based courses is either infrequent or lacking completely. As noted by Cavalli and Lorenzi "distance learning becomes active learning and the lecturer is the instructor who moderates activities, whose principle objective is to facilitate learning." (Cavalli and Lorenzi 2000). In this way it is hoped that students will undergo a change in their learning culture, achieving a different approach to study compared to their classroom peers.

COURSE STRUCTURE

The informatics course was offered in two modes:

- Classroom Mode: students attend 48 hours of face to face teaching, with extra hours of tutoring in the labs and individual tutoring with lecturers.

- Blended Learning Mode: students undertook 24 hours of face to face teaching and a recommended 24 hours online. As the mode was principally aimed at students who also worked or students who were not able to attend the University during the day, classes were held in the evening or on Saturday mornings. On reviewing the course, it became apparent that the designated 24 hours online was an indicative figure only, as participation online by students and tutors greatly exceeded this number during the course.

The two modes were run parallel with each other, within the same time frame of three months (October, November, December), with the same contents and same teaching materials. Students from both modes took the same exams during the module. The exam consisted of three parts, held at the end of part one, part two and part three of the course.

In the blended learning mode, students were given a specific login name to access Lotus LearningSpace. LearningSpace allows students to participate in forums (asynchronous discussions), undergo tests that are corrected by LearningSpace itself (assessments), complete tasks assigned and evaluated by the tutor (assignments) and have access to frequently asked questions (FAQ). Assignments were given a mark by tutors which, in conjunction with the assessments undertaken, allowed students to have an ongoing evaluation of their progress throughout the course. Tutoring sessions held at the University focused on problems encountered online coupled with activities in the PC labs.

Students who took the classroom mode were also offered the opportunity to use the online environment. A specific module was created online, not shared the blended learning group and classroom students were given a generic login. Lesson outlines were posted, exercises and their answers were given and students were able to participate in their own discussion groups, albeit with a generic login name. Students were free to choose at the start of the course which mode they preferred to subscribe to. The majority of students still preferred to choose the traditional classroom delivery, with over 500 students enrolled in the classroom mode and 92 students enrolled on the blended learning mode. Although not all the students who enrolled actually frequented the classroom mode, this is indicative of the number of students who took the first part of the exam after part one of the module. Most of the students who took the course in either mode were in the first year, though about 20% were in later years of their degree course. For almost all students this was their first experience of using an online environment, and with the timetabling of the blended learning classes in the evenings and at the weekend, this may explain why the numbers on the blended learning mode are relatively small compared to the classroom cohort. Students new to the online environment may need more encouragement before choosing a course to see the advantages and benefits offered therein.

STUDENT PROFILE

In both groups the majority of participants were female, with a slightly higher percentage of males in the blended learning group than in the classroom group. Most of the students within the Faculty of Economics are female (63% in the Classroom group and 58% in the Blended Learning Group.) , as the High Schools which supply the students tend to have a higher female component. While most of the classroom students were attending the first year of University, in the blended learning group students from other years make up almost half the group (Tab. 1)

Table 1: Year of degree course

	Classroom	BL
First year	78,7%	51,7%
Not first year	21,3%	48,3%

As has been shown in other studies (Gilroy et al, 2001) the distance learning environment tends to be favoured by non traditional students, that is students who may be older, who work or who have other obligations, such as family, that do not allow easy access to traditional classroom courses.

At the beginning of the course both sets of students underwent a test to assess their basic computer skills. Interestingly, the blended learning group generally demonstrated a higher level of basic PC skills than did their classroom counterparts (Tab. 2) Again this suggest that students need to feel comfortable with the Internet and computers before they feel ready to tackle an online course and may not desire to do so unless other circumstances weigh on their choice.

Table 2: PC skills at start of module

	Classroom	BL
Average test result (%)	60%	71%

PARTICIPATION AND EXAM RESULTS

In order to evaluate the participation of the students within the blended learning group, an active participant was considered one who posted a question or reply in the forums, or completed an assignment, assessment or test online. Of the 92 students who enrolled in the blended learning mode, 41 actively participated online in forums, corresponding to a percentage of about 45%. As regards the classroom group, an average of

approximately 300 students attended lectures, a percentage from the original enrolled figure of 537 of almost 56%. Both modes reflect a typical occurrence within the Italian University system, with a high number of students who initially enrol on a course, followed by a relatively low number of students who actually attend.

With regards to exam results, a notable difference can be seen between blended learning students who actively participated and those who did not (Tab. 3).

Table 3: Participation and results of blended learning group (BL)

	Exam passed	Exam failed	Exam not taken	Total
Students who actively participated	28,3%	8,7%	6,5%	45%
Students who did not actively participate	15,2%	10,9%	30,4%	55%

As can been seen, students who actively participated were more likely both to take the exam and to pass. Over 70% of those who participated and took the exam passed. Students who participated actively in LearningSpace almost certainly gained a greater advantage from having more contact with both the lecturer and their peers. On average, blended learning students took 5 online assessments and in questionnaire submitted after the end of the course, those who took more assessments throughout the module stated that these were beneficial to them in tracking their progress, highlighting weak and strong points and helping them to prepare for the exam. As both student cohorts took the same exam, it is useful to compare the final marks obtained (Tab. 4).

Table 4: Exam scores

	Classroom	BL
% of students who passed course exam	65%	69%

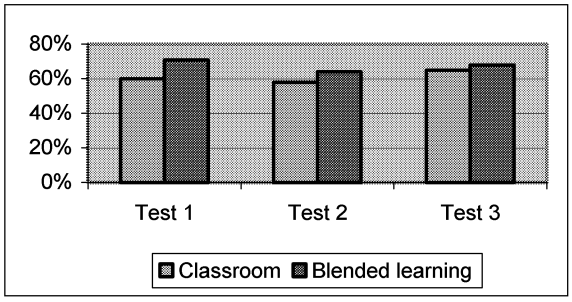


Figure 1: Exam scores for both groups

As can be seen from Figure 1, the blended learning group compared favourably with the classroom group, in all three parts of the exam.

A QUANTITATIVE ANALYSIS OF THE ONLINE ENVIRONMENT

As noted earlier, both modes were given access to online support through a VLE where participants could download material or actively participate in a virtual classroom through discussions and tests. In effect the online environment can be broadly divided into two categories:

- a depository for learning materials (documents from the lecturer, exercises, answers etc)
- the virtual classroom (forums, delivery and correction of exercises, tutor observation and interacting with tutors and students)

Access logs were available where students' behaviour could be evaluated once in the VLE, for example, demonstrating how often a student participated in a discussion or downloaded material or completed an online test (assessments). From Table 5 it can be seen that the blended learning group were far more active participants within LearningSpace than their classroom peers, especially with regards to participating in discussions. A far higher number of blended learning students posted questions and contributed to threads than in the classroom group, demonstrating a more collaborative attitude towards their studies and fellow students. Students in the classroom group made use of the VLE in a far more passive way, generally keeping to downloading materials that complemented classroom activities, without properly engaging in discussions amongst themselves or with the lecturer (Tab. 6). For the blended learning group, participation in discussion rooms increased significantly as the course continued. As Dr WG Lockitt says of his research into individual learning styles and e-learning "all learners identified 'hands-on' as the preferred method of learning….the overriding feature of the term hands-on given by respondents focused on the involvement of the learner in the learning process and the perception that their input was being integrated into the learning process itself." (Lockitt 2002). Once students see the benefits of sharing information with peers and lecturers they become more willing to post up question and replies and this case is no exception.

Table 5: Access logs for blended learning group (BL)

Blended Learning Group	Download materials	Virtual Classroom	Other
October	88,2%	9,6%	2,2%
November	84,8%	13,5%	1,7%
December	67,6%	25,2%	7,2%
Total for 3 months	*83,7%*	*13,5%*	*2,8%*

Table 6: Access logs for classroom group

Classroom group	Download materials	Virtual classroom	Other
October	89,1%	6,1%	4,8%
November	91,3%	6,5%	2,2%
December	92,1%	5,5%	2,4%
Total for 3 months	*90,2%*	*6,1%*	*3,7%*

A QUALITATIVE ANALYSIS OF FORUMS

As was to be expected, students in the blended learning group made far greater use of discussion rooms (forums) than their classroom counterparts. Over 120 discussions were initiated within the forum, with an average of 4 postings per student. The online group showed themselves to be fairly resistant to this form of communication with only 23 postings being initiated by students. The fact they were only given a generic login and also had far greater access to the lecturer in the classroom were almost certainly contributing factors to their lack of interest in the forums. Lecturer's attitudes within each group is also noticeably different. The blended learning lecturer began the discussions with a welcome note explaining the use of the forum and the advantages of using them. The classroom students received no such message, although an introductory seminar in the classroom on the use of forums was given. In both groups however, the majority of postings were directed specifically at the lecturer, demonstrating that the attitude of using forums as a glorified email still persists. This is especially notable in the classroom group where no students posted a reply to a question. The lecturer discerns this lack of collaborative approach by the students and tries to elicit some responses by asking students to actively contribute to discussions intuited by their peers, but to no avail. In the blended learning group, where a collaborative attitude gained ground more quickly and students could see advantages of it from an early stage, participants showed themselves to be more willing to communicate with each other and with the lecturer. As Gilly Salmon noted it is not a case of merely transferring classroom attitudes onto the web and hoping that discussions will flow, but that "students still need the 'champions who make the learning come alive – the e-moderator" (Salmon 2000). However it should be noted that neither group responds well to postings where the lecturer is not present and no thread exists where the students interact solely with each other.

In both groups the main topics discussed where logistical problems, how to get access to materials, how to attach files etc. When the lecturer from the blended learning group posted up questions on the students experience of the online environment they responded with enthusiasm, listing problems, advantages and general experience. It would appear that the students prefer to have the lecturer as overseer to their writings, and do not feel the need to communicate with each other through this medium when no lecturer is present. As Salmon again comments reflecting on Putman "new users search for rules and recipes early in the learning process. The best way to help them is offer a start, then 'stand back' and gradually let the user embed in his or her own experience." (Salmon 2000). Students in the blended learning group may feel that as time is a limitation for them, (most of them are full-time workers or have families), that they prefer to interact directly with the lecturer in order to be more time efficient. It is difficult to judge whether the discussions in the blended learning group go beyond a mere email type scenario, just as it is difficult to judge how many students read the discussions without participating actively in them, thus still getting some advantages from the forums. However it has been noted that the presence of the lecturer is fundamental to communication in both cases. If one wishes students to communicate more autonomously, it is necessary to state this explicitly at the start of a course. One student comments " elearning reduces distances, but it's still better to hear explanations in class!". Students from both modes do not feel confident enough to completely remove themselves from the lecturer's watchful eye and it becomes the lecturer's duty to create an environment where students feel comfortable, especially at the start of a module.

e-learning allows me to plan my time better	171
e-learning gives me more flexibility	168
I would definitely take another course with an elearning element	167
I would have preferred more assignments to auto evaluate my progress	165
I like the fact that I am able to look at various topics at once	161
……	…..
I don't post up questions as I'm afraid of being judged	35
I didn't find the assignments useful	34
I feel too isolated	31
I thought using LearningSpace would have been easier	27
I would have preferred a course without an elearning element	13

Figure 2: Some responses to questionnaire by blended learning students

Students on the blended learning course were asked to complete an online questionnaire at the end of the course with various statements relating to elearning where they had to indicate whether they agreed or disagreed. Fig. 2 demonstrates the five most popular and most unpopular answers.

As can be seen, students appreciate the salient factors of an online environment and do not feel isolated when using it. They are comfortable with the technology and

would prefer more online tests, a valuable part of the earning process. It is interesting to note that most would take another course with an online blended learning component. As noted earlier, one of the motivations behind the course was to offer students an introduction to elearning in the hope that this would form the basis of a practice that would remain with them throughout their University career and possibly beyond. From the experiences of the blended learning group here, it may be said that the course is a success in this respect.

GENERAL OBSERVATION ON THE CLASSROOM AND BLENDED LEARNING MODE

The blended learning mode activated a new approach to learning amongst the participants that has far greater consequences than the traditional classroom course. Students rather have begun to develop autonomous learning styles and ways of managing their time that will stand them in good stead for further courses with online elements. Whereas the classroom students tended to use Learning Space as a "library" where they could download materials, much in the same way they would collect sheets of handouts at the end of lecturers, the blended learning group really began to interact with each other and the lecturer in a way that is completely new to them. The VLE can be used as a tool to stimulate intellectual curiosity. There still remains the difficulty of creating a true virtual classroom where students ask each other questions and learn from each other. However as Cavalli and Lorenzi note, students begin to demonstrate this ability in subsequent years of the course, after greater experience of the online environment has shown the benefits of peer learning (Cavalli and Lorenzi 2000). The positive results obtained by the blended learning students in the exam are significant enough to grant validity to the mode itself. It is important to remember that this is almost certainly the first experience of a non traditional course for the blended learning students and the ability to make use of tests online assessments certainly contributed to positive exam results for these students. An online environment pushes students to continually verify their own progress, providing a stimulus towards learning.

CONCLUSIONS

With blended learning used extensively with one group and to some extent with the other, some changes were seen in the students. Students gained advantage from the course and the blended learning students especially seem keen undertake this mode of study again. Aspects unique to distance learning environments, such as the use of assessments are popular amongst students and should be made available to be used on an ad hoc basis, as and when students wish. It was also seen that an online environment requires significant changes of attitude on the part of the students and lecturers. If students are to develop a collaborative mentality of peer learning, they need both careful guidance and space to make the forums their own. Lecturers should be aware that novice users of an online environment may need a great deal of guidance before they are ready to participate in discussions without the presence of the lecturer and react accordingly. A delicate balancing act of participating just enough is required to create an effective forum. From the experiences seen here, some useful recommendations can be made for future use in courses that combine traditional teaching methods with an online aspect and to other colleagues who may be considering introducing e-learning into their curriculum:

- Group activities created especially for an online environment may serve to foster a collaborative attitude amongst students.
- The lecturer's role should be re-evaluated within the online environment at the start of the course.
- Particular attention should be paid to forums as these are the most suitable part on the online environment for collaborative work. Lecturers could post up questions that require some deeper thought from students and make it more explicit that students are expected to answer.

REFERENCES

Cavalli, E. and Lorenzi, A. 2000. "Methodology and technology for e-learning". In *Le Tecnologie dell'Informazione e della Comunicazione come motore di sviluppo del Paese.* Proceedings of 38th AICA 2000 conference, Taormina, Italy.

Gilroy, P.; Long, P.; Rangecroft, M.; and Tricker, T. 2001. "Evaluation and the invisible student: theories, practice and problems in evaluating distance learning provision". In *Quality Assurance in Education* Vol 9 number 1, pub. MCB University Press.

Gnudi, A., Lorenzi, A. 2002. "E-learning to acquire the basic ICT skills for first-year universitary students", In *Proceedings of 2002 Eden Annual Conference*, Granada, Spain.

Lockitt, W.G. 2002. "Individual learning styles and e-learning". In *Proceedings of EDEN second research workshop*, Germany.

Putman, R. W. 1991. "Recipes and Reflective Learning". In *E-Moderating: The Key to teaching and Learning Online*. Kogan Page, London.

Salmon, G. 2000. "E-Moderating: The Key to teaching and Learning Online". Kogan Page, London.

Tele-Education and Blended Learning in complex Networks of Competence

Christian-Andreas Schumann, Kay Grebenstein and Jana Weber
Mitteldeutsche Akademie für Weiterbildung e.V.
c/o University of Applied Sciences Zwickau
D-08056 Zwickau
Germany
E-Mail: christian.schumann@fh-zwickau.de

KEYWORDS
Education and Training using Multimedia, CBT, Teleworking, Teleprocessing and Telepresence

ABSTRACT

Whole organisations, multitudinous specialists, developers and producers work with and for each other in order to advance, develop and transfer information, knowledge, experiences, know how and technologies.

Networks of competence can be built wherever networking, a thematic focus or the intensification of research and development cooperation is in demand to improve the status quo and to increase business revenue.

Competence cells form the basic components for competence networks. The successful development in tele-education and blended learning will be focused on the extension of these networks based on growing competence of the educational cells.

In the present paper there is shown the fundamental as well as the special importance of competence cells when realising knowledge transfer via lifelong learning in SME.

INTRODUCTION

The recent development in Tele-Education is just to understand in the context of national and international, interdisciplinary corporation of public, non-profit and business units. The expenditures of the required resources to determine the subjective knowledge, to transfer it into evaluated know how of the organisation, and to include it into learning programs such as CBT (Computer Based Training) or WBT (Web Based Training), they are so enormous that there is a strong constraint for organisations to cooperate in tailored networks of specialists, developers and users. These networks are well known as networks of competence based on so called competence cells (Mueller 2003). Each cell is a separate organisational unit characterised by independence in law and economy, special knowledge, skills, performance, competence and interfaces.

The separate cells form temporary or constant networks of competence in the framework of business activities and processes in order to be successful in the market. The theory of networks of competence was developed especially for the production networking, but it is valid for networking in tele-education and blended learning, too. The Institute for new forms of education of the University of Applied Sciences Zwickau and the Academy of Further Education of Middle Germany are in this sense two public or non-profit competence cells, respectively. They are included in different processes of training and education and therefore forced to build up networks of competence with other partners. The following sections of the paper include an overview of the very complex activities and tasks of both units in the world of new education realised in networks of competence.

KNOWLEDGE MANAGEMENT, E-BUSINESS AND E-LEARNING IN CELLS AND CHAINS

Corporation and networking of education, training and learning have to be managed. Besides the general management and project management especially the management of data, information, skills as well as knowledge is essential. The common information management is supplemented by the knowledge management. Knowledge is the ability to define patterns of facts serving the preparation and the execution of actions and decisions. The similar relation is remarkable in the case of the general business with regard to the management of e-based processes and functions. The e-business is added to the common business and related to the e-learning activities embedded in the world of public, non-profit and profit business processes and workflows, respectively.

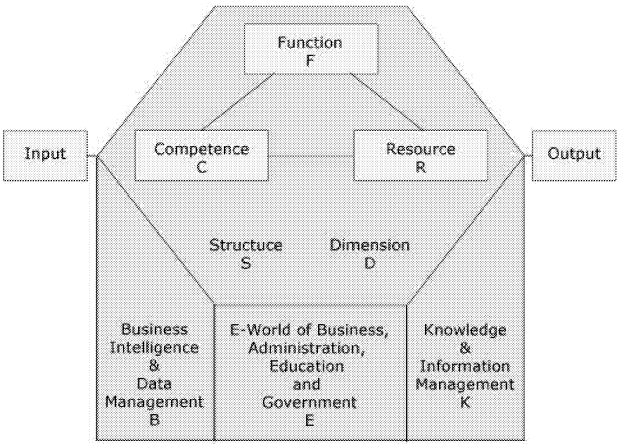


Figure 1: Generic Model of Educational Competence Cells

Both sides of management and business including the special fields of information management and knowledge

management as well as e-business and e-learning are influenced by and related to business intelligence. Business intelligence is the ability to generate knowledge by analysing the state of the art in the own business units and networks as well as in the competitive ones. All these aspects are focused on the competence cell for education and learning as part of the competence networks. (Fig. 1)

The competence cells of education and learning are arranged in chains consisting of development, supply and application modules. Due to the scale of competence of the organisational unit, the cells are single or multi-functional. The multi-functional cells include integrated functions and processes of development, supply and/or application.

According to the integrated functions and processes, the competence cells are put in its proper place in relation to the other cells forming the network of competence in all. The networking is one of the main aspects guaranteeing the success of the business, educational or learning unit, respectively. The competence network based on special competences of the integrated cells is able to support the whole chain beginning with the development, via supply services up to the application. (Fig. 2)

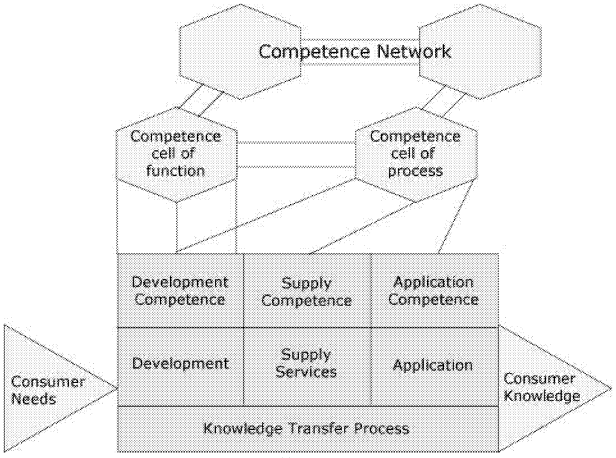


Figure 2: Competence Cells in the Competence Network

The Institute for new forms of education of the University of Applied Sciences Zwickau has the consulting and management competence for the whole chain from analysing the demands for further education up to the learners support. The Academy of Further Education of Middle Germany is characterised by special competences in development of courses and contents in e- and blended learning as well as in supporting learners and companies in the application of new opportunities in web-training and network context. Both institutions are multi-functional but forced to corporate in networks of competence in order to offer more complex services for the customers.

LEARNING ENVIRONMENT, CONTENT DEVELOPMENT AND QUALITY MANAGEMENT

The access to the world of e-learning is usually realised by web portals opening the gate to the learning space. Behind the portal, in the learning space, the supporting tools as learning platforms are arranged. In the past the Institute for new forms of education of the University of Applied Sciences Zwickau developed one prototype of platform in order to gain the necessary knowledge for choosing the right product for educational processes and training programs offered by the competence cells alone or in competence networks. Recently, the two competence units use several learning platforms, such as Blackboard, Distance Learning System, ILIAS open source as well as Saba Learning Enterprise (Baumgartner 2002), referred to the target group and the kind of the network of competence. The objective is to consume the existing commercial standard products of learning platforms in order to be able to focus the activities of the cells on the main competences and the core processes.

The platform is rather the frame for tele-education. The content in the learning space is possibly the most crucial aspect as it is often complicated and mostly expensive to produce the right content for the right target group in the right design and structure because of the enormous number of fuzzy factors of influence in educational processes. Therefore, the total quality understanding and management, accompanying the whole process from the development via the supply services up to the application, is essential.

The Institute for new forms of education deals with new kinds of distance education and blended learning, tele-teaching and tele-working, etc. The main competences are developed in new distance education courses, for instance in Business Engineering, Business Informatics, and Industrial Management, in content development, for instance for Enterprise Resource Planning, and in organisation of knowledge transfer in the university and in networks of competence.

The Academy is a competence cell dealing with the core processes of content development, the creation of competence networks, the optimisation of educational and business processes, the regional knowledge transfer, etc. Recently, further main competences were generated in the projects concerning media competence, regional knowledge transfer in innovative plant planning, quality management and assurance, international business relationships, business and market intelligence, analysing and optimisation of intercultural business processes, business and operational models for educational systems, standard contract design in corporation networks of SME, etc. (Fig. 3)

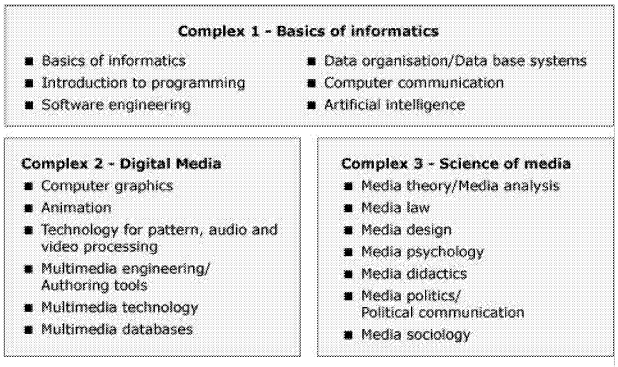


Figure 3: WBT Program "Media Competence for SME"

General and special quality management rules were developed and applied for all these activities. Thus, an expert group accompanies the processes of content development and application. The results of R&D are used in practical educational programs in the networks of competence including partners from science, administration as well as industry.

Usually, standard products such as the mentioned platforms, the adobe product family or authoring systems like Macromedia are used to generate efficient and high-quality products and offers for the customers. But for the complex task of tele-education the workflow systems are essential, too, and some special systems solutions have to be developed by the competence cell itself especially in order to improve the performance of the whole system.

WORKFLOW MANAGEMENT AND SPECIAL SOFTWARE SOLUTIONS

One of the biggest problems in realising E-learning projects is their complex production. In order to be able to create good learning contents it is very important to consider all aspects like an ideal project management and a consistent quality assurance. Therefore, it is necessary to have access to special information management systems.

The so called workflow management solutions are used in the industrial production to optimise processes and to avoid redundancies. The scientists K. Hales and M. Lavery define the term workflow management software as follows:

"Workflow management software is a proactive computer system which manages the flow of work among participants, according to a defined procedure consisting of a number of tasks. It co-ordinates user and system participants, together with the appropriate data resource which may be accessible directly by the system or off-line, to achieve defined objectives by the set deadlines. The co-ordination involves passing tasks from participant to participant in correct sequence, ensuring that all fulfil their required contributions, taking default actions when necessary."

Unfortunately, the concept is not completely transferable to the creation of learning applications as the content is being focused on and not the process itself. That is why, so called content management systems are being applied to produce E-learning contents. (Fig. 4)

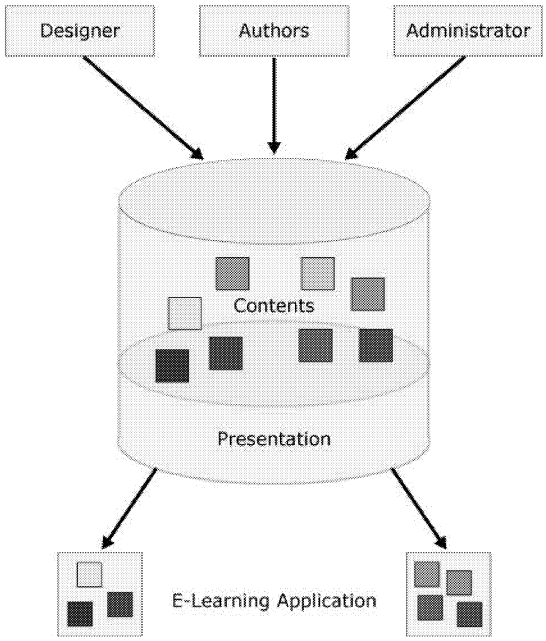


Figure 4: Content Management for E-Learning Applications

A classically programmed WBT application is static. It is only once produced and you have to intervene directly into the source code if there are any textual changes required. Apart from changing contents, changes of the layout are also necessary in order to adapt these applications to new data and conditions. Using dynamic Internet presences is preferred to avoid this disadvantage regarding WBTs with high change frequencies.

The contents and the layout of content management systems and a dynamic Internet presence are standing separated. The content, e.g. texts, pictures, animations as well as information about the structure, is located in an own data base. As general basis act data storages, e.g. XML or data bases, cooperating via interfaces with the Web server in order to provide and publish saved contents. The finalisation of content and layout just follows the call of the learning application whereas the content is provided via queries of data being embedded with the help of special script programming languages into a specific presentation medium for instance HTML. The above mentioned separation of content and layout assures various advantages. Via interfaces content management systems offer the possibility to change contents dynamically and to extend them. Thus, authors do not need to know anything about programming and layout designing. The complete content is produced automatically by the system. Furthermore, it is possible to change the layout or even the presentation format. On the other hand the user gets a general survey of the completely included knowledge because of the data management of all contents in content

management systems. Consequently, the user can compose individually the own learning application depending on the learning situation.

MARKET ACCESS AND PERSPECTIVES

The application of the produced content modules takes place in different types of education, but with regard to the academic target group exclusive for higher education. Therefore, the special marketing concept was developed. It describes in detail the target groups, market situation, the business models, the application objectives, the development and application perspectives, etc. There are also the main fields of application characterised as undergraduate and graduate studies, class room instruction and distance learning, primary and further education and training.

Recently, the content is well structured and modularised such as a construction set. Therefore, the instructor is able to generate individual learning programs by separate modules. This is sufficient for the proprietary application by the mentioned competence cells as well as by their business partners. The further decomposition and administration by the content management system will improve the ability for more individualisation of the offered contents being the precondition for offering and delivering content elements or content sub modules to competence cells of new competence networks. The recent state of the art in combination with the planned modifications permits the access to each common learning platform and corporation in the following competence networks and educational programs:

- Further education in media competence for SME
- Special programs for regional knowledge transfer for SME
- Bi- and multilateral corporation programs with business partners
- Part of online BA and MA programs in university networks
- Part of an international MBA university program
- Part of undergraduate and graduate university courses
- Part of an educational market place such as of Saxony
- Part of a Saxon educational portal as university platform for e-learning.

Due to the characteristics of the competence cells, the above mentioned educational units are involved in competence networks of developer's applicants in the regional, national and international context.

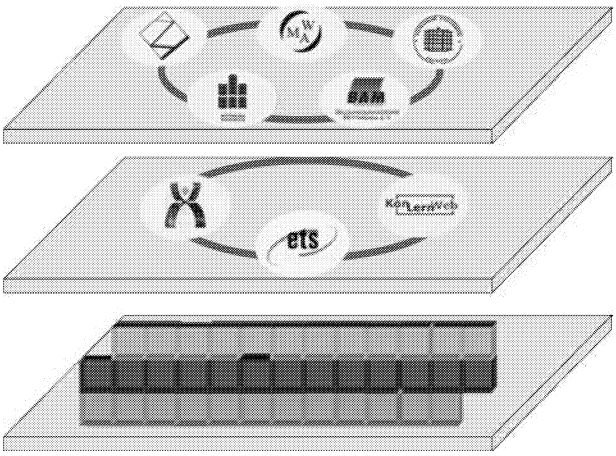


Figure 5: Content Modules, E-learning Platforms and Competence Cells in their Network

The main objectives for the future are:

- Further standardisation of the content description and process interfaces
- Development of new content management systems for the individualisation of educational programs
- Completion, modification and actualisation of contents
- Extension of competence networks by improving the competence of the cells
- Adoption of the surfaces and the design to different national educational standards
- Methodology for individualised mixed applications in blended learning.

The future for tele-educational competence cells will be in the corporation in complex networks of competence.

SUMMARY

Competence cells are the elements for competence networks. The successful development in tele-education and blended learning will be focused on the extension of these networks based on growing competence of the educational cells.

The knowledge transfer will be promoted by new methods, contents and software in relation to the web-based education and training. Therefore, new kinds of content structures, supply services and applications will be involved in tele-education and blended learning.

There are a lot of partial solutions already in practice. The best practice applications are used to advance the competence of educational cells and of integrated networks in general. The main objective is the improvement of the learner situation and the educational processes.

REFERENCES

Baumgartner, P.; Häfele, H. and K. Maier-Häfele. 2002. „E-Learning Praxishandbuch, Auswahl von Lernplattformen". Studienverlag Innsbruck.

Müller, E. and R. Riedel. 2003. "Planning, Scheduling and Producing in non-hierarchical Competence Networks". In *5th*

International Workshop on Human Factors. Proceedings. Limerick, Ireland.

Hales, K. and M. Lavery. 1991. Workflow Management Software. Ovum Report, London.

Schumann, Chr.-A. 1996. "Computer simulated Plant". In *European Simulation Symposium*. Proceedings. Genova, Italy.

Schumann, Chr.-A. et al. 1998. "Cross-discipline co-operation in education with international and industrial support". In *90th Anniversary Jubilee Seminar on Engineering Education*. UNESCO International Centre for Engineering Education (UICEE) & The University of Wismar. Proceedings, Wismar, Germany.

Schumann, Chr.-A. et al. 1998. "Distance Education as Part of modular-design system for lifelong learning". In *XV. IFIP World Computer Congress*. Proceedings. Vienna/Budapest, Austria/Hungary.

Schumann, Chr.-A. 1999. "Development and Using of Regional Network for Open Learning and Distance Education". In *19th ICDE World Conference on Open Learning and Distance Education*. Proceedings. Vienna, Austria.

Schumann, Chr.-A. et al. 1999. "Hypermedia Based Distance Learning Systems and their Applications". In *The Eight International Conference Information Systems Development ISD'99*. Proceedings. Boise, Idaho, USA.

Schumann, Chr.-A. et al. 2000. "Closing the Gap between School, University, and Further Education Systems in the process of lifelong learning". In *EDEN - European Distance Education Network: Open Classroom Conference*. Proceedings. Barcelona, Spain.

Schumann, Chr.-A. et al. 2000. "Humans and their Images in the Real and Virtual World of Learning, Training, and Working". In *The Wanderstudent 2000, International Colloquium*. Proceedings. KU Leuven, Belgium.

Schumann, Chr.-A. et al. 2001. "Improvement of Supply Chain Management by Knowledge Management in the Automotive Industry". In *ATT by SAE*. Proceedings. Barcelona, Spain.

Schumann, Chr.-A. et al. 2002. „Integration des Wissensmanagements in das Supply Chain Management am Beispiel der Automobilindustrie, Modelle im E-Business". In *4. Paderborner Frühjahrstagung, Fraunhofer-Anwendungszentrum Logistikorientierte Betriebswirtschaft*. Proceedings. Paderborn, Germany.

Schumann, Chr.-A. et al. 2002. "Impact of the Reorganisation of the Enterprise Information System by Knowledge Management Methods and Tools on the Supply Chain". In *SAE Automotive & Transportation Technology Congress*. Proceedings. Paris, France.

Schumann, Chr.-A. et al. 2003. „Regionaler Wissenstransfer", In *LEARNTEC2003*. Proceedings. Karlsruhe, Germany.

Schumann, Chr.-A. et al. 2003. "Process Support and Quality in WBT Content Development". In *EDEN Conference*. Proceedings. Rhodos, Greece.

Schumann, Chr.-A. et al. 2003. "Media Competence for the Management (SME)", *EDMAN 2003*. Proceedings. Brno, Czech Republic.

E-LEARNING TOOLS

IMPROVEMENT OF AN E-LEARNING PLATFORM THROUGH THE ANALYSIS OF USAGE PATTERNS

Roberto Cappuccio
Francesca Di Bono
Alberto Sillitti
Giancarlo Succi
Center for Applied Software Engineering
Free University of Bolzano-Bozen
39100 Bolzano
Italy
{rcappuccio, fdibono, asillitti, gsucci}@unibz.it

KEYWORDS
eLearning, metrics collection, GQM

ABSTRACT

The market for remote learning is one of the most competitive markets in the IT industry and it is rapidly reaching maturity. Many organizations have moved out of the start-up phase of e-learning, characterized by the evaluation of available tools, and have become mature and demanding practitioners.
In such a situation, in order to reach and to maintain supremacy, competitors have to constantly improve their products, adding new features to satisfy the current needs of the customers or to anticipate their evolution.
This kind of constant adherence to the needs of the customers can only be accomplished through the continuous and exhaustive analysis of usage patterns gathered from currently used tools.
In order to be effective, the collection of the metrics has to be performed automatically and transparently. Then, special statistical analysis methods should be developed in order to get relevant information out of the collected data.

INTRODUCTION

Constant focus on user has been the leading paradigm during the development phase of the Teleacademy project and the experience gathered gives now the opportunity to analyze and discuss this approach.
The Teleacademy project has been partially funded by the European Social Fund (ESF) and developed by the Center for Applied Software Engineering (CASE) of the Free University of Bolzano/Bozen in collaboration with the Business Innovation Center - Alto Adige (BIC).
The main goal of the project consists in the development of a platform enabling distance learning and the gathering of metrics regarding usage patterns.
Recent surveys, presented in the following sections of this paper, demonstrate that a negative perception of the overall quality of e-learning solutions is still dominant among the users despite of the huge efforts made by the LMS (Learning Management System) suppliers.
The main reason for this disappointment is the awareness reached by the users about their needs and requirements which are not completely satisfied by actual solutions.
A user-driven approach in the development of new solutions for e-learning could drive the efforts of the suppliers in the directions demanded by the users.
The paper is organized as follows: section 2 analyzes the state of the art; section 3 describes the user centric approach; finally, section 4 draws the conclusion.

STATE OF THE ART

As mentioned before, the market for remote learning is rapidly reaching its maturity.
After a first phase characterized by the suppliers' attempt to impose their solutions focused only on technology, the market shows that roles have been inverted: now it is the user that demands solutions tailored on its real needs.
In the summer of 2003, brandon-hall.com conducted a research in order to evaluate the state of the e-learning industry. This research, conducted through online surveys and telephone interviews, gave the results summarized in Table 1 (Nantel 2003).

Table 1: Summary of the Results of Brandon-hall.com Research

6.8%	We're in the planning stages of implementing e-learning in our organization.
12.62%	We're testing the e-learning waters by providing a few e-learning courses to our learner community. These may be off-the-shelf courses or we may be converting some conventional training content to e-learning. We do not use a learning management system (LMS).
38.83%	We're providing either custom-built or off-the-shelf, self-paced e-learning courses and tracking results using a learning management system.
26.70%	We're providing self-paced e-learning courses, tracking results using a learning management system, and

	providing live e-learning sessions using virtual classroom applications.
15.05%	We're providing self-paced e-learning courses, tracking results using a learning management system, and providing live e-learning sessions using virtual classroom applications. In addition, we are using a content repository such as a learning content management system (LCMS) to manage our content.

The report "Quality and eLearning in Europe" published by Bizmedia (Massy 2002), based on a survey conducted in five European languages (English, French, German, Spanish, and Italian) in April 2002, shows that 61% of all respondents rated the overall quality of eLearning negatively - as 'fair' or 'poor', in particular ´fair´ and ´poor´ rating were given by 59% of the EU public sector respondents and 72% of the EU private sector respondents.
The most important criteria for evaluating quality in eLearning are in order of priority:

- Functions technically without problems across all users
- Has clearly explicit pedagogical design principles appropriate to learner type, needs and context
- Subject content is state of the art and maintained up-to-date
- Has a high level of interactivity

A very small number of respondents from only three countries gave an excellent rating. These were Germany, Ireland and UK. A somewhat larger number of respondents from nine countries rated overall quality 'very good'.
At the bottom end, respondents from 10 EU countries included overall ratings of 'poor', and 50% or more of respondents from the UK, Italy, and Germany rated overall quality only 'fair'.
75% of French respondents gave responses of 'fair' or 'poor' to overall quality.
Only 2 German respondents gave a rating of 'excellent' and 'very good' (one for each), although nearly 28% rated overall quality as 'good'. Similar to total respondents rating overall quality 'poor' or 'fair', 61% of Germans also gave this rating.
Similar to France, over 75% of Italian respondents rated overall quality as 'poor' or 'fair' and there were no ratings of 'excellent' or 'very good'.
Spanish respondents also gave very negative ratings with 74% stating 'poor' or 'fair'.
More positively, the 'poor' or 'fair' ratings in the UK came from 57%, with 12% giving a 'very good' or 'excellent' and 28% giving a rating of 'good'.

From the suppliers' perspective, one thing is immediately evident: the great number of solutions thought for solving various needs related to distance learning. Many of them are really excellent but often in specific fields only: compliance to international standards for the delivery of courseware, availability of tools enabling virtual reality, automatic collection of usage pattern metrics and their statistical analysis, availability for free and elimination of the classical temporal and physical barriers that hinder the adoption of a life-long learning style.
Given the results of many comparative evaluations of the available tools, no one of them offers all these features at the same time.

A detailed comparison of the features offered by the chosen web-based learning systems is the starting point to take advantage from the existing products and speed up development.
For this purpose, the following research is based on the outcomes provided by Edutech published in March 2003 in the Swiss Virtual Campus (SVC) Platform Evaluation Report (Edutech 2003).
The criteria adopted for the selection of the tools to be evaluated are the following:

- Previous usage of the tool by the Free University of Bolzano/Bozen (Faculty of Applied Computer Science)
- Popularity of the tool and its acceptance and usage by higher education institutions

According to these criteria, the selected eLearning systems are: WebCT Vista 1.2, Teleacademy (developed by the CASE), Clix 5.0, Blackboard ML, IBT Server 6.1, Qualilearning/Luvit 3.5 and Globalteach.
No one of the selected platforms includes all the tools required for a complete remote learning solution.
As the SVC Platform Evaluation Report states, "the evaluation shows that every product does have its specific strengths, where it outdoes most of its competitors, whereas it is lacking in other aspects. A direct comparison is therefore nearly impossible".

Table 2: Summary of the Strengths and Weaknesses of the evaluated products (Edutech 2003)

WebCT Vista 1.2	
Strengths	**Weaknesses**
• Powerful communication and student tools • Efficient file handling due to integrated file manager with WebDAV support • Numerous flexible authentication models	• Cannot copy-paste URLs, no browser bookmarks, browser's "Back" and "Reload" buttons don not work • incomplete font scaling, difficult printing

supporte • Large institutions and consortia can hierarchically organize groups, courses, sections... • Good documentation	• Limited SDK- limited extensibility • Limited layout control • Incomplete or missing support for eLearning specifications
Blackboard ML	
Strengths	**Weaknesses**
• Clean, easy-to-use interface • Powerful "virtual classroom" tool • Good possibilities for interoperating with other systems • Extensible system through building blocks program • Good documentation	• Limited customizability of look and feel • No internal resource or file manager • Frame based display
Clix 5.0	
Strengths	**Weaknesses**
• Large palette of tools • Good support for external content • Syllabus/learning plan with branching options • Powerful rights management system • "Mandaten" concept (one installation for several units with their own courses)	• Creating a course is complex process • No hierarchical content structure • No FTP or WebDAV support • No search functions in contents • Limited support for eLearning specifications
Globalteach	
Strengths	**Weaknesses**
• Excellent support for SCORM 1.2 (debugger included) • Efficient and flexible customization possibilities • Extensible framework with various documented APIs • Complete technical documentation	• Authoring tool is a Window application • Administration tool is a Window application • Server runs on MS.net infrastructure only
IBT Server 6.1	
Strengths	**Weaknesses**
• Full XML/XSL support • Support for SCORM compliant learning modules • The entire system can be fully customized	• Many features are only available through programming • Difficult to use for course designers • System cannot really used out-of-the-box
• Server runs on all operating systems with Java Virtual Machine • Good multi-language support • Clean and modern technical design • Modular architecture • Runs on all modern browsers	• Running/customizing the server requires skilled and experienced staff • Complex system, difficult to get used to it • Small user base in universities
Qualilearning Luvit 3.5	
Strengths	**Weaknesses**
• Nice menu based interface • Some interesting didactical functions (e.g. feedbacks on each document) • Powerful rights systems • Good statistics on course usage • Support for some e-learning specifications	• Frame based display • Lack of browser toolbars • No WebDAV or ftp support for file upload • No search function • No info about problems with zip archives, limited online support and system extensibility
Teleacademy	
Strengths	**Weaknesses**
• Global and Faculty news/events tool • Free • Nice, clean, easy-to-use interface • Useful additional forums (buying/selling, accommodations, car trips, get for less, job ads, living in South Tyrol)	• No course management (contains only lecture's videos) • No fundamental tools: quiz editor, calendar, document archive, grade-book • Limited support for eLearning specifications

THE FOCUS ON USER

The approach adopted for the assessment of the effectiveness of the Teleacademy e-learning platform is the Goal Question Metric paradigm (Basili et a.), a method to guide the definition and exploitation of a goal driven measurement program used in Software Engineering as a technique for the creation of a set of metrics in such a way that:

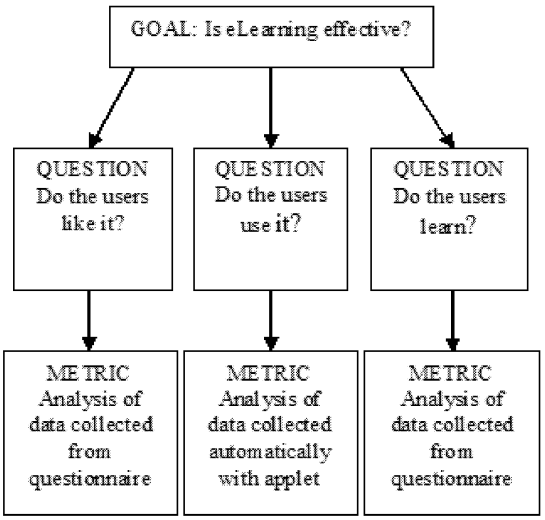
- Resulting metrics are tailored to the organization and its goal.
- Resulting measurement data play a constructive and instructive role in the organization.
- Metrics and their interpretation reflect the values and the viewpoints of the different groups affected (e.g., developers, users, operators).

GQM defines a measurement model on three levels:

- Conceptual level (goal): A goal is defined for an object, for a variety of reasons, with respect to various models of quality, from various points of view, and relative to a particular environment.
- Operational level (question): A set of questions is used to define models of the object of study and then focuses on that object to characterize the assessment or achievement of a specific goal.
- Quantitative level (metric): A set of metrics, based on the models, is associated with every question in order to answer it in a measurable way.

For the Teleacademy Project, the resulting GQM is the one summarized in Figures 1.

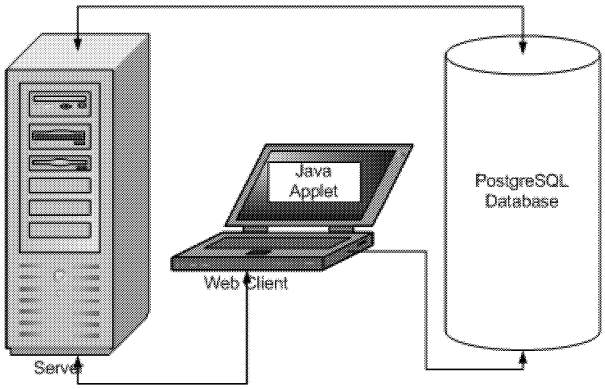
Figures 1: Summary of Teleacademy Projects GQM



As shown in the GQM table above, two kinds of questions are formulated: qualitative and quantitative.
The collection of answers to qualitative questions is achieved through questionnaires distributed to a sample of the system's users.
The collection of data regarding quantitative questions has to be accomplished automatically and transparently in order to be perceived as less invasive as possible.

The chosen technique to gather data automatically from Teleacademy consists in inserting a lightweight Java applet in every page sending records to an external database.
This differs completely from the approach adopted in the previous version of the Teleacademy Portal in which it was responsibility of the web server to keep track of user's activity.
This first approach presents the drawback of not keeping track of user's changes of focus, that is: it does not notices when a user keeps a page open but works on another application (such as a word processor, a spreadsheet, etc.).
Through the use of an "invisible" applet on every page a higher granularity of tracking can be accomplished without user's notice.
Figures 2 shows a simplified picture of the implementation framework including the Java Applet running on client.

Figures 2: Activity Tracking: Implementation Framework



Information gathered through the analysis of collected metrics, other than being used for the evaluation of the effectiveness of the platform, accomplishes another fundamental goal: give the developers the opportunity to improve the system.
Through the analysis of quantitative data, it is possible to adapt the platform to the real user needs:

- Improve usability of the interface
- Allow personalization of the user experience
- Improve interaction between tools
- Eliminate superfluous tools
- Improve the most used tools
- Suggest best practices (based on the analysis of usage patterns adopted by successful students)

Through the questionnaires is possible to obtain a different kind of information (of qualitative nature), such as:

- Need for a new tool
- Need for new features in an available tool
- Need for improved interface
- Need for new media supports
- Need for different instructional design

CONCLUSION

The goal of the Teleacademy project is the development of a customizable solution for remote learning and to gather metrics related to its use.
In order to build a complete platform, the most popular commercial applications have been analyzed and their features have been compared. The result is a list of

requirement that will be fulfilled during the development phase.
This is in any case only a starting point. Further development of the platform will be completely user-driven, that is, the adopted methodology should give the possibility to constantly adapt the system to the real needs of the users which are evolving continuously.

ACKNOWLEDGEMENT

The project has been partially supported by the European Social Fund. Special thanks to Barbara Repetto, Head of the ESF office in the Province of South Tyrol.

REFERENCES

Basili, V.R., G. Caldiera and H.D. Rombaci. "The Goal Question Metric Approach." Available at: http://wwwagse.informatik.unikl.de/pubs/repository/basili94b/encyclo.gqm.pdf

Berk J. 2003. "Learning Measurement: It's Not How Much You Train, But How Well." *The Elearning Develper's Journal*, (Nov), 1-8

Edutech. 2003. "Evalation Reports." (Jan) Available at: https://www.edutech.ch/edutech/tools/ev2.php

Edutech. 2003. "SVC Platform Evaluation Report." (March) Available at: http://www.edutech.ch/edutech/tools/ev2.php

Lockee, B., M. Moore and J. Burton. 2002. "Measuring Success: Evaluation Strategies For Distance Education." *Educase Quarterly*, No.1 (Jan), 20-26.

Massy J. 2002. "Quality and eLearning in Europe." *Bizmedia 2002*.

Nantel R. et a. Novembre 2003. *Athoring Tools 2004. A Buyer's Guide to the Best E-Learning Content Development Applications.* Brandon-hall.com

BIOGRAPHY

ROBERTO CAPPUCCIO was born in Bolzano (Italy) in 1967. From 1991 to 2001 he worked as an independent software consultant and developer.
Since 2002 he works for the Center for Applied Software Engineering (CASE) at the Free University of Bolzano where he leads the development team of the Teleacademy project.
In 2004 he will obtain his bachelor in Applied Computer Science from the Free University of Bolzano.
His current research interests are distance learning and data mining.

AN INTRODUCTORY STUDY ON DEVELOPMENT OF A SYLLABUS GENERATOR

Zeynel Cebeci
Department of Informatics
Çukurova University
01330 Adana, Turkey
E-Mail: zcebeci@cukurova.edu.tr

Fuat Budak
Department of Environmental Engineering
Fac. of Arch. & Engineering, Çukurova University
01330 Adana, Turkey
E-Mail: fbudak@cukurova.edu.tr

KEYWORDS
Syllabus, XML, ASP, Learning and teaching software, Authoring Tools

ABSTRACT

Syllabus preparation is one of the most required tasks performed in schools. In academic institutions a course syllabus is considered as a basis for a common understanding between instructor and students. Syllabus gives initial and working information such as name, credits, prerequisites, instructors, materials, policies, objectives, schedule of the course to students.

Generally, academic staff builds a variety of different syllabus documents which are traditionally written according to a template given by administrations of the schools. Unfortunately this kind of work is time-consuming. Furthermore, it results with a high cost in collecting, printing and delivering the syllabi. In case of using the online systems and tools will dramatically reduce the physical works and the costs in syllabus preparation and delivering.

In this study, a syllabus generator called as "syllabuserXML Engine" (or shortly "SXE") which was developed with use of core technologies such as Active Server Pages and XML is introduced and discussed in aspects of efficiency and usefulness in preparing course syllabi. SXE can be used for producing and publishing the online versions of syllabi.

INTRODUCTION

Generally, instructors give the information about course to each student on first day of class. It is very important for clarifying the points needed by students. A syllabus should communicate to students what the course is about, why the course is taught, where it is going, and what will be required of the students for them to complete the course with a passing grade (Altman and Cashin 1992).

According to search results performed on some search engines i.e., Google, Altavista and with "Syllabus Finder" by Cohen (2003), there is a remarkable large variation among the syllabi that were published by the institutions around the world. The reason of these variations is due to lack of specifications for syllabus preparation and inadequate number of the tools available online, additionally. In fact, according to search results made on the Web, it was obtained that there is a limited number of tools and/or solution publicly available on the Internet. A prime study to develop a working environment was carried out by Black (2002) with use of his syML language for syllabus preparing. Another approach offered by Cebeci and Budak (2004_a) was used for building XML-based tool content in their study. In addition to these, there are intensive efforts to use XML-based syllabus generators in some programs, individually.

In order to contribute to learning tools, this study aims to introduce an engine that enables syllabus building in XML by use of Microsoft's (MS) Active Server Pages (ASP) technology.

ARCHITECTURE AND WORKING METHOD

In order to set up SXE, distribution package *(sxe.zip)* should be extracted on an IIS-compliant Web server. Then the directories listed in Table 1 must be created and the contents should be copied into relevant directory. Then, *config.xml* file must be edited for user environment and Web server. To configure the configuration file *config.asp* script can be used, or editing can be done with any text editor manually.

Table 1: Directory Structure of SXE

Directory	Use	Files
engine	Hosts the engine scripts and files	All script and configuration files
working	Temporarily hosts htm and xml files for syllabus	Empty at begin
repository	Hosts all syllabi files permanently	syllabi.xml sview.xsl all htm and xml files for each syllabus

SXE consists of engine scripts, configuration files, XSLT view templates, style sheets, picture files listed in Table 2. The functional model of SXE is also illustrated in Figure 1.

SyllabuserXML Engine (SXE) was based on xml-formatted syllabus model described by Cebeci and Budak (2004), and a pionering tool developed by Cebeci and Budak (2004).

The main component of SXE is syllabi manager (SM) that performs building new syllabus documents and transferring them to syllabi repository. SM also organizes for other necessary processes such as viewing, updating and deleting the existing syllabi in repository.

Table 2: Components of SXE

Component	Function	Files in Group
Syllabi manager	to build, transfer, copy, update, list, view, delete and synchronize the syllabus files	*smanager.asp* *sbuild_ui.asp* *sbuilder.asp* *slist.asp* *sviewer.asp* *supdate.asp* *sdelete.asp* *stransfer.asp* *ssyncronize.asp* *scopy.asp*
Configuration	to set up directories, and language ans server details	*sconfigure.asp*
Templates	syllabus viewer template in XSLT	*sview.xsl* *syllabi.xsl*
Predefined data and images	files for repository, configuration, formatting, and emblem	*syllabi.xml* *config.xml* *syllabuser.css* *logo.jpg*

Configuration component is used to enter the name and properties of the server in which SXE will be set up, and which directories will be used for hosting the engine scripts, working files and syllabi repository. The elements necessary to determine these details are entered into *config.xml* file located in engine directory. In order to use SXE, first, this file should be configured correctly.

Third component includes only one file named *sview.xsl.* This provides a predefined student view of syllabi documents in XSLT. This file can be recoded for a desired view for printing course catalogs, guides etc.

Finally, build-in files are *config.xml*, *syllabuser.css* and *logo.jpg*. *Config.xml* contains the relevant information about server and directories which will be used by SXE. *Syllabuser.css* file includes the cascading style sheet definitions for using with SM. *Logo.jpg* is an institutional emblem image which can be replaced with users' chosen image with same name.

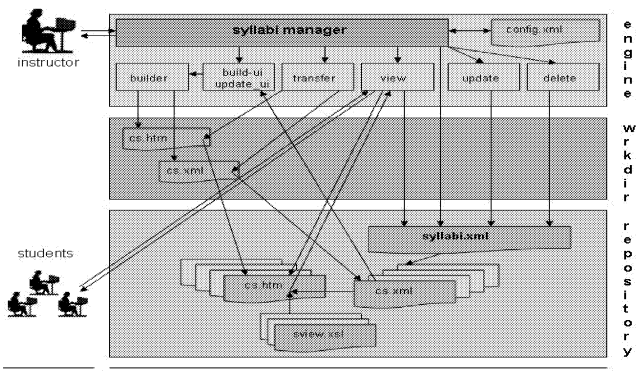


Figure 1: Functional Model of SXE

In SXE to organize syllabi SM (*smanager.asp*) is called on the server. This file is located in engine directory and URL of engine depends on the path defined by system administrator. When SM is called by instructors it displays number and title of the courses in repository like in Figure 2. SM uses *syllabi.xml* file located in repository directory of SXE. In each line of display, on the right side, the process buttons for add, copy, update, view, delete and transfer the existing and new syllabi.

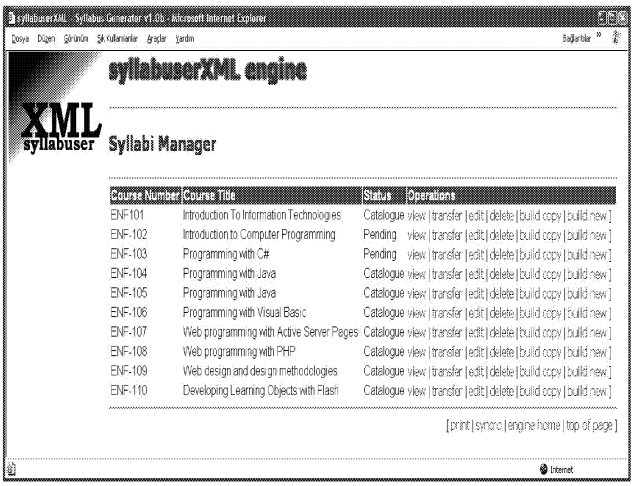


Figure 2: An Example Screenshot for SM View

Syllabus builder of SXE mainly consists of two parts: User interface and builder. When the instructor requests for building a new syllabus, SM sends a Web form (*sbuild_ui.asp*) to client browser. When all information is filled in form and pressed "Build Syllabus" button, SM's syllabus builder script (*sbuilder.asp*) is run automatically. The builder creates two files named as *cs.htm* and *cs.xml* which contains syllabus display codes and syllabus xml information, respectively. Here, "cs" stands for the file name for a specific course which is replaced with the relevant course number entered in syllabus build interface by instructors, i.e., for the course number *enf106,* the produced files will be *enf106.htm* and *enf106.xml.* The file with *xml* extension contains the syllabus information in XML format which is structured according to suggested model by Cebeci and Budak (2004_b). The file with *htm* extension contains codes for calling the relevant *xml* file and *xsl* file (view template), and showing the syllabus information to the users.

Sometimes, copying an existing syllabus may be required to reproduce a new file which has similar syllabus information with the existing one. This could make easier to create a new file when compared to creating from scratch. Thus, copy option of SM can be applied to create a starting file for only editing the different information to be filled in.

All *htm* and *xml* files produced for each course syllabus are temporarily stored in a working directory defined in *config.xml.* These files are marked with a "pending" status tag in SM view. In order to publish these syllabi to the students they should be transferred to the repository

directory defined in *config.xml*. For this task, transfer option is run by clicking "transfer" on SM view, and then transferring tool (*stransfer.asp*) will move the syllabus files (*cs.htm* and *cs.xml*) of the course selected from working directory to repository directory, permanently. The syllabi files which moved into repository directory are shown with a "in Catalog" mark on SM view.

In order to view a syllabus listed on SM view, the students can call syllabus viewer (*sviewer.asp*) of the engine, SXE. In this case the relevant *htm* file is called from repository and the information in its associated *xml* file sent to the students. *Htm* files created for each syllabus contains the codes to parse *xml* file according to template and rules defined in viewer template in *sview.xsl* file located in repository directory.

In order to update a syllabus, the update script is used. This is a part of SXE which calls the relevant *xml* file from repository, and parses and passes syllabus information to user interface (*update_ui.asp*) for performing necessary updates by instructors. Once the information is updated, this user interface calls the builder script of SXE to save the edits to working directory, again.

Delete option in SM is used to remove any syllabus from repository. When this process is invoked, it deletes *htm* and *xml* file associated with the selected course.

CONCLUSION

SXE was developed as freeware syllabus generator, so that might be hoped that it will contribute to prepare course syllabi efficiently and quickly. The engine is still under beta test and constantly under development. For benefiting it, the engine can be accessed and experienced at the URL: *http://cebeciz.cukurova.edu.tr/tools/syllabuserxml.*However, currently, SXE is working only in English, its multilingual version will be worked in near future.

Since SXE was developed fully in ASP and XML technologies, currently it can be used only in MS Windows platforms with Internet Information Server software installed. If the scripts are converted to other scripting languages like PHP, Perl or Python it can be easily benefited for other platforms such as Linux because the files and architecture are based on XML format.

Creating syllabi files in XML format with SXE enables to move them for processing by different systems and applications. This interoperable working style provides that sharable syllabi documents can be re-used for various purposes and for different views.

REFERENCES

Altman, B.H & W.E. Cashin. 1992. "Writing a syllabus". *Idea Paper No 27*, Center for Faculty Evaluation & Development. Kansas State University.

Black, D.A. 2002. *SyML - the Syllabus Markup Language* (accessed at http://icarus.shu.edu/dblack/SyML/ on 16th Jan, 2004).

Cebeci, Z. and F. Budak. 2004a. "A tool to create XML-Based Course Syllabus Documents". (unpublished, submitted to *Computers & Education* on Jan 2004) .

Cebeci, Z. & F.Budak. 2004b. "An Approach to Structuring a Syllabus Document in XML. (unpublished, submitted to *Journal of British Educational Technology* on Jan 2004) .

Cohen, D. 2003. *Syllabus Finder.* (accessed at http://chnm.gmu.edu/tools/syllabi/ on 15th Jan, 2004). Center for History and New Media, George Mason Univ., VA, USA.

AUTHOR BIOGRAPHY

ZEYNEL CEBECI was born in Giresun province located in Black Sea region of Turkey in 1960. He completed his secondary education in High School of Teacher Education for Primary Schools in Trabzon province in 1978. Then he was graduated from Faculty of Agriculture (FAg) of the Cukurova University (CU), Adana-Turkey in 1983. He completed his M.Sc. and Ph.D. in Biometry and Genetics Unit at the Institute of Basic and Applied Sciences (IBAS) of CU in 1985 and 1990, respectively. He has worked as software developer, research assistant and assistant professor in private businesses and CU from 1985 to 1999. He was awarded and assigned as a full professor in 1999 at CU. Currently he works an academic staff member in the Biometry and Genetics Unit of FAg and IBAS, and teaches courses on information and communication technologies. In addition, he has been working as the director of the Computer Research and Application Center of CU since 2000. He is also head of the Department of Informatics of CU for 2 years. Dr. Cebeci studies on programming languages, software development and applications on genetics (computational biology), statistics and biometrical methods and algorithms. Recently, he carries out researches and implementation works on e-learning systems and technology.

FUAT BUDAK was born in Bitlis province located in Eastern Anatolia Region of Turkey in 1961. He was graduated from Department of Soil Sciences of Faculty of Agriculture (FAg) at the Cukurova University (CU), Adana-Turkey in 1983. He completed his M.Sc. and Ph.D. in Agricultural Economics Division at the Institute of Basic and Applied Sciences (IBAS). Then he also studied MBA program at the Franklin University, Ohio-US. Currently he works an academic staff in the Department of Environmental Engineering, and teaches courses on information and communication technologies and environmental economy. He studies on software development and applications for environmental sciences. Currently, he also involved in research works on e-learning systems and technology at CU.

KNOWLEDGE IN DISTANCE LEARNING SYSTEM

Malgorzata Nycz and Barbara Smok
Wroclaw University of Economics
53-345 Wroclaw, ul.Komandorska 118/120,
Poland
E-mail: {malgorzata.nycz,barbara.smok}@ae.wroc.pl

KEYWORDS
Knowledge, knowledge management, distance learning, CAI

ABSTRACT

The paper presents teaching/learning models, distance education environment characteristics, knowledge assets in distance learning system, knowledge management in distance learning. At the end of paper some aspects pro and contra distance learning have been shown.

INTRODUCTION

Nowadays knowledge is treated as a one of the most valuable assets and its value will be increasing. New information and communication technologies offer better, faster and more powerful tools and techniques and from the other hand these new possibilities create new needs. It is impossible to possess all necessary and required knowledge but it must be easy and fast to find what is needed. When realizing the didactic process students and teachers are not at the same place, they communicate through internet, students can follow courses accessible in the net, they learn in their own tempo and when they want, etc.

TEACHING/LEARNING MODELS

In traditional education, students and teachers are at the same place and at the same time. Students listen to the teacher when he/she carries on the lesson/lecture. Teacher is the source of knowledge delivered to students. When digesting the didactic material students can use additional means as manuals, books. Teacher makes the assessments of students. The quality of education process depends on the teacher's quality, his knowledge, personal character feature and ability to teach others. In distance learning system students and teachers are geographically separated, they communicate through internet, students have unlimited access to didactic materials, they can learn in their own tempo; learning time also depends on student preferences. They decide when are ready to pass the exam from given subject (Nycz and Smok 2003), (Nycz 2002). The gravity point has been moved from a teacher to the system. Teacher is rather as the instructor for students, theirs buddy but not a teacher in traditional model. Assessment of students is realized by both teacher and system, but the teacher is responsible for final results. The main differences between models of education has been shown in fig.1.

Model / Feature	Traditional model of education	Distance education system
Main knowledge source	teacher	knowledge bases in education system, any knowledge sources accessed via internet
Additional knowledge	books, manuals, audio and video materials	traditional sources, teacher
Assessment	only by teacher	system and teacher who is responsible for final assessment
Quality of education	dependents on teacher's quality, his level of knowledge, his ability to share his knowledge	depends on electronic knowledge sources quality and other didactic materials

Fig. 1. Types of education model

Building of knowledge in student's mind is a process of problem solving that makes student to be active, innovative and developing his experiences. In distance learning student can be active in building knowledge within the following cooperation forms: common learning within the team, interactive process of group building of knowledge, active participation in generation and selection of information, knowledge construction in context of other students points of view. The teacher's task is to be a supervisor of learning process and to monitor the progress in learning (Bielecki 2002). In distance education system knowledge resources are of very various form but mainly they are in form of modular didactic modules. They allow student to learn particular portion of didactic material. Passing to the next module may be dependent on whether he has positively completed the previous one. The same is with courses. Usually student can access three types of modules: modules covering material of particular subject, tests and exercises modules, help modules. The didactic process can be organized in various ways. The most common one is based on didactic paths (Piasta 2003). Student can be directed on a given path in result of e.g. enter test, that is often built in shape of decision tree. Directing on paths may also appear in result of poor results of tests after modules and student either has to return to a given module or extra explanations in form of help have to be presented to him. Within the repetition, student can be asked either to repeat the whole module or to go through additional explanations and maybe to repeat more difficult parts. Realization of auxiliary material covers large range of tasks: extra tasks to be solved, manuals and other books from a given discipline, possibility to ask by email colleagues or teacher for help. The realization of help materials is possible when either a student chooses the help option or he unsuccessfully tried several times to solve a test. In such a situation, the control may suggest additional explanations, exercises or reading appropriate books. The help option is under special supervising of control due to the fact that maybe not all suggested exercises have to be done. In that case student obtains the suggestion to stop the help mode and return to "normal" work or he decides to stop it (Peraya 2003). Obtained results are collected in the system. To enter the next course, the student has to pass the final test of the previous one. Sometimes the system enables the student to enter new course without requesting to pass final test of previous course. This situation occurs when the student wants only to read the material covered by the course, to check his knowledge from the course but not to be assessed by the system, only for him. The final assessment of the student is the teacher's task who can be supported by the system. This support can be realized e.g. either using simple mechanisms delivered by the system or using specialized external tools that infer about student's progresses in education process and generate a piece of advise about the final assessment.

THE DISTANCE EDUCATION ENVIRONMENT CHARACTERISTICS

The education environment can be seen from different points of view. The communication and application aspects have been those that are often taken into consideration (Ismangil 1998). In such an approach we distinguish the communication and the application layers. The first one can be understood as the communication and network infrastructure, which is essential to realize the distance education. The communication infrastructure can be realized in many different ways, e.g. using so-called agents or virtual classroom. The communication is in both directions between student and teacher. *Virtual classroom* can be defined as the education environment that expands traditional teaching/learning environment. It has been done not only by using distance, interactive teaching/learning but also by exploitation of multimedia via internet (Baborski and Nycz 1999). Student accesses into library files that are situated either in "its" school or collected in libraries outside school. The communication layer plays also a role within the didactic process: it enables the contact between student and between student and teacher when student sends its works for assessment or asks for help. *The agent technology* is used when solving problems by some objects called agents (Russel 1995). Application of this technology is still in progress, but some first achievements are very promising. Especially, when artificial intelligence aspects are introduced into realization of didactic process. In intelligent education environment, the agents communicate with each other, cooperate by exchanging messages, and student can access larger knowledge assets then in traditional learning process. The education environment may also be organized according to *hypertext* approach where the education aspect rather then technologic or communication has been underlined. Hypertext allows the student elastic navigation through the didactic material. Hypertext has been used for long time in didactic computer supported systems and can be a good input point when preparing modern system supporting education.

KNOWLEDGE ASSETS IN DISTANCE LEARNING SYSTEM

Knowledge can be defined as a set of information that enables drawing conclusions from premises (Baborski and Nycz 1999). Premises can be a situation description, set of facts, input conditions in dynamic model or other sets of information. A set of information being knowledge can have various forms. This way we came to difference between knowledge base and database. Database is a set of fact descriptions made using a data model. Knowledge base is database plus rules of inferring from data. Knowledge is a one of the most valuable assets in any didactic system. New information and communication technologies offer better, faster and more powerful tools and techniques, and these new possibilities create new needs. Knowledge can be seen as a kind of social activity or as a way of communication. The teacher's role has been changed. There are some approaches to knowledge modelling. Nonaka and Takeuchi (Nonaka 2000) distinguish two types of knowledge: tacit and explicit. Socialisation means that knowledge can be transmitted within personal contacts. Knowledge externalisation is the expression of tacit knowledge using commonly accepted and used forms. Hidden knowledge is also knowledge of domain expert; lack of this knowledge can make impossible solving many domain problems. Explicit knowledge is knowledge in shape of different process descriptions or suggestions about different ways of performing given tasks. This kind of knowledge can be easily codified. As it is said in (Nonaka 2000), new knowledge is generated when tacit and explicit knowledge influence each other. New knowledge creation is a repeated cycle presented in fig.2.

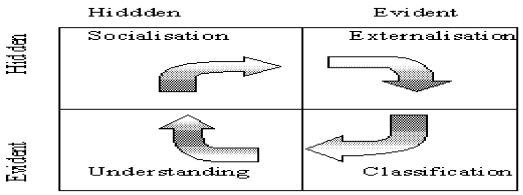


Fig.2. The Nonaka's model of knowledge
Source: (Nonaka 2000)

One of the most valuable assets of any didactic process is knowledge accessible in the system. It's correctness, actual and completeness implies the quality of this form of education as well as its further development. Knowledge, as any other assets in the system should be managed. Knowledge is collected in knowledge bases. It is of different nature and can be organized in various ways. Knowledge management in the education system can be understood as in any other system and covers such areas as knowledge acquisition, its maintenance, accessing, updating and verifying (Baborski and Nycz 1999). Knowledge should be actual, complete, certain, consistent as far as it is possible, etc. Knowledge comes from different sources. Among them one can distinguish: didactic courses and modules prepared by teachers, bibliography, books and manuals accessible in both traditional and electronic ways, didactic materials in form of audio and video, exchange of knowledge between students and/or institutions, etc. Knowledge covered by courses or other didactic units may quickly become not actual. The problem of knowledge updating and acquisition of new knowledge is very difficult and complicated. The methods of automated acquisition as well as methods of data mining are used when solving this problem. New knowledge must be verified either by teacher or by special verifying procedures.

KNOWLEDGE MANAGEMENT IN DISTANCE LEARNING

Distance learning management system allows creation, storing, updating and management of didactic material used within the education process (Koolen 2001). It realizes the approach that enables multiuse of didactic units or knowledge, offers collected modules and courses and should fit international standards (Peraya 2003), (Singh 2002). The education system bas been required: to collect only positively assessed didactic material, to enable building new course/units by teachers without programming by them, only using sheets accessible in the system. Database should be opened to be flexible and both teachers and students should accept formats of presented material (Hayhoe 2002). The main idea of the distance learning management system presents fig. 3.

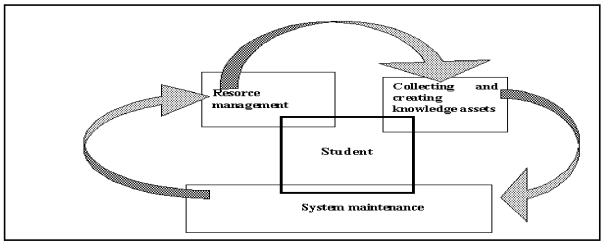


Fig.3. Idea of distance learning management system
Source: based on (Nycz 2003)

From the student's point of view it is not important how the education environment has been organized. He is interested only in interactive access to large resources of knowledge necessary within learning. In fig. 4 we present how such an environment can look like.

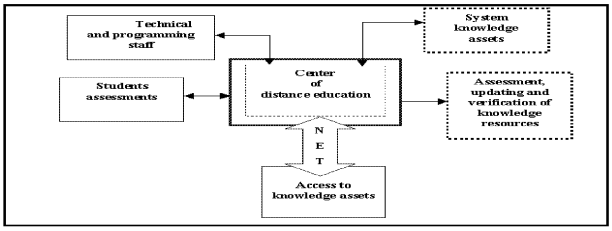


Fig.4. Center organization for distance learning
Source: [(Piasta 2003)

Knowledge management is not only the technological issue but also business and organisational one and is connected with organisation, its culture and values. On the other hand, technology is a tool that enables knowledge conservation, ordering and exploitation. Knowledge management is often connected with marketing of different products (in this case - didactic modules). But it is impossible either to buy a strategy of knowledge management or copy without understanding the assets and processes in a given organisation. In particular, implementation of concrete tools without analysis of possible to achieve business goals and technological and organisational barriers previously done are in longer perspective senseless.

DISTANCE EDUCATION PRO AND CONTRA

Distance learning can be seen in categories of advantages and civilization progress. Its positive aspects are following: the possibility of access to education for students that they would not have in traditional forms of education, larger choice of courses and accessible options in learning process, resources are not stored in one place, the lack of strict timetables of lessons and exams, student can learn in his own tempo and when he wants, many students can use the knowledge assets (didactic material), the possibility of in-line cooperation with other universities or institutions, fast communication among geographically dispersed students, students and teachers. As negatives of distance education can be treated the following aspects: both students and teachers must be familiar with computers and must have access to computers in network, negative impact on social connections between people due to the fact that students have approximately rare personal contacts with colleagues and teachers, some problems may occur until standards in communication language cannot be accepted. It depicts the language in which the courses have been prepared to be understandable for students from abroad, unsolved problems with intellectual properties of prepared didactic materials, distance education must be supported by appropriate institutions in the larger scale in unification of student assessment requests, regional distance education centers have to be organized not only within universities but also in larger scale. When looking at all pros and contras we observe that these contras should not have important impact on distance education development.

References

Baborski A and M. Nycz. 1999. "Nauczanie multimedialne a wiedza", in: Prace Naukowe AE im. O.Langego we Wrocławiu, Wrocław, 1999, in Polish

Bielecki W. T. 2002. „Założenia dla systemów e-lerning", in Polish

Hayhoe G.F. 2002. "Evaluating Distance Learning in Graduated Programs: Ensuring Rigorous, Rewarding Professional Education", http://www.puw.pl/elerning.html, from October 25, 2002

Ismangil B.P. 1008, "Towards Standardisation of Distributed Open Learning Environments", http://www.dcs.shef.ac.ukl~m6bpi/project/report.html#toc1 from August 5, 1998

Koolen R. 2001. "Learning Content Management Systems – the Second Wave of eLearning", Grand Rapids, Michigan, USA, 2001

Nonaka I. 2000. "Kreowanie wiedzy w organizacji", Poltext Warszawa 2000, in Polish

Nycz M. 2002. „Zarządzanie wiedzą w systemach nauczania otwartego", in: „Systemy wspomagania organizacji SWO 2002", H. Sroka (ed.), Akademia Ekonomiczna w Katowicach, Katowice 2002, in Polish

Nycz M. 2003. "Nauczanie wirtualne a wiedza", in: „Komputerowo wspomagane zarzadzanie, R. Knosala (ed.), WNT, Warszawa 2003, in Polish

Nycz M. and B. Smok. 2003. "Distance Education as a Way to Meet the Challenges of the XXI Century", SympoTIC'03 Joint 1st Workshop on Mobile Future & Symposium on Trends in Communications, Proceedings, IEEE CS Section, Bratislava, Slovakia, 26-28 October 2003

Peraya D. 2003. "Distance Education and the WWW", http://www.puw.pl/elerning.html from January 18, 2003

Peraya D. 2003. "Distance Education and the WWW", http://www.puw.pl/elerning.html from January 18, 2003

Piasta Z. 2003. http://zdzich.wsh-kielce.edu.pl/~zbylu/teksty/e_modele.html form February 11, 2003, in Polish

Russel S. and P. Norvig. 1995. "Artificial Intelligence. A Modern Approach", Prentice Hall, Upper River, New Jersey 07458,1995

Singh H. 2002. "Demystifying e-learning standards", http://www.puw.pl/elening, from November 7,2002

Biographies

Malgorzata Nycz, Ph.D. Eng. is a lecturer in the Faculty of Management and Computer Science, Wroclaw University of Economics, Poland. Dr Nycz has authored over 50 publications mostly oriented on intelligent systems topics and distance learning issues. In recent years she has been engaged in the international project "Knowledge Acquisition and Intelligent Distributed Learning in Resolving Managerial Issues" with Belgium. Her current research is in intelligent systems with focus on knowledge discovery from databases and modern education including e-learning processes.

Barbara Smok, Ph.D. is a lecturer in the Faculty of Management and Computer Science, Wroclaw University of Economics, Poland. Dr Smok has authored over 45 publications mostly oriented on databases, intelligent systems topics and distance learning issues. In recent years she has been engaged in the international project "Knowledge Acquisition and Intelligent Distributed Learning in Resolving Managerial Issues" with Belgium. Her current research is in intelligent systems with focus on data warehouses, knowledge bases, and modern education including e-learning processes.

TELE-COMMUNICATION APPLICATIONS

MPEG-21 SESSION MOBILITY FOR HETEROGENEOUS DEVICES

Frederik De Keukelaere
Davy De Schrijver
Saar De Zutter
Rik Van de Walle
Multimedia Lab
Department of Electronics and Information Systems
Ghent University
Sint-Pietersnieuwstraat 41, B-9000 Ghent,
Belgium
E-mail: Frederik.DeKeukelaere@ugent.be

KEYWORDS
Session Mobility, MPEG-21, Digital Item Adaptation

ABSTRACT

Nowadays, multimedia is becoming an important part of our live. Every day, the number of people that are having multimedia experiences is augmenting. To allow this growth in multimedia experiences, a new set of devices has been developed. Each of those devices has different terminal and network capabilities, or even different functionalities. As a result of this growing variety of devices, more and more people tend to use several devices for multimedia consumption. Having a wide set of devices for multimedia experiences, results in a demand for seamless switching between devices, better known as session mobility. In this paper, we discuss how session mobility can be realized between heterogeneous devices using MPEG-21 technology. First, we give an overview of the difficulties that occur when doing session mobility between heterogeneous devices. After this problem statement, we give a detailed discussion on how to overcome those difficulties by using MPEG-21. Throughout this paper, we will demonstrate how different parts of MPEG-21 can be integrated into a complete MPEG-21 compliant multimedia framework that facilitates session mobility between heterogeneous devices.

INTRODUCTION

Multimedia is becoming an important part of our lives. Every day, a lot people are having multimedia experiences. For example, while being at home, the possibility to consume multimedia tends to be ubiquitous. We can watch television, listen to the radio, record video using highly advanced camcorders, etc. Even when we are leaving our houses, multimedia stays with us during our journey. For example, multimedia appears as digital advertisements, in flight movies or simply some background music at work.

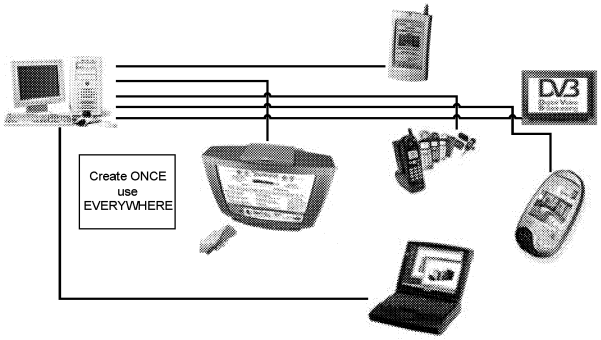


Figure 1: Universal Multimedia Access

Together with this upcoming availability of multimedia, we have a large set of new devices capable of bringing this multimedia experience to end users in a broad environment. Portable audio players, portable video players, powerful hi-fi devices and multimedia PCs are all creating an environment in which it is possible to have access to multimedia content, anywhere at any time.

This concept, accessing multimedia anywhere, at any time and on any device, is better known as Universal Multimedia Access (UMA) (Perkis et al. 2001, Vetro et al. 2003). The goal of Universal Multimedia Access, as presented in figure 1, is to create content once and afterwards allowing access to that multimedia content, anywhere, on any device, using any type of network and at any time.

People nowadays have a wide variety of devices at their disposal for multimedia consumption. Therefore, they tend to switch more often between those devices. To optimize the user experience when switching between devices, it is feasible that such transfers occur transparent without requiring complex user interactions. In this paper we describe how a low complexity interoperable framework allowing transparent transfer of multimedia sessions between devices with different capabilities can be created using MPEG-21. But before describing the MPEG-21 Session Mobility solution, let us have a look at the general problem definition and the difficulties that

occur when transferring multimedia sessions between heterogeneous devices.

DIFFICULTIES WITH SESSION MOBILITY BETWEEN HETEROGENEOUS DEVICES

Transferring a multimedia session from one device to another, generally known as session mobility, has been studied in various domains. While some have focused on the mobility of sessions between applications e.g., between browsers (Song et. al 2002), others have focused on protocols that allow session mobility (Handley et. al. 1999, Schulzrinne et al. 2000).

To realize successful transfer of a multimedia session between two devices, it is possible to use the following three step protocol:

1. Collect session information
2. Transfer session information
3. Process information and continue session

It is straightforward to use this simple protocol in an environment where a session is transferred between devices with the same characteristics. However, currently we are moving towards a Universal Multimedia Access environment in which we have to deal with a variety of devices and session mobility between those devices tends to become more difficult. In this paper, we will address three difficulties that can occur when doing session mobility between heterogeneous devices.

The first difficulty that needs to be solved is the difference in terminal characteristics between devices. For example, watching a video on a terminal with a large screen and then transferring this video session to a terminal that does not support such a large screen will most likely require the adaptation of the video to a lower resolution. Other possible differences in terminal capabilities can cause similar problems for session mobility, such as differences in processing power, different availability of codecs and different battery capacity.

A second difficulty when using a wide set of devices, is the difference in the networks that connect different devices. Similar to differences in terminal capabilities, differences in network capabilities can result in extra complexity. For example, switching from a device with a broadband connection to a device with limited bandwidth will likely require switching to content encoded at a different bit rate in the new session. On the other hand, switching from a device with low bandwidth to a device with high bandwidth will most likely not result in the impossibility to use the same content on the new device. However using the same content would probably be a sub-optimal solution, because

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS">
    <Item>
        <Descriptor>
            <Statement mimeType="text/plain">
                Live: Throwing Copper
            </Statement>
        </Descriptor>
        <Choice choice_id="select_track">
            <selection select_id="track 1"/>
            <selection select_id="track 2"/>
        </Choice>
        <Component>
            <Condition require="track 1">
                <Resource ref="Top.mp3" type="audio/mp3"/>
            </Condition>
        </Component>
        <Component>
            <Condition require="track 2">
                <Resource ref="Alone.mp3" type="audio/mp3"/>
            </Condition>
        </Component>
    </Item>
</DIDL>
```

Figure 2: A music album Digital Item

the new device has the possibility to receive higher quality content. Other differences in network characteristics can also have similar impact on session mobility; examples are differences in error rate of the network and different packet loss ratios.

As a final difficulty for session mobility between heterogeneous devices, we would like to address the interoperability of the messages that are sent between the different devices. To make it possible to do session mobility between devices with different characteristics, it is required to have a platform independent and lightweight language with a common, proprietary or standardized, representation for the messages carrying session information. Lack of a common representation, results in the impossibility to reconstruct the session on the new device based upon the session data of the originating device.

MPEG-21

ISO/IEC 21000, better known as MPEG-21 (Burnett et al. 2003), is the latest standard of the Moving Picture Experts Group (MPEG). In contrast to other MPEG standards, which are mostly targeted at audio and video compression, MPEG-21 is designed to become a generic framework for multimedia production and consumption.

Today, a typical multimedia production and consumption chain involves a large set of tools and standards for generating and consuming multimedia. Because MPEG-21 not only encapsulates data encoding and presentation, but also includes dynamical data adaptation, processing, and rights management, it will most likely succeed in reducing the

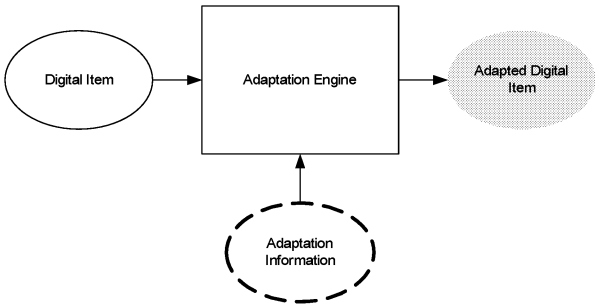


Figure 3: Digital Item Adaptation

multimedia production and consumption chain. MPEG-21 describes how different parts of multimedia technologies interoperate with each other and provides a set of description schemes that can be used in a process in which content is adapted to a specific set of terminal and network characteristics.

Because of the broad scope of MPEG-21, it is divided into several parts, of which Digital Item Declaration and Digital Item Adaptation are the most important parts in the context of session mobility. Most parts of MPEG-21 make intensive use of the Extensible Markup Language (XML) (Sperberg-McQueen et al. 1998) to describe their information.

MPEG-21's fundamental unit of distribution and transaction is the Digital Item (DI). A Digital Item is defined in MPEG-21 as a structured digital object with a standard XML representation, identification and associated metadata. A Digital Item can be seen as a composition of different resources like audio, video, text and metadata such as MPEG-7 descriptions, terminal capabilities and rights expressions.

In figure 2, a music album DI is described in the Digital Item Declaration Language. This Digital Item Declaration (DID) (ISO/IEC 2003) is a digital representation of a music album of the group "Live". The Item contains different components which represent album tracks. The descriptor describes the content of the Item (i.e., the music album) and the components contain resources with a reference to the place where the actual music data is stored (i.e., an mp3-file of the music track). In this example, the user has the ability to choose which track of the album he prefers to play. The choices are located in the choice-tags and based on the selections, the desired resources become available.

MPEG-21 Session Mobility is a part of ISO/IEC 21000-7 Digital Item Adaptation (DIA) (MPEG 2003). The main idea behind DIA is that Digital Items are subject to dynamic adaptation throughout their lifecycle. A high level adaptation process, which is typically used in DIA, is sketched in figure 3. In this adaptation process, the central part is the Adaptation Engine. This part performs

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS">
    <Item id="item_01">
        <Choice choice_id="resolution">
            <Selection select_id="qcif"/>
            <Selection select_id="cif"/>
        </Choice>
        <Component id="qcif">
            <Condition require="qcif"/>
            <Resource mimeType="video/mpeg"
              ref="foreman_qcif.mpg"/>
        </Component>
        <Component id="cif">
            <Condition require="cif"/>
            <Resource mimeType="video/mpeg"
              ref=" foreman_cif.mpg"/>
        </Component>
    </Item>
</DIDL>
```
Figure 4: Choices in Digital Item Declarations

the adaptation of the Digital Item based on different inputs.

The first input is the original Digital Item. This Digital Item is declared in the Digital Item Declaration Language, and typically contains the multimedia data, together with some additional metadata. The second input to the Adaptation Engine is the adaptation information, typically providing information about the context in which the original Digital Item will be used. This information is standardized within the Digital Item Adaptation specification. The result of the adaptation process is an adapted Digital Item.

This adaptation process can be mapped to session mobility in the following manner. Transferring a multimedia session from one device to another device can be seen as the start of a new multimedia session on the target device (i.e., loading the original Digital Item) plus adapting that multimedia session, based on the information from the session on the originating device (i.e., the adaptation information). The result of the adaptation is the continued session on the target device, (i.e., the adapted Digital Item).

With regards to the technical aspects of session mobility, the Digital Item Declaration and the Digital Item Adaptation are the most important parts of MPEG-21. The other MPEG-21 parts that will become important when session mobility is implemented in a real life scenario are the parts of MPEG-21 that are related to Digital Rights Management (DRM). Within MPEG-21 the following parts will realize a DRM framework for multimedia: 21000-4 Intellectual Property Management and Protection, 21000-5 Rights Expression Language and 21000-6 Rights Data Dictionary.

OVERCOMING DIFFERENCES IN TERMINAL AND NETWORK CHARACTERISTICS

The discussion about solving the difficulties for session mobility between heterogeneous devices, will be split up into two sections. In this section we are going to discuss the difficulties of the differences in terminal and network characteristics between two devices. In the next section we are going to address the problem of having a common format for session information.

The main problem for differences in terminal and network characteristics is the fact that moving a session from one terminal to another one will often result in the impossibility to continue exactly the same session as the session on the first device. This is often the case due to the limitations of the hardware of the new device. Displaying a high resolution video on a low resolution display is simply impossible. When transferring in the opposite direction, it is probably possible to display the low resolution video on the high resolution video screen but this will not make optimal use of the hardware of the new device. In many cases this will not be accepted by the end user or at least it will be experienced as sub-optimal.

To solve this problem we use the Choices mechanism from the Digital Item Declaration Language. As mentioned earlier, it is possible to include choices in MPEG-21 Digital Item Declarations. Until now we have used Choices to choose between different parts of a Digital Item. For example, in figure 2 we could choose between different tracks of a music album. In the context of session mobility, Choices can be used to choose between different formats of multimedia content. For instance, in figure 4 the included choice allows the user to choose between two different movies, one in QCIF (176x144) format, and one in CIF (352x288) format. Similar to this type of choice, it is possible to include choices in the DID that are related to any type of terminal or network characteristic.

Figure 4 gives an example of a choice that is related to the terminal characteristics of a device. To be more specific, the choice in question is related to the size of the display. By including a choice in a DID, it is possible to create a Digital Item in such a way that it is possible to use the same Digital Item on devices with different capabilities. The only difference between the consumption of the multimedia content on one device and the consumption on another device will be the state of the choices that are included in the DID, i.e., the QCIF resolution will be selected on the terminal with the QCIF display, and the CIF or the QCIF resolution will be selected on the terminal with a CIF display.

For session mobility, the possibility to include choices in DIDs allows us to overcome the difficulties caused by differences in terminal and network characteristics. To allow transfer of a multimedia session with content at CIF resolution to a device with a QCIF display, it is only required to add a choice in the DID (QCIF or CIF). After transferring a session from the CIF device to the QCIF device, that choice needs to be reconfigured and after reconfiguration successful transfer has occurred.

```
...
<Selection select_id="qcif">
  <Descriptor>
    <Statement mimeType="text/xml">
      <DIA>
        <Description xsi:type="TerminalCapabilitiesType">
          <TerminalCapabilities
              xsi:type="InputOutputCapabilitiesType">
            <Display>
              <Resolution horizontal="176" vertical="144"/>
            </Display>
          </TerminalCapabilities>
        </Description>
      </DIA>
    </Statement>
  </Descriptor>
</Selection>
...
```

Figure 5: Automated Choice reconfiguration

The information included in figure 4 does not allow a machine to automatically choose between the QCIF and CIF format. The information about QCIF and CIF is contained within the select_id of the Selection element. From a machine's point of view, a select_id with the value "qcif" is semantically the same as a select_id with an arbitrary value, e.g., the value "a". The only functionality of the select_id with value "qcif" in the figure is that it makes the Component containing the QCIF video conditionally available because that Component contains a Condition element with a required attribute that has the value "qcif". A full discussion on the different elements of the Digital Item Declaration Language can be found in 21000-2 Digital Item Declaration (ISO/IEC 2003).

To make it possible to automatically configure Choices, it is necessary to include additional information in the Choices, more specific in the Selection elements that allow the run-time configuration of a DID. Additional information can be included in Selection elements using the Descriptor and Statement elements of the DIDL specification. Those DIDL elements allow the inclusion of metadata in Digital Item Declarations. Figure 2 contains an example of a Descriptor with a Statement element containing descriptive information about the content of the album, it contains the title of the music album.

Similar to the inclusion of descriptive data about the content it is possible to include descriptive data about the terminal and network characteristics. Those descriptions can give information about the requirements that need to be satisfied before a certain Selection can be chosen. Figure 5 gives an example of how to include additional information within a Selection element that allows machines to interpret the requirements for choosing a

Selection element correctly. The additional information included in this figure is Digital Item Adaptation information. A part of the Digital Item Adaptation specification standardizes a set of descriptors, called Usage Environment Descriptors, which allow authors to create descriptions of a usage environment. Usage Environment Descriptors make it possible to describe the terminal and network characteristics for a certain device. For session mobility, it is possible to use those Usage Environment Descriptors to allow correct interpretation, especially by machines, of the requirements for choosing a Selection element. Figure 5 contains an example of a Usage Environment Descriptor, describing the display of a terminal. In this example the Selection element can be interpreted as follows: "to choose this Selection, your terminal has to have a display with a horizontal resolution of 176 lines and a vertical resolution of 144 lines". This requirement is equal to the following requirement: "your terminal has to be able to display QCIF format". Using this additional information a machine can automatically make the Selections in a DID and configure that DID accordingly.

For session mobility, the configuration of Choices can be useful when transferring sessions between devices with different capabilities. For example, suppose we have a PC platform on which we are currently running a MPEG-21 multimedia session. The DID that is used in this session contains a set of Choices that allow the DID to be configured for both a PC and a PDA platform. Suppose now that the session from the PC needs to be transferred to the PDA. To realize such a transfer, the PC collects the information about the session, including the information about the Choices and sends that information to the PDA platform. The PDA then processes the session information, loads the Digital Item, and restores the state of the configuration of the Choices from the originating device. At this point the PDA tries to continue the session that was transferred from the PC. However, because of the differences in terminal and network characteristics of the two devices the continuing of the session fails. At this point the PDA can use the additional information in the Selection elements to reconfigure the Choices in such a way that it can actually continue the session. When the PDA tries to continue the adapted session, this time it succeeds, albeit this time with video at a lower bitrate and at a smaller resolution.

A COMMON FORMAT FOR SESSION INFORMATION

The next difficulty that needs to be addressed before session mobility can be realized successfully between heterogeneous devices is the format for the session information. Such a format needs to be lightweight and transparent. This will enable that format to be used on a variety of devices, going from powerful pc platforms, to more restricted, in terms of memory and processing power, platforms such as mobile phones.

Because of the fact that we have been using the XML based DID format for the multimedia content, it would be feasible to use an XML based format for the session information as well. The ideal solution would be to use the same format, being the DID format, to store both the multimedia content and the session information. Using the same format for both purposes, allows the creator of multimedia terminals to reuse its software for both the session mobility part as for the content consumption part of the terminal and therefore reduces the required footprint for such terminals. This is especially important for devices with restrictions on processing power, memory, battery consumption, etc.

Within MPEG-21 Digital Item Adaptation a format for session information has been developed. This format, which is based on the generic DID format, allows the creation of session mobility Digital Items. Those DIs contain the necessary information about multimedia sessions allowing the reconstruction of such sessions based on the session mobility Digital Item.

For the example in figure 4, a session mobility Digital Item, conform the Digital Item Adaptation specification, can be created using the following steps. Initially an empty Digital Item is created using the DIDL elements DIDL and Item.

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS">
    <Item>
    </Item>
</DIDL>
```

In the second step, a Component/Resource child combination is added as a child of the Item from the first step. This Component/Resource contains a SessionMobilityTarget element which references the URI for the content DI. For the example we suppose the content has been placed on a local web server.

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS"
      xmlns:sm="urn:mpeg:mpeg21:2003:01-DIA-SM-NS"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:dia="urn:mpeg:mpeg21:2003:01-DIA-NS">
 <Item>
  <Component>
   <Resource mimeType="text/xml">
    <dia:DIADescriptionUnit
      xsi:type="sm:SessionMobilityTargetType"
      ref="http://192.168.0.1/cdi.xml" />
   </Resource>
  </Component>
 </Item>
</DIDL>
```

For each Choice element in the content DI there is an Annotation within the Item containing an Assertion (ISO/IEC 2003) that captures the current configuration-state of the Selections in the Choice of the content DI.

Suppose that the configuration of the Choice was a CIF resolution.

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS"
      xmlns:sm="urn:mpeg:mpeg21:2003:01-DIA-SM-NS"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:dia="urn:mpeg:mpeg21:2003:01-DIA-NS">
 <Item>
  <Component>
   <Resource mimeType="text/xml">
    <dia:DIADescriptionUnit
     xsi:type="sm:SessionMobilityTargetType"
     ref="http://192.168.0.1/cdi.xml" />
   </Resource>
  </Component>
  <Annotation target="http://192.168.0.1/cdi.xml#item_01">
   <Assertion target="http://192.168.0.1/cdi.xml#resolution"
              true="cif"/>
  </Annotation>
 </Item>
</DIDL>
```

As a final step in creating this session mobility Digital Item we create a Descriptor within the Item containing the information about the multimedia resource that is currently being consumed. For example, the location of the current media stream (e.g., the URI), the position in the current media stream (e.g., 50 sec) and the status of the current session (e.g., pause). This information is stored in a SessionMobilityAppInfo element of type SessionMobilityAppInfoType. The resulting session mobility Digital Item can be found in figure 6.

Based on the information in this session mobility Digital Item, it is possible to reconstruct the multimedia session that was previously running on the originating terminal. It is possible to determine what content was being consumed using the Component/Resource elements, i.e., the Digital Item located at "http://192.168.0.1/cdi.xml". It is also possible to configure the Choices in that Digital Item, i.e., CIF resolution was chosen. If necessary these choices can be reconfigured to adapt the Digital Item to the new terminal characteristics. Finally, it is also possible to restore the state of the resource consumption based on the information that is contained in the Descriptor/Statement children, i.e., the playbackstatus indicates the resource is being played, at a media time of 1.234898 seconds.

CONCLUSIONS AND FUTURE RESEARCH

In this paper we have discussed the difficulties that can rise when doing session mobility between heterogeneous devices. We have located three different causes for such difficulties. The first cause is the difference in terminal characteristics of heterogeneous devices. The second cause is the difference in network characteristics between heterogeneous devices. The final difficulty we discussed is the definition of a common format for expressing session mobility information.

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS"
      xmlns:sm="urn:mpeg:mpeg21:2003:01-DIA-SM-NS"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:dia="urn:mpeg:mpeg21:2003:01-DIA-NS">
 <Item>
  <Descriptor>
   <Statement mimeType="text/xml">
    <DIADescriptionUnit xsi:type="sm:SessionMobilityAppInfoType">
     <sm:ItemInfo target="http://130.130.88.219/cdi.xml#cif">
      <PlayerStatus xmlns="urn:be:ugent:mmlab:sm">
       <PlayBackStatus>isPlaying</PlayBackStatus>
       <MediaTime>1.234898</MediaTime>
      </PlayerStatus>
     </sm:ItemInfo>
    </DIADescriptionUnit>
   </Statement>
  </Descriptor>
  <Component>
   <Resource mimeType="text/xml">
    <dia:DIADescriptionUnit
     xsi:type="sm:SessionMobilityTargetType"
     ref="http://192.168.0.1/cdi.xml" />
   </Resource>
  </Component>
  <Annotation target="http://192.168.0.1/cdi.xml#item_01">
   <Assertion target="http://192.168.0.1/cdi.xml#resolution"
              true="cif"/>
  </Annotation>
 </Item>
</DIDL>
```

Figure 6: A session mobility Digital Item

To overcome those difficulties when doing session mobility between heterogeneous devices, we investigated how MPEG-21 technology could be used to solve the different problems. MPEG-21, which is in fact a new and upcoming multimedia standard, aims to be used on a wide set of devices. To address the difficulties that occur when multimedia consumption is done on such a variety of devices, MPEG has developed the Digital Item Declaration Language. This language allows the declaration of Digital Items that are suited for consumption on devices with different terminal and network characteristics. To be more specific, the Choice elements in the DIDL make it possible for the content creator to include different Choices that allow the configuration of a Digital Item in such a way that it can be consumed on a variety of devices.

For session mobility we have investigated how the Choice mechanism can be used to allow the (re)configuration of Digital Items after a transfer of a multimedia session. We have demonstrated how this can be done by human interaction and without human interaction. To allow automated configuration of Choices, we have demonstrated how Digital Item Adaptation descriptions, more specifically, Usage Environment Descriptions, can be used for dynamic reconfiguration of the Choices without human interaction.

As a final part of the paper we addressed the problem of a common format for session information. Within MPEG-21 Digital Item Adaptation, such a common format has been developed based on the Digital Item Declaration

Language. With this language, it is possible to create standardized session mobility Digital Items that will be correctly interpreted by MPEG-21 compliant terminals.

In this paper we have demonstrated how Digital Item Declaration and Digital Item Adaptation can be integrated to a complete MPEG-21 compliant multimedia framework that facilitates session mobility between heterogeneous devices.

Currently there are two other parts of MPEG-21 under development that can become more important for session mobility between heterogeneous devices. MPEG-21 Digital Item Processing and MPEG-21 Scalable Video Coding are those parts of MPEG-21 that most likely will allow MPEG-21 Session Mobility to reach its full potential. Further study on the integration of those parts of MPEG-21 in our current MPEG-21 Session Mobility framework can therefore be considered as future research.

ACKNOWLEDGMENTS

The research activities that have been described in this paper were funded by Ghent University, 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

Burnett, I., R. Van de Walle, K. Hill., J. Bormans, and Fernando Pereira, "MPEG-21 Goals and Achievements", *IEEE Transactions on Multimedia*, IEEE Computer Society, October-December 2003, p.60-70.

ISO/IEC, "ISO/IEC 21000-2:2003 Information technology -- Multimedia framework (MPEG-21) -- Part 2: Digital Item Declaration," March 2003.

Handley, M., H. Schulzrinne, E. Schooler, and J. Rosenberg,"SIP: session initiation protocol," *Request for Comments 2543*, Internet Engineering Task Force, March 1999

Perkis, A., Y. Abdeljaoued, C. Christopoulos, T. Ebrahimi, and J. Chicharo, "Universal Multimedia Access from Wired and Wireless Systems," *Birkhauser Boston Transactions on Circuits*, Systems and Signal Processing, Special Issue on Multimedia Communications, Volume 20, May-Aug 2001, p. 387-402.

Moving Picture Experts Group, MPEG Home Page, http://mpeg.telecomitalialab.com/

Moving Picture Experts Group "ISO/IEC 21000-7 FDIS Part 7: Digital Item Adaptation", ISO/IEC JTC1/SC29/WG11 N6168, Waikoloa, December 2003.

Schulzrinne, H., and E. Wedlund, "Application-Layer Mobility using SIP", *Mobile Computing and Communications Review*, Volume 4, Number 3, pp. 47--57, July 2000.

Sperberg-McQueen, C. M., T. Bray, and J. Paoli, The Extensible Markup Language (XML), http://www.w3.org/XML, 1998

Song, H., H. Chu, S. Kurakake, "Browser Session Preservation and Migration," *Proceedings of WWW2002*, May 2002.

Vetro, A., C. Christopoulos, and T. Ebrahimi, "Universal Multimedia Access," *IEEE Signal Processing Magazine*, 20 (2) 16-16, March 2003

MODELING JINI-UPnP BRIDGE USING RAPIDE ADL

Ahmed Sameh
Dept. of Computer Engineering
The George Washington University
Washington, DC 20052
Email: sameh@gwu.edu

Rehab El-Kharboutly
Dept. of Computer Science,
The American University in Cairo
P.O.Box 2511, Cairo, Egypt
Email: sameh@aucegypt.edu

ABSTRACT

The exploding deployment of network enabled mobile devices, along with the expansion of networked services have created the need for users to easily manage these devices and services and also to coordinate with one another. Service Discovery Protocol (SDP) enables networked devices, applications, and services to seek out and find other complementary networked devices, applications, and services needed to properly complete specified tasks. A variety of Service Discovery Protocols have been proposed by the market and academia, including Jini, UPnP, SLP, Salutation and Bluetooth. For these protocols to co-exist, they should exhibit interoperability features. A number of bridging techniques have been proposed and implemented. Efforts have been on going to analyze these bridges from an architectural point of view. A most suitable means for such purpose is Architecture Descriptive Languages (ADLs). ADLs, like Rapide, enable the simulation of distributed systems such as Service Discovery Protocols. In this paper we propose a one directional bridging system (Jini-UPnP Bridge). To validate the proposed system, we model and simulate the bridge using Rapide ADL simulation and analysis toolset. We perform a number of simulation tests and use the Rapide Poset viewer to analyze the simulator's output Poset tree of events. The bridge overhead, compared to a non-bridged native Jini service was found to be about 93.5%. The bridge performance was measured under both light and heavy network loads. Under light loads the bridge achieved 0.071% improvement, while its performance has degraded 0.034% under heavy load. The bridge performance was also measured when bridging multiple services. The results fall in reasonable ranges from 1.00079s to 1.00143s for the overall bridging time. To further validate our model, we performed a set of experiments to test communication failures.

KEYWORDS
SDP, ADL, Jini, UpnP, Rapide

INTRODUCTION

The number of networked services is expected to increase enormously in the incoming era. Other than traditional services (e.g. printing, scanning and faxing), new networked-services for business purposes, such as network based computational systems, or light weight services, such as restaurant directories and translators, are becoming available and highly important. For an effective use of these services, users should have means for direct and easy access to these services. Service Discovery Protocol (SDP) presents an attractive solution for services discovery and coordination (Bettstetter and Renner 2000).

One of the main factors of judging the efficiency of a given SDP is its ability to interoperate with other SDPs. Interoperability is a vital issue since it would enable services and clients with different service discovery protocols to communicate and interact with one another. Some of the SDPs use a proxy or bridge as a solution to enable services that don't support their SDP to nevertheless have role in their federations.

In this paper, we present a new approach for bridging between Jini and UPnP. We use architectural modeling to develop a Jini-UPnP Bridge. We validate our work by carrying out a series of simulation tests and experiments on the executable architectural model. Initially, we set a hypothetical topology of Jini and UPnP clients and services in addition to our proposed Jini-UPnP bridge. This setup is used to verify that the Jini-UPnP Bridge is capable of registering a UPnP Service that offers a JiniFactory, with the Jini Lookup service. The basic functionalities of The Jini-UPnP Bridge are tested and verified. We assess the performance of the Jini-UPnP Bridge through a number experiments including: 1- measuring the overhead of bridging a UPnP service versus direct registration of a Jini native service, 2- measuring the performance of the bridge under both light and heavy network loads, 3- deducing the performance of the bridge on bridging multiple UPnP JiniFactory services. Moreover, we performed the set of experiments conducted by Dabrowski and Mills in (Dabrowski and Mills 2001) to test the behavior of our hybrid-bridging environment in cases of communication failures. We compared their results to ours to validate the correctness of our model (El-Karboutly 2002).

This remainder of this paper is organized as follows: In Section 2, we describe the proposed Jini-UPnP bridging technique. First we give a high level design view and then we present some implementation details. Our tests and experimental work is discussed in Section 3. We conclude in Section 4.

THE PROPOSED Jini-UPnP BRIDGE

One of the main factors of evaluating and judging any of the available SDP protocols is the extent to which it allows for interoperability. A bridge between UPnP and Jini has not been investigated before; though it has been mentioned as possibility in a number of references (IBM 1999) (Richard 2000) (ADL 1997).

Both Jini and UPnP introduce the concept of bridging a foreign network device as part of their specifications. Jini refers to it as a network proxy (Luckham 2001). While UPnP refers to it explicitly as a UPnP Bridge (Wang 2003). In both SDPs, the bridging concept is based on introducing a foreign device to the SDP environment through the use of a representing entity that speaks on its behalf (a bridge).

The choice of bridging Jini and UPnP is based upon the fact that both protocols, though similar at the core functionality level, have dissimilar points of strength. Both Jini and UPnP support the same set of basic SDP operation, including service advertisement and service discovery. They both support the concept of leasing for registered services and support eventing and notification mechanisms for updating service information. Jini a centric protocol, based on the presence of a central cache manager, is an example of three-party protocols, which cannot function without a Lookup Service. On the other hand, UPnP is decentralized and is more of a peer-to-peer communication model. Compared to Jini, UPnP is a lightweight protocol. This is due to the fact that Jini requires the presence of a JVM for all its entities. Bridging between Jini and UPnP will enable thin services that don't have a JVM to announce their services to Jini clients. Jini's most attractive feature is the ability of downloading services driver's or proxy, which enables easy and direct usage of the service.

Our work is built on the concept of a Jini network proxy described in Jini Device Architecture and is based on the efforts of Eric Guttman in (Guttman and Kempf 1999). A Jini-UPnP Bridge is an entity that enables services that support UPnP protocol to be reachable by Jini clients. For Jini clients, Jini-UPnP is a transparent layer that they are unaware of. The UPnP services that are advertised via the bridge are treated as native Jini services.

The proposed Jini-UPnP Bridge is modeled as a special network node that can communicate with other network nodes in both Jini and UPnP protocols. It mainly acts as a *Service User* (i.e. Control Point) in UPnP environment and a *Service Manager* (Service) in Jini environment. It waits for announcements made by UPnP devices and services that are willing to advertise their presence to the Jini clients and acts as a representative, almost a mirror for them in the Jini environment.

The first order of business of the proposed bridge is to prepare an appropriate entry for UPnP services, in the Jini Lookup Service. This involves primarily setting the appropriate attributes required and creating a service object as part of Jini service's registration.

UPnP services that are willing to advertise their presence to Jini clients are not required to have a JVM installed. They are mainly required to have a ***Jini driver Factory*** (Guttman and Kempf 1999). A Jini driver factory is a (*.jar) file that bares a manifest for the advertised service. A Java Archive File (*.jar) file is used to bundle multiple files into a single archive file. Typically a JAR file contains the class files and auxiliary resources associated with applications.

The proposed bridging process is done through the following steps:

The Jini-UPnP bridge searches the UPnP reachable entities to find devices and services that have Jini driver Factory or waits till it receives announcements made by Jini driver Factory services.
Once a Jini driver Factory service is found, the Jini-UPnP bridge obtains a complete description of the service including attributes, GUI URL and control URL.

The URL of the Jini driver factory is composed by extending the control URL with a unique identifier. The Jini driver factory is downloaded using GET method over HTTP.

The Jini-UPnP bridge performs attributes transformation from UPnP format to Jini format to prepare for service registration. Upon successfully translating the entire service attributes and obtaining the Jini driver factory, the Jini-UPnP bridge registers the discovered service with **Jini Lookup Service**. Using the Jini driver factory, the bridge creates a service object that is used for registration. Registration is done by sending a join request with all necessary attributes to **Jini Lookup Service** that adds the new service to its cache.

Whenever a Jini client needs our bridging service, it contacts **Jini Lookup Service** and downloads the instantiated object that is used to drive the service. Like any typical Jini service, the Jini-UPnP bridge should be equipped with JVM to be able to participate in the Jini SDP.

The first step in modeling our bridge is to set a hybrid Service Discovery environment, where different services and clients speak different service discovery protocols. This means that we would have n Jini services, m Jini clients, e Jini lookup services , p UPnP services and q UPnP clients, where n,m,e,x,p,q are natural numbers > 0 and by setting them we define our topology. This topology would be ADL modeled such that entities are able to perform normal service discovery operations with no conflicts.

Having the two NIST Rapide models for Jini and UpnP (Dabrowski and Mills 2001) , we merged the two models into one model with both Jini and UPnP interfaces and main modules in preparation to build our proposed bridge. The proposed Jini UPnP bridge is basically a network node that acts as a *UPnP SM* in UPnP environment and a *Jini SU* in Jini environment. It's basic sub modules are the basic components of UPnP SM and Jini SU models, in addition to sub modules that perform bridging.

The main sub modules of Jini-UPnP Bridge architecture are:

UPnP Service User (UPnP SM): is a modified implementation of the UPnP SM entity that also includes ***UPnP Local Cache Manager*** and the ***UPnP SU Filter***. The ***UPnP Local Cache Manager*** is modified such that it handles attribute translation from UPnP to Jini and also Jini driver factory download.

Jini Service Manager (Jini SM) : is a modified implementation of the Jini SU that communicates directly with the UPnP SM module of the bridge to receive bridged services.

In normal UPnP SU, the local cache Manager module is an interface for the internal cache of the SU. It handles UPnP discovered service records, notifications and events. In our bridged model it also handles the functionality of managing a cache for the Jini driver factory of the discovered Jini Driver Factory services. It implements the interface MANAGED_RESOURCE_JAR which exposes two methods: SUGetJar that requests downloading a jar file for a given Jini Factory service, and SMJarResponse which is the response to a SUGetJar request. MANAGED_RESOURCE_JAR is represented in Rapide ADL as follows:

```
TYPE MANAGED_RESOURCE_JAR IS INTERFACE
ACTION
 OUT
    SUGetJar
     (SU_ID, SM_ID : IP_Address;    -- Source SU, target
SM
      QueryIssueTime : TimeUnit;    -- time query issued
      URLField : Integer);          -- This should be a URL
or a  Device ID for  identification purposes
 IN
    SMJarResponse
     (SM_ID, SU_ID : IP_Address;      -- Sending SM,
Receiving SU
      UniqueID      : Integer;        -- Unique Identifier for
SD
      Jar  : String;                  -- a dummy string representing
the  downloaded file
      TimeStamp : TimeUnit);
 END;
```

Upon discovering the presence of a UPnP Service that provide a Jini Driver Factory, the Jini UPnP Bridge; first retrieves its complete description and downloads its jar file and then advertises its presence to the Jini Lookup Service. To perform the last functionality, Jini UPnP Bridge uses the interface ADVERTISE_SERVICE. ADVERTISE_SERVICE is responsible for propagating discovery of new service, change of a currently discovered service and deletion of a service to the JINI SM sub modules of the bridge. It is called by the Bridge Local Cache Manager sub module and implemented by the Jini Service Repository sub module.

ADVERTISE_SERVICE interface is presented as follows in Rapide ADL:

```
TYPE ADVERTISE_SERVICE IS INTERFACE
ACTION
OUT AddNewService(?Service_ID : Integer; -- ID of the
service
ServiceType,          -- service type /name
ServiceAttributes,    -- service attributes
ServiceAPI,           -- service Proxy and APIs
ServiceGUI : String; -- service GUI
NLeaseTime,
NDuration : TimeUnit
- lease duration),
    ChangeServiceEv (?Service_ID : Integer; -- ID of the
service
   ServiceAttributes:String -– new service
Attributes ),
    DeleteServiceEv   (?Service_ID : Ind_Service_ID;   --
Service ID
    ExpireOption : String       --Expire Option);

END; --ADVERTISE_SERVICE
```

A UPnP service that wishes to be used by Jini clients through our Jini UPnP bridge, should provide a Jini driver factory. The Jini UPnP Bridge issues an HTTP Get command to download the Jini driver factory file. A change was necessary to the UPnP SM Rapide Model for providing this functionality. The MANAGED_RESOURCE_JAR interface, introduced in the last section, is added to the UPnP Service Manager Model to be implemented by the UPnP SM_Repository sub module.

The overall Rapide model for a hybrid SDP environment with Jini UPnP Bridge consists basically of six different types of network entities: Jini SM, Jini SU, Jini SCM, UPnP SU, UPNP SM and Jini UPnP Bridge. Each of these modules implements the basic functionally of UPnP and Jini SDP Protocols. The Jini UPnP Bridge modules implements protocols of Jini SM and UPnP SU in addition to bridging functionality.

On the **network level**, the Jini-UPnP bridging environment consists of network nodes that are connected through communication links. Communication links are

mainly TCP/IP and UDP connections that are used for multicasting and unicasting messages. These communication links are modeled in our Rapide ADL as separate entities representing different multicasting and unicasting functionality.

The six network nodes: Jini SM, Jini SU, Jini SCM, UPnP SU, UPNP SM and Jini UPnP Bridge consist of **major functional components.** These are shown on the **Entity Major Functions** layer or the third layer from top. For Example the ***Jini Service Manager*** entity consists of ***a Service Repository*** and ***SCM discovery*** modules.

The lower level in the architecture shows the **main functional subcomponents**. These are the main components that carry out the main functionalities in the system. Some of these subcomponents are modeled as a Rapide interface and are implemented by different higher level models, while the rest are implemented as independent low level functionality modules. The main functional subcomponents of the ***SCM Discovery*** module, which is a basic module required in all Jini entities, is divided into three groups: Direct Discovery Protocol subcomponents, Aggressive Discovery subcomponents and Lazy Discovery subcomponents. Subcomponents that implement **Lazy Discovery Protocol** are: the ***Announcement Responder,*** which listens passively for announcements from entities that the SCM may wish to discover, the ***Announcer*** subcomponent, whose role is to send announcements to entities that may wish to discover the SCM to which it belongs, the ***SCM API Server,*** which provides service interfaces (APIs) to discovering entities after the initial response by the discovering entity to the SCM announcement and the ***Executive*** subcomponent whose main task is to control switching between aggressive, lazy and directed discovery.

Jini-UPnP BRIDGE TESTING and PERFORMANCE MEASURES

The next step after modeling the bridging between Jini and UPnP is to verify that the basic functionality of the bridge is correct through simulation tests. The Rapide toolset provides a set of compilation and runtime execution tools whose output is a simulation of the Rapide architectural model. The output of the simulation could be analyzed in various ways, including constraint checking, analysis for surprises and depiction of behavior. We chose to analyze the output of our simulation using the **P**artial **O**rder **S**et (Poset) browser. Poset browser enables us to view how a given architectural design behaves. It represents casual event simulations in a DAG form, nodes representing events and directed arcs representing causality.

In each of our tests, we first establish initial conditions by constructing a topology of Jini and UPnP basic entities in addition to the Jini-UPnP Bridge. The following tests have been conducted and proven successful: 1- testing to validate that initial discovery and advertisement activities in our hybrid environment of both Jini and UPnP entities, function correctly, 2- testing a complete scenario of bridging a Jini Service to examine the correctness of the bridging process, 3- testing that the proposed UPnP Jini Bridge successfully propagates changes that occur in the JiniFactory service to the SU Jini clients that have previously discover it, 4- testing to confirm that the JiniFactory service shutdown is propagated successfully to Jini SCM through Jini-UPnP Bridge.

We have conducted five experiments to measure the performance of the proposed Jini-UPnP bridge. In the following we discuss and report only four of them, naming: 1- measuring the overhead of bridging a UPnP service verses direct registration of Jini native service, 2- measuring the performance of the bridge under both light and heavy network loads, 3- deducing the performance of the bridge on bridging multiple UPnP JiniFactory services. Moreover, we performed the experiments conducted by Dabrowski and Mills in (Dabrowski and Mills 2002) to test the behavior of our hybrid-bridging environment in cases of communication failures. We compared their results to ours to validate the correctness of our model (El-Karboutly 2002).

The usage of a bridge in a hybrid system implies the presence of an overhead in time and resources. We are interested in measuring the overhead of bridging a UPnP service compared to having that same service as a native Jini service. The overhead is measured in terms of time and the number of messages exchange.

The following table shows the most relevant parameters and values for our experiment.

	Parameter	**Value**
General Parameters	Simulation overall time	3600s
	Node Startup Delay	1-15 s uniform
Behavior in both Jini and UPnP architectures	Polling interval	180s
	Registration TTL	1800s
UPnP specific behavior	Announcement interval	1800s
	Msearch query interval	120s
	SU purges SD	At TTL expiration
Jini specific behavior	Probe interval	5s (7 times)
	Announce interval	120s
	SM or SU purges SD	After 540s with only REX
Jini UPnP Bridge specific behavior	Jar file size	11Kb
Transmission and processing	UDP transmission delay	10 µs constant

delays	TCP transmission delay Per item processing delay	10-100 µs uniform 10 µs for cache items 10 µs for other items

Table 1 Jini-UPnP Rapide Model Input Parameters

First, we ran the Jini Rapide model with a topology of one Jini Service Cache Manager (SCM), two Jini Service Users (Jini SUs) and one Jini Service manager (Jini SM), where one of the Jini SUs requests a service of the same type as that offered by the Jini SM. We measure the time taken and the number of messages exchanged since the Jini SM starts up and until the Jini SU receives the service description. Next, we run our Jini-UPnP Bridged model with a topology of one Jini SCM, two Jini SU, one Jini SM, one Jini-UPnP Bridge, one UPnP SU and two UPnP SM. The time taken by a Jini SU to discover a requested UPnP service is measured. This time value is the sum of the time taken for Jini UPnP Bridge to discover the services; the time the bridge registers this service with the Jini SCM and the time the Jini SCM forwards the service description to the interested Jini SU.

Measurements for Jini are done on two stages; first we measure the time taken for Jini SM to register with SCM and the number of messages needed. We assume that SCM discovery has already taken place. The time taken for this operation, as shown in the results is **TIME TAKEN 1 = 0.064s** , and the number of messages exchanged is four messages (NUM MSGs 1 :4). The second stage is where the SCM starts matching the newly added service description to the available SU requests. Two messages are exchanged for this operation to complete and the total time needed is **TIME TAKEN 2 = 0.00081s**. Thus the total time for the whole operation starting with SM registration to SU discovery takes **TOTAL TIME = 0.06481s** on average.

Bridging a UPnP SM service to be reachable for Jini SUs is done in three stages. First the Service SM is discovered by the Jini-UPnP bridge, then the bridge registers the service with Jini SCM. The time taken for a Jini-UPnP bridge to discovery and obtain the complete description of Jini Factory service is **TIME TAKEN 1:1.00132s** where five messages are exchanged in this operation. Secondly, the bridge registers the newly discovered service with the SCM by exchanging two messages in **TIME TAKEN 2:.00022**. The last stage is where the SCM matches the added service to the notification for services that SUs have registered with the SCM earlier. This operation exhausts about **TIME TAKEN 3: 0.00061s**. The total time consumed in the process of bridging **TOTAL TIME = 1.00215s**

Comparing the results for a native Jini service to that of bridging the service through Jini-UPnP Bridge, it is clear that the bridging process has an overhead of about **0.93734s** or a 93.5% overhead.

Network Bandwidth is a main factor in the behavior of any distributed system. The performance of different entities in a SDP is very much affected by network delays as a main parameter. In our model for Jini-UPnP Bridge, we simulate network bandwidth by having network delay as one of the main model input parameters. Parameters are defined for unicast and multicast delays between any pair of nodes and also for the network as a whole. The following tests record the effect of varying network delays on the performance of UPnP-Jini Bridge.

In the pervious experiment we were interested in measuring the overhead of bridging a service in terms of time and number of messages. We fixed the TCP/IP network delay to a typical network delay value of 10-100 µs uniform. To measure the performance of the Jini-UPnP Bridge in a light loaded network, we repeat the experiment done in the previous section with the same input parameters, yet changing the TCP/IP network delay to **10-30 µs uniform.** The results would be compared to those obtain in the pervious section. We repeated the experiment ten times to compute the average overall time taken by the bridge.

Compared to the results obtained in the previous experiment, the bridge performance increases about 0.071 % with a less loaded network (i.e. higher bandwidth) of 10-30 µs uniform delay. The results show an improved value for the time of registration with the bridge from 1.00132 s in normal network to 1.000617 in a less loaded network. We are more interested in the last time value **(Overall Time)** since the time taken to download the Jini driver factory is a factor of it. The results are up to our expectations since an overall improvement in time delay is noticed.

To measure the performance of Jini-UPnP Bridge in a congested network, we apply the same experiment with a higher network load with the same input parameters, yet changing the TCP/IP network delay to **80-100 µs uniform.** The results would be compared to those obtain in case of typical network delays. We repeated the experiment ten times to compute the average overall time taken by the bridge.

Compared to the results in normal network condition that are obtained in the previous experiment, the bridge performance degraded about 0.034 % with a congested network (i.e. low bandwidth) of 80-100 µs uniform delay. The result is as expected since the effect of having a low bandwidth is of direct effect on the time taken to transfer messages and to download Jini driver factory. The overhead in time is more obvious in the time taken for registration with the bridge, as downloading the Jini driver factory file is a factor in it.

A UPnP client (UPnP SU), in a pure UPnP environment, is capable of discovering and communicating with multiple UPnP Services at the same time. Also, a Jini

Service Manager (Jini SM) could advertise and register the availability of more than one service. Our UPnP-Jini Bridge is primarily composed of both a UPnP SU and Jini SM. Thus a UPnP-Jini Bridge is capable of bridging more than one UPnP service and registering it with the Jini SCM at the same time. We are interested in testing this capability of our modeled Jini-UPnP Bridge to bridge successfully multiple services at the same time and also to depict the effect of multi-service bridging on the Bridge performance.

In the previous experiment, we've chosen a topology with one UPnP Service (UPnP SM) that offered a Jini Factory and is bridged using the UPnP Jini Bridge. In this experiment, we conduct a topology of five UPnP SMs to be bridged, one UPnP Jini Bridge, one UPnP Service User, one Jini SCM, two Jini SUs and one Jini SM. We assume the same input delays and parameters presented above. We record the time taken for a Jini SU to discover a requested UPnP service. This time value is the sum of the time taken for Jini UPnP Bridge to discover the services; the time the bridge registers this service with Jini SCM and the time the Jini SCM forwards the service description to the interested Jini SU.

The results obtained are not uniform, yet they fall in a certain time range, for example the overall time taken by UPnP-Jini bridge to bridge a given service ranges between 1.00079 s and 1.00143 s. These results are expected since the behavior of the bridge is a function of the number of events it receives at the same time and the way it schedules the incoming events. The results fall in reasonable ranges and are close to the results obtained in case of bridging one service. These results are also dependent on the time each node starts announcing its service. Nodes that announce their services consecutively with a small time variant (e.g Nodes 2, 3), cause high frequency of events on the bridge, which results in degradation in the bridge performance and higher delay values.

CONCLUSION

The problem we addressed in this research is enabling thin servers and lightweight devices to offer their services to Jini clients through passive and indirect registration using our proposed Jini-UPnP Bridge. This problem has been addressed before by using SLP instead of Jini (Guttman and Kempf 1999), yet the bridging between Jini and UPnP has not been investigated before in SDP research literature.

We modeled and simulated our solution using Rapide ADL toolkit. Modeling is an approach for designing quickly, efficiently and correctly. It allowed us to control the quality and performance. We've chosen Rapide ADL to benefit from the set of modeling and simulation tools it offers. We used architectural models of Jini and UPnP as a basis to create hybrid discovery environment including both Jini and UPnP and to design and model our proposed bridge. For testing and simulating the bridge, we created a hypothetical topology of Jini and UPnP clients and services in addition to our proposed Jini-UPnP bridge. We simulated the topology to verify that the Jini-UPnP Bridge is capable of registering a UPnP Service that offers a JiniFactory, with the Jini Lookup service. The Jini-UPnP Bridge is tested for cases where the bridged service is updated or deleted. A number of performance experiments have been done on the bridge.

REFERENCES

ADL 1997, "Using Architecture Description Languages (ADLs) to Improve Software Quality and Correctness in Dynamic Distributed Systems" http://www.itl.nist.gov/div897/ctg/adl/sdp_projectpage.html

Bettstetter, C. and C. Renner, 2000, "A Comparison of Service Discovery Protocols and implementation of the Service Location Protocol", *In Proceedings of EUNICE 2000, Sixth EUNICE Open European Summer School, Twente, Netherlands.*

Dabrowski, C. and K. Mills, 2001, "Analyzing Properties and Behavior of Service Discovery Protocols using an Architecture-based Approach", *Proceedings of Working Conference on Complex and Dynamic Systems Architecture.*

Dabrowski, C. and K. Mills, 2002 "Understanding Self-healing in Service-Discovery Systems," *ACM Workshop on Self-Healing Systems*, Charleston.

El-Kharboutly, R. 2002, "Modeling Jini-UpnP Bridge Using Rapide ADL", M.Sc. thesis in Computer Science, The American University in Cairo.

Guttman, E. and J. Kempf, 1999, "Automatic Discovery of Thin Servers: SLP, Jini and the SLP-Jini Bridge," *Proc. 25th Ann. Conf. IEEE Industrial Electronics Soc. (IECON 99), IEEE*, Press, Piscataway, N.J.

IBM 1999, white paper, "Discovering Devices and Services In Home Networks".

Luckham, D. 2001, "Rapide: A Language and Toolset for Simulation of Distributed Systems by Partial Ordering of Events," http://anna.stanford.edu/rapide .

Richard, G. 2000, "Service Advertisement and Discovery: Enabling Universal Device Cooperation," *IEEE Internet Computing.*

Wang, O. 2003, "Interoperability of COM/DCOM objects with CORBA objects by using DCOM/CORBA Bridge and their performance analysis", http://www.engr.sjsu.edu/fatoohi/wang-report/abstract.html

SLEEPING IN JAVA

Gunther Stuer, Kurt Vanmechelen, Jan Broeckhove, Tom Dhaene
Dept. of Mathematics and Computer Science
University of Antwerp
Middelheimlaan 1, 2020 Antwerp, Belgium
gunther.stuer@ua.ac.be

KEYWORDS

Java, Performance Modeling, Timing.

ABSTRACT

When simulating the behavior of fine grained distributed systems or networks, the domain specific activities are often replaced by pauses. When programming these simulations on Linux, obtaining accurate pauses is not as straightforward as one might expect.

This contribution compares three Java pausing techniques: Java's standard `sleep` method, busy waiting and RTC (real time clock) interrupt counting. Furthermore, a statistical analysis of their accuracy is presented.

INTRODUCTION

As the complexity of the systems being studied increases, it becomes more and more important to build accurate models. Using these models, the original system can be simulated and the effect of various parameters can be evaluated.

When simulating the behavior of fine grained distributed systems or networks, domain specific activities are often replaced by pauses [1][2]. The length of these pauses has to be accurate in order to have reliable results. As we will show in this contribution, this is not always as straightforward as one might expect when running the simulation on a standard Linux platform. Three different techniques will be discussed in detail and a statistical analysis of their accuracy will be presented. The experiments were conducted on a 1.7GHz, 256MB, SuSe 8.2 Linux computer with a 2.4.20 kernel. We have used Sun's Java 1.4.1 as principal Java Runtime Environment, with native method calls through JNI where necessary.

THE EXPERIMENTS

To test the accuracy of the different sleep implementations, we have built a framework based on the *strategy* design pattern [3]. This pattern decouples an algorithm from its host by encapsulating the algorithm into a separate class. More specific, an object and its behavior are separated and put into two different classes. This allows one to switch the algorithm being used at any time without needing to modify the code of the host. In this paper, we apply this for different implementations of the sleep method.

In our framework, all tested implementations are encapsulated in a class implementing the `Sleeper` interface. This interface contains only one method declaration, which will execute a sleep with a duration of `nrOfMillis` :

```
void sleep(long nrOfMillis)
```

Depending on a command line argument, the appropriate sleep implementation is instantiated and an experiment is performed.

In what follows, T will denote the *requested* sleeping time in milliseconds and A the *actual* time slept in milliseconds. A is measured by two `System.currentTimeMillis()` calls that immediately precede and follow the `sleep` call. A `System.currentTimeMillis()` call returns a long integer containing the difference, measured in milliseconds, between the current time and midnight, January 1, 1970 UTC. The timing granularity of `System.currentTimeMillis()` on our Linux platform is 1 millisecond.

We will make 100 measurements of A for every value of T. Since we want to avoid dependencies on periodic sampling, we do not use a simple loop to perform the 100 measurements. Instead, we have used shell scripts to iteratively invoke a `main` method that executes the desired sleeping behavior. A will be measured for succesive values of T incrementing with 1ms for $0 < T \leq 100$, with 100ms for $100 < T \leq 1000$ and 500ms for $1000 < T \leq 4500$.

We use MA to denote the mean value of A over the 100 samples and $AERR$ to denote the mean absolute error of T in respect to A.

JAVA THREADS

The first sleep implementation we will discuss is the default Java sleep method. As it is the default sleeping technique in Java, it is by far the most used one. Unfortunately though, it is also the worst one.

METHOD

A process may be scheduled to sleep for a certain amount of milliseconds by issuing a static `Thread.sleep` call. This call will preempt the Java thread that executes the statement.

RESULTS
The graph in figure 1 shows a scatter plot of the requested and actual sleep times. As the observer behaviour is the same for all values of T, for clarity, only the segment of $0 < T \leq 100$ is shown. When one interprets the data presented in this graph in terms of the mean of A one finds :

- If T = 0 then MA = 0
- If $(T \ mod \ 10) = 0$ then $MA = T + 5$
- If $(T \ mod \ 10) \neq 0$ then $MA = ((\lfloor T/10 \rfloor)+1)*10+5$

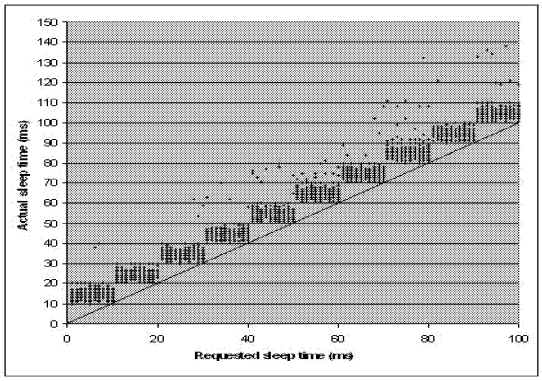


Figure 1: Scatter plot showing requested versus actual sleeping time using a Java `Thread.sleep()` call.

In summary we conclude that the actual sleep time using `Thread.sleep` is given by rounding T to the next multiple of 10 milliseconds, augmented with a variable amount of time that has a mean value of 5 milliseconds, and a range of 0 to 9 milliseconds.

We have also observed that scheduling these sleeping processes in a periodic manner, i.e within the body of a loop, results in the variable time having a constant value of 9. In this case, a sleep with period of 16 milliseconds, for example, always has an actual sleeping time of 29 milliseconds.

This behavior can be explained by the fact that the Linux scheduler has a granularity of 10 milliseconds. Hence all scheduling requests are rounded to a multiple of 10. Furthermore, since the requirements state that `Thread.sleep(x)` should sleep *at least* x milliseconds, this value is rounded upwards. Finally, when a process is notified by the scheduler in time quantum a, the process will only be in the runnable state in quantum $a + 1$. The extra amount of oversleeping within the 0 to 9 millisecond range, is thus determined by the time instance within the quantum a on which the sleep call was issued.

When sleeping in a loop, the new sleep statement will always occur at the very beginning of the process' time quantum, resulting in the apparent constant addition of 9 milliseconds to the already rounded requested sleep time. It is therefore clear that the accuracy of the standard Java pause technique is inadequate for fine grained pauses.

BUSY WAITING

For the second technique, we use busy waiting under a soft real-time scheduling policy to obtain a higher accuracy.

METHOD
The system `time.h` header contains a C function called `nanosleep`. In general, the accuracy of this sleep function is comparable to that of the Java method since it has the same dependencies on the kernel's scheduling accuracy. However, the process's scheduling policy may be set to `SCHED_FIFO` or `SCHED_RM`. The process is then scheduled as a *soft real-time process*. Under these scheduling policies, the `nanosleep` function will perform *busy waiting* if it is instructed to sleep for $T < 2$. This busy waiting procedure guarantees accurate sleeps within the order of microseconds if $T < 2$. To sleep for a period of n milliseconds we will call `nanosleep` n times in a for loop, with a sleep duration of 1 millisecond.

Since a soft real-time process has priority over all standard system processes that run under the `SCHED_OTHER` scheduling policy, the sleeping process must run under root permissions.

Because the sleep technique needs to be usable within a Java context, we use the Java Native Interface (JNI) to call the native sleeping code depicted below :

```
JNIEXPORT void JNICALL
  Java_BusySleeperNC_sleep
  (JNIEnv *, jclass, jlong x)
{
    int max = sched_get_priority_max(SCHED_FIFO);
    struct sched_param p;
    p.sched_priority = max;

    sched_setscheduler(0, SCHED_FIFO, &p);

    struct timespec t,r, start, end;
    t.tv_sec = 0;
    t.tv_nsec = 1*1000*1000;

    for(int i = 0; i < x; i++) {
        nanosleep(&t, &r);
    }
}
```

RESULTS
The graph in figure 2 shows $AERR$ when using the busy wait method in the 0-100 millisecond range.

In this range, the method shows high accuracy. The mean error is 0.07 ms with a standard deviation of 0.16 ms. However, if we look at the absolute error of higher sleeping durations in figure 3, we observe that `nanosleep` returns too early and that the error increases with increasing T. This observation was also made in [4] when evaluating the accuracy of busy waits on a Linux 2.4.0-test6 kernel.

To correct this, we time the amount of time slept in the native method and apply a correction afterwards. The adjusted code is depicted below:

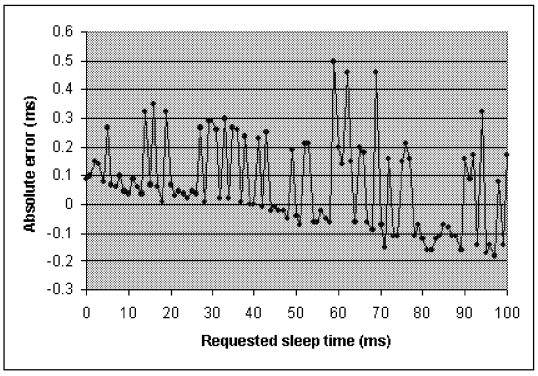


Figure 2: Absolute error on sleep method using busy waiting, in the 0-100 millisecond range.

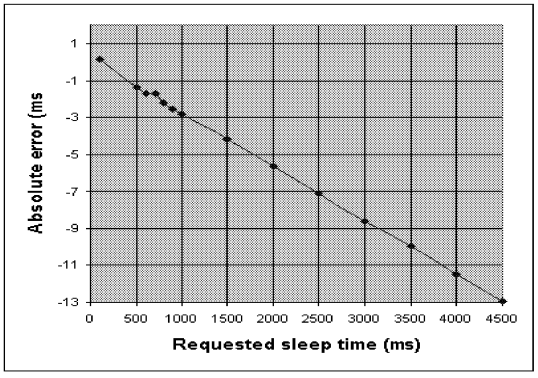


Figure 3: Absolute error on sleep method using busy waiting, in the 100-4500 millisecond range.

```
JNIEXPORT void JNICALL
Java_BusySleeper_sleep (JNIEnv *, jclass, jlong x)
{
    int max = sched_get_priority_max(SCHED_FIFO);
    struct sched_param p;
    p.sched_priority = max;

    sched_setscheduler(0, SCHED_FIFO, &p);

    //Configure 1 millisecond time structure
    struct timespec t,r, start, end;
    t.tv_sec = 0;
    t.tv_nsec = 1*1000*1000;

    clock_gettime(CLOCK_REALTIME, &start);
    for(int i = 0; i < x; i++) {
        nanosleep(&t, &r);
    }
    clock_gettime(CLOCK_REALTIME, &end);

    //Compute extra sleep needed in microseconds
    long microDiff = x*1000 -
      ((end.tv_nsec - start.tv_nsec)/1000 +
      (end.tv_sec - start.tv_sec)*1000*1000);

    for(int i = 0; i < (microDiff)/1000; i++) {
        nanosleep(&t,&r);
    }
}
```

This method exhibits higher accuracy in the whole range 0-4500, the mean absolute error is 0.0097 ms with a standard deviation of 1.502 ms. The relatively high standard deviation is caused by 39 outliers with mean of 25.2 milliseconds and a standard deviation of 2.63 milliseconds. In the sample population of 11700 samples, they represent 0.33% of the number of cases.

A disadvantage of this method is the fact that the process must run under root permissions. Another, more prominent, disadvantage is the amount of CPU usage. Busy waiting fully blocks all other system processes that are not scheduled under a soft real time policy. In general, this approach cannot be used in multi-threaded applications.

RTC INTERRUPT COUNT

The Real Time Clock interrupt count is a sleeping technique which depends on the presence of a chip with real time capabilities, as is the case in almost all modern computers.

METHOD

Nearly all PC's contain a MC146818 compatible chip which offers real time clock services independent of the CPU. In Linux one can access this clock through the `/dev/rtc` device. In order to interface with the clock, one includes `mc146818.h` and `ioctl.h` system header files. The clock can be configured to generate interrupts with a rate between 64 Hz and 8192 Hz. The interrupt rate is required to be a power of 2. Only processes with root privileges are allowed to increase the clock's frequency beyond 64 Hz, a restriction that dates from the days when the generation of thousands of interrupts represented a high system load.

This method counts the number of generated interrupts in an infinite while loop, computing the elapsed sleeping timespan on every iteration. If this span exceeds the required sleeping time x, the loop is exited. In every iteration, a read operation is issued on the clock which blocks until an interrupt arrives. If RTC interrupts have already arrived, the method returns an unsigned long and blocks. The return value contains the interrupt status in the lower byte, the remaining higher order bytes contain the number of interrupts generated since the last call to read.

Following native code performs sleeps by counting the RTC clock's interrupts :

```
JNIEXPORT void JNICALL
Java_IRQSleeper_sleep(JNIEnv *, jclass, jlong x)
{
    if(x == 0) return;

    int RTCHandle = open("/dev/rtc", O_RDONLY);
    long irqCount=0;
    int FREQ = 1024;
    unsigned long data;
    int ret;

    // Set the interrupt frequency to 1024 Hz
```

```
    retval=ioctl(RTCHandle, RTC_IRQP_SET, FREQ);

    // Enable periodic interrupts
    ret = ioctl(RTCHandle, RTC_PIE_ON,0);

    while(1) {
        ret = read(RTCHandle, &data,
                   sizeof(unsigned long));
        irqCount += data/256;
        if((double)irqCount*1000/FREQ > x)
        { break; }
    }

    // Disable periodic interrupts
    retval = ioctl(RTCHandle, RTC_PIE_OFF, 0);
    close(RTCHandle);
}
```

RESULTS

Theoretically, we should achieve millisecond precision with an interrupt frequency of 1024 Hz. The graph in figure 4 shows the 5% trimmed mean of the method's measured absolute error. The sample population contained 127 outliers with values in the 20-50 millisecond range. They have a mean of 34.6 milliseconds and a standard deviation of 14.6 milliseconds. They represent 1.09 % of the total sample population and are thus filtered out by the 5% trimmed mean norm showed in the graph.

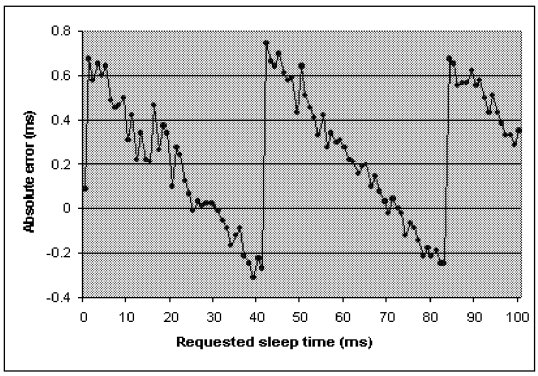


Figure 4: Absolute error on native sleep method using IRQ counts, in the 0-100 millisecond range.

For sleeps between 0 and 100 milliseconds, an accuracy within the 1 ms range is achieved. The graph shows periodic characteristics that are caused by the upward rounding of the necessary number of IRQ ticks for a given sleeping interval. For example, let I be the number of ticks required to sleep for a period of T milliseconds. For T=20, we have I=20*1.024=20.48. This results in a rounding error of 0.52 IRQ ticks or 533 μsec. The graph in figure 5 illustrates this theoretically computed rounding error. The measured results in figure 4 are less stable because of the limited timing granularity of `System.currentTimeMillis()`. This was confirmed by timing the native code using high resolution C timers. The results of these timings matched the pattern in figure 5 more accurately.

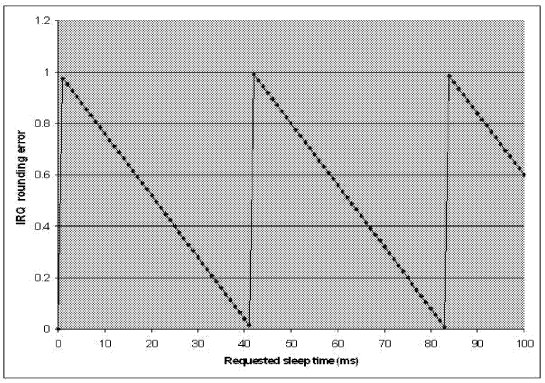


Figure 5: IRQ rounding error in the 0-100 millisecond range.

We have observed, as in the busy waiting case, that for large values of T, this sleeping method returns too early. This is illustrated in figures 6 and 7. A possible cause of this drift could be the fact that the RTC clock generates interrupts faster then 1024 Hz. This would also explain why, for values of T that require an integral number of IRQ ticks, the error is maximal instead of minimal. All values of T sampled in figure 7 require an integral number of IRQ ticks and should therefore result in a minimal error. The graph in figure 6 however, illustrates that for T=4000 the error is maximal. As in the busy waiting case, we can apply a correction in the native code using high resolution timers to achieve millisecond accuracy for large values of T.

This technique induces a CPU load of 0.3-0.7% on 1024 Hz interrupts, rising to 3 - 4% on 8192 Hz interrupts which provide higher accuracy.

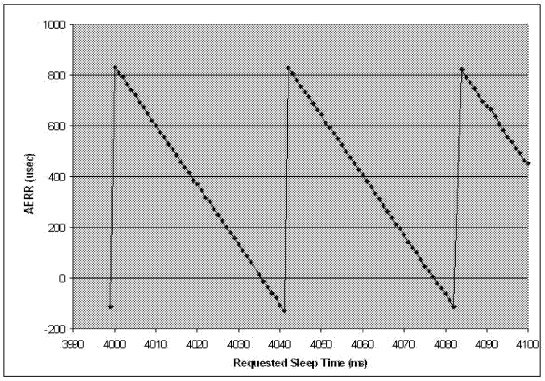


Figure 6: Absolute error on native sleep method using IRQ counts, in the 3999-4100 millisecond range.

OTHER METHODS

Apart from the methods presented, there are other possibilities to achieve more accurate sleeping behavior on a Linux platform. One could 'sleep with slack', i.e record 'overslept' time and compensate by subtracting this time from the next sleeping period. Of course, this method does

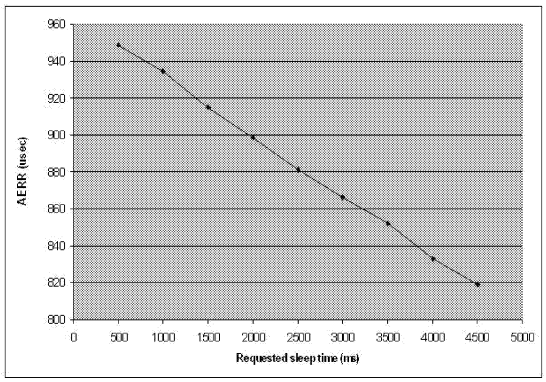


Figure 7: Absolute error on native sleep method using IRQ counts, in the 1500-4500 millisecond range.

not increase the accuracy of an individual sleep request. Another possibility is to pre-compensate, i.e. measure the standard oversleeping time and subtract this from the requested sleeping period.

Another possibility is to patch the kernel. The UTIME [5] patch developed at Kansas University achieves high accuracy without increasing system load by reprogramming the timer chip to generate an interrupt on the next foreseeable event. If such an event does not exist within a timeframe of 10 ms, the timer generates an interrupt at 10 ms in the future. The Kansas University Real Time Linux patch [6] uses the UTIME patch as a basis for providing real time process scheduling capabilities on a Linux system. It provides an extensive API for scheduling processes with real time accuracy.

The UTIME patch is still susceptible to the activity of system processes such as disk access and memory page management. It has been shown [4] that a UTIME patched kernel can still exhibit a maximum delay of 125 milliseconds on a periodically scheduled nanosleep of 50 milliseconds, when high disk activity takes place. To rectify this situation, the 'Low Latency' kernel patch [7] inserts additional preemption points into the Linux kernel at locations where long latencies may arise due to non interruptible system calls. 'Rapid Reaction Linux' [4] combines the UTIME and Low latency patches, achieving mean delays within the order of 30 microseconds, while keeping the maximum delay below 5 to 10 milliseconds under heavy disk load.

SUMMARY

The standard Linux kernel's scheduling delay varies from 10 to 20 milliseconds under low load conditions. Since most simulations require higher scheduling accuracy, alternate sleeping techniques had to be investigated. While several kernel patches exist that deliver high accuracy characteristics, they were deemed inappropriate because users are not always prepared or able to patch the kernel. Furthermore, their availability is also limited to a small number of supported kernels.

After reviewing three non-intrusive techniques, we conclude that the RTC interrupt count technique is the best choice. This approach is highly portable, accurate within the order of 1 millisecond, and has low CPU overhead.

*

REFERENCES

[1] P. Hellinckx, G. Stuer, et al : Dynamic Problem-Independent MetaComputing Characterisation Applied To The Condor System, *Proceedings of Modelling and Simulation 2003*, pp 262-269, Napels, Italy, 2003.

[2] F. Hancke, T. Dhaene, et al : A Generic Framework for Performance Tests of Distributed Systems, *Proceedings of the European Simulation Multiconference 2003*, pp. 180-182, Nottingham, UK, 2003.

[3] E. Gamma, et al : Design Patterns, Addison-Wesley, 1995.

[4] A. Heursch, H. Rzehak: Rapid Reaction Linux: Linux with low latency and high timing accuracy, *5th Annual Linux Showcase & Conference*, Oakland, California, USA, (2001).

[5] Kansas University: UTIME Patch - Micro-Second Resolution Timers for Linux, `http://www.ittc.ukans.edu/utime/`, (1997).

[6] Kansas University: KURT-Linux, Kansas University Real-Time Linux, `http://www.ittc.ku.edu/kurt/`.

[7] Ingo Molnar: Linux Low Latency Patch for multimedia applications, `http://people.redhat.com/mingo/lowlatency-patches/`.

TELE-COMMUNICATION SECURITY AND PRIVACY

E-LEARNING AS A TOOL: FRAMEWORK FOR BUILDING AN INFORMATION SECURITY AWARENESS PROGRAMME FOR A LOCAL TELEOPERATOR

Jorma Kajava and Rauno Varonen
University of Oulu,
Department of Information Processing Science, Linnanmaa,
FIN-90014 Oulu University, Box 3000, FINLAND,
E-mail: {Jorma.Kajava, Rauno.Varonen}@ oulu.fi

KEYWORDS: Information Security Awareness, e-learning

ABSTRACT

Being one the key skills in our modern society, learning and its counterpart, teaching, are currently becoming an important business area. Online education is assuming a central role in education, but at a cost: greater vulnerability to security threats. This paper brings together these two strands by discussing a framework for an e-learning system designed to provide information security education to the employees of a small telecommunication operator.

INTRODUCTION

In this presentation, we discuss teaching and learning in the modern environment, which is currently being transformed through the increasing use of distributed solutions. The Internet and its user-friendly interface are bringing real meaning to global co-operation in education. The globalisation process that has greatly altered business and manufacturing is now affecting education. Unfortunately, some of the problems that have inflicted commerce and industry are now beginning to be felt in education.

Information security has been thought of as concerning mainly business and commerce. As higher education was not commercial in nature, it did not attract uninvited third party interest. This situation is about to change, however, as e-Learning is rapidly becoming a big business at the global level.

A recurrent claim in the on-going security discussion in 2003 has been that information security is not an important topic in e-learning, as it only relates to wasted operating time. We do not agree with this kind of thinking, we want to define information security as a key function in future e-learning environments.

E-LEARNING AND SECURITY

Information security considerations usually start with the basic dimensions of security: confidentiality, integrity and availability [Parker 1981], but in the e-learning community this discussion tends to begin with copyrights. Who owns the materials used in e-learning? Are they the property of the individual authors, or perhaps of the educational institutes, journal publishers, newspapers or software producers? These are important considerations *per se*, but they should not overshadow other, security-related, aspects of the problem.

As a working environment, higher education is characterized by openness. However, the large-scale introduction of e-learning and the concomitant move into environments, such as the Internet, necessitate strong protection methods. What do these changes involve for openness and the much touted academic freedom?

And what is the nature of the Internet as a working platform? Who has responsibility over it?

Other significant security considerations relate to security guidelines, access controls, personal identification numbers, rules for network users, end-user education and security management. And there are problems caused by viruses, spam mail, hackers, intruders, criminals and terrorists.

Previously, the focal question in information security concerned the balance between usability and security. Now the main problem is the balance between security and the protection of privacy.

Without proper security solutions in place, e-learning loses much of its attraction and compromises its value. Thus, e-learning must be accompanied by appropriate security measures. Enhancing security requires that users of e-learning environments are aware of secure routines and follow given guidelines.

One emerging form of providing security education is through online programmes in e-learning environments [ISO-IEC 1995].

INFORMATION SECURITY AWARENESS

On the basis of a series of practical studies, we have come to the conclusion that there are several stages of awareness [ISO-IEC 1994, 1995]. Within practically every organisation, there are people at every stage, and the success or failure of IT security awareness correlates with either progress upwards or a relapse downwards. The stages of IT security awareness are:

1. drawing people's attention to security issues,
2. getting user acceptance,
3. getting users to learn and internalise the necessary information security activities.

The first stage includes drawing people's attention to information security-related issues and trying to catch their interest. The second stage involves user acceptance; having got the end-users' attention, it is important that security educators also get the employees to accept the IT security policy of the organisation. Finally, at the third stage, the end-users should have internalised the instructions they have received, and they should take corrective measures in accordance with the organisation's security policy. In this paper, the term 'awareness' includes all the aspects mentioned above.

IT security awareness should be comprehensive, well-organised and systematically executed from the start. In addition, the efficiency of all actions should be measured to ensure the on-going development of the organisational IT security awareness programme. As for our own IT security awareness programme in the university environment, we have come to the realisation that a wide range of tools and methods are necessary to implement security awareness for different people in different environments and at different stages of awareness. Security education is needed to convince every user of the importance of following guidelines and to make them aware of the consequences of intentional violations of information security. Education is also needed to ensure that the achieved level of awareness (as defined above) will be maintained. Various awareness raising methods, such as campaigning and the so-called Hammer theory, are needed to provide incentives for end–users and to refresh the importance of these factors in the minds of people. And finally, awareness, comprising education and training, should ensure that people internalise security guidelines and abide by them in their daily work.

The awareness programme of any industrial organisation should follow the framework developed in [NIST 1995]:

1. Identify Programme Scope, Goals and Objectives
2. Identify Training Staff
3. Identify Target Audiences
4. Motivate Management and Employees
5. Administer the Programme
6. Maintain the Programme
7. Evaluate the Programme.

The awareness programme should be targeted for at least four different groups. These are: top management, IS management, end-users and IT/IS specialists. Of course, there is no exact formal classification for these groups and, as a result, the end-user group, for example, remains relatively vague.

The IT security awareness programme should be implemented at all levels of the organisation, starting from the top management, who should be made aware of the need to establish and maintain an organisational security policy. Then comes the creation of a security model including IT security policies, the allocation of responsibilities, etc. IT security awareness programmes are essential in keeping users in the "security team" and in ensuring the overall success of the organisation's security strategy. All evidence shows that, to function satisfactorily, a security programme must find support in all parts of the organisation. Moreover, the top management has to accept security issues whole-heartedly and allocate resources and appropriate financial support to IT security. As stated earlier, the effectiveness of the measures taken during an awareness programme should be evaluated as objectively as possible. The problem is that most organisations do not provide feedback or measure the success of their programme. The security management should react to feedback and consider necessary improvements. Feedback should be based on organisational and end–user viewpoints and on the established results of particular security measures.

The educational dimensions of a security awareness programme should include at least the following points [ISO-IEC 1995]:

1. Objectives behind the corporate IT security policy as well as an explanation of the policy's guidelines and directives
2. A risk management strategy leading to an understanding of risks and safeguards
3. IT security programme/plan to implement and check safeguards
4. Basic information protection needs
5. Establishment of a classification system concerning the protection of information
6. Policy of reporting and investigating breaches of security or attempts thereof

7. Implications of security incidents for end–users and the organisation
8. Procedures, responsibilities, job descriptions
9. Security audit/compliance checks
10. Change and configuration management
11. Consequences of not acting in an authorised manner (including disciplinary actions).

THE INTERNET AS A PLATFORM

The convergence of computer and telecommunication systems created the Internet with its immense potential for enjoyment, recreation, learning and so forth. The educational use of the net has exploded during the past few years and continues to do so at an exponential rate. Web-based learning provides a valuable tool to educators who have not been slow to take advantage of the possibilities it offers. Online instruction opens up the world's largest store of information to students of all ages and comprises thereby an indefinitely rich teaching resource. But there is a snake in the paradise.

The Internet allows instructors and students to access its resources and to interact with each another in ways undreamed of ten years ago. But the openness of the net is also its Achilles heel: uninvited users, human or otherwise (viruses, for example), can also sneak in.

The Internet is not inherently a safe place. The basic reason for this state of affairs is that the net was not originally designed as a global information network. Rather, it has almost uncontrollably acquired its current functionality and evolved into its present size. The main focus has been on usability and ease of access with relatively little concern for security [Davis 1994, Pfleger 1997, Stallings 1995, 2000].

Further, the development of the Internet is spontaneous. No authority presides over the net or designs its further development. Consequently, it is evolving in several directions simultaneously. And no one knows what the end result is going to be, what the net is going to be like, say, in ten years time. Such lack of planning introduces an element of unpredictability into any long distance plans involving the net.

This evolution has been facilitated by the fact that no one is in charge of the Internet. Thus, no organization can be held liable for security on the net with the result that there has been no pressure to institute any measures to safeguard its security. Basically, users are at liberty to do whatever they want - as long as they don't get caught.

The global nature of the Internet adds to its vulnerability. Anyone with a malicious intent has a world wide playground at their fingertips. They can test their ideas and even solicit help from likeminded people, who may inadvertently become accomplices. And it doesn't stop there. People with a malicious intent may be far outnumbered by the simply curious, the thoughtless and the show-offs.

Another of the weaknesses of the Internet stems from its global nature. By crossing national borders, the net bypasses national legislations. What may constitute a crime in one country is not necessarily illegal in another country. Thus, even if authorities managed to track down the perpetrators of a particular offence, they may live in a country where the offence is not penalisable.

Currently, the responsibility for protecting information and assets lies chiefly with the owners of the information or assets. Large organizations tend to be well aware of the risks involved and have adopted appropriate measures, whereas smaller organizations sometimes try make a saving in the wrong place by taking perfunctory action.

Schools and other educational institutes often belong to the latter group. Security issues are not their first concern when installing computer systems and educational software. Our experience is that it is generally individual teachers who initiate e-learning experiments and conduct courses involving the use of the Internet. More often than not, these teachers are primarily interested in the pedagogical and didactic quality of their teaching, security issues are very much on the backburner.

What safeguards there are, are left at the responsibility of the institution. The same is true of university-based education, individual teachers and instructors are more worried about usability than security. Ensuring a proper level of security is the task of the administration and the university computer centre and has little bearing on the teachers. We believe that although large institutions such as universities, polytechnics and such usually have organized safeguards in place and are reasonably well protected against threats, also individual teachers and other personnel should have a basic understanding of necessary security issues and should follow secure practices.

Universities are generally associated with a great degree of freedom. The introduction of control measures translates into a reduction in the liberties academics are used to enjoying. Yet, such measures are essential for the operation of all academic and other organizations. We could use the term transparent university to refer to a university which has implemented computer-based controls. A case in point is a group of control measures known as deterrence functions. Their significance lies in the fact that if university students or employees do

something that is forbidden, this action can be traced back to them.
Control measures such as deterrence functions constitute a limitation of freedom. However, most working environments, especially universities, must have the highest level of trust to be successful. This trust can only be achieved through security. Thus, we have freedom on one hand and trust on the other, and they must be balanced against each other. Another balance has to be found between security and usability. The most difficult part of the equation, however, is to get the acceptance and cooperation of all employees and students.

One area of this research relates to the human mind. Training and education empower people's values, norms and tacit knowledge. The use of deterrence functions in computer-based controls paves the way for the transparent society. By accepting the fact that computers exercise a form of control over our society, we have the possibility to discuss the balance between controls and ethics as well as the balance between legal society and terrorism.

SMALL SCALE INDUSTRY EXPERIENCE

This chapter discusses aspects of an e-learning based information security project of a small organisation in the telecommunication service sector. What is special about this organisation is that it has outsourced all software development work. All 500 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 [Kajava 2003 (a, b)].

Understanding information security issues from the technical point of view is an advantage that these employees have. Nevertheless, since they do not have a wider perspective on other aspects of security, including 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. This chapter seeks to answer the question of what technical tools the company could use to solve this dilemma.

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 is 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 propose 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.

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.

CONCLUSIONS

Constructing an information security awareness programme for a small company requires mapping the initial knowledge level of the employees regarding security issues. On the basis of this mapping it is then possible to design a learning environment to meet their established needs.

An easy and often adopted approach is to bring in an external consultant. However, outsiders cannot solve all problems, they can only introduce a framework within which companies can try to work out solutions. A sustainable solution to information security issues is founded on the knowledge, skills and attitudes of company personnel. What an external consultant can accomplish is to extract these internal resources. In essence, it all boils down to trusting the abilities and experiences of personnel. Solving security problems and retraining a secure working environment is their responsibility.

A key factor in keeping the required resource level low is to automate the learning process. In small and medium-sized enterprises this involves working out core security issues, laying down best practices and presenting them in a way that everyone can understand regardless of level of prior knowledge or experience. In the project described here, an attempt was made to automate the learning process, measure the success of each individual employee through an automatic questionnaire and register the results in a database. The ultimate purpose is to develop

an online learning system that provides maximum learning experiences, while requiring a minimum amount of human intervention - apart from the efforts of the learners.

To sum up, the computer is just a device, but it can be transformed into a great facilitator of learning. It is important that this learning process is conducted in a manner that, in addition to being interesting and educational, respects the learners as persons.

REFERENCES

DAVIS, P.T., Complete LAN Security and Control. Windcrest/MacGraw-Hill. New York. 1994.

ISO-IEC-27, Guidelines for the Management of IT Security (GMITS): Part 1 – Concepts and models for IT Security. 1994.

ISO/IEC JTC1/SC27, Guidelines for the Management of IT Security (GMITS). 1995.

Kajava, J. : Security in e-Learning: the Whys and Wherefores. European Intensive Programme on Information and Communication Technologies Security, (IPICS'03). The Fourth Winter School. Information Society Technologies (IST/EU), Infotech Oulu and Oulu University. Series A, Research papers 32. 8 - 16 April 2003. Oulu, Finland.

Kajava, J., Pyy, J., Tuormaa, E., Heikkinen, I., Mukari, J., Rämet, T.: Information content for education produced by TiKo project. Oulu Telecom Plc. 16.5.2003. Oulu. (in Finnish).

The NIST handbook, An Introduction to Computer Security, NIST special publications in October. 1995.

PARKER, D. B., Computer Security Management, Prentice Hall, Reston, USA. 1981.

PFLEEGER, C. P., Security in Computing. Second Edition. Prentice Hall. Upper Saddle River. 1997.

STALLINGS, W., Network and Internetwork Security Principles and Practice. Prentice Hall. Englewood Cliffs. 1995.

STALLINGS, W., Network Security Essentials - Applications and Standards. Prentice Hall. Upper Saddle River. 2000.

E-LEARNING AND INFORMATION SECURITY. Evolutionary versus Revolutionary Approach

Jorma Kajava and Rauno Varonen
University of Oulu,
Department of Information Processing Science, Linnanmaa,
FIN-90014 Oulu University, Box 3000, FINLAND,
E-mail: {Jorma.Kajava, Rauno.Varonen}@ oulu.fi

KEYWORDS: Information Security, e-learning, Information technology.

ABSTRACT

Computer generations follow each other at 10-year intervals. These generations are marked by revolutionary advances in technology. Security solutions, on the other hand, are evolutionary in character. They consist typically of incremental improvements, usually in response to some threat. The main question posed in this paper is the path that e-learning should follow. Will security solutions be superimposed on e-learning platforms gradually as the need arises, or will security be an integral part of these new learning environments? The paper looks at the risks involved in adopting the former approach and discusses aspects of secure online education.

INTRODUCTION

By being implemented incrementally, by small steps, information security solutions can be regarded as evolutionary adaptations to changes in the environment. One reason for this piecemeal approach is that it is almost impossible to come up with a foolproof solution guaranteeing absolute security. As a consequence, the development of information processing is essentially a continuous process based on small improvements in earlier designs.

Information technology, however, has experienced a major change at about 10-year intervals. Following each change, computer technology has undergone a total transformation and the ways of exploiting computers have developed radically. In everyday parlance, these periods are known as computer generations. The step from one generation to another is a revolutionary process.

The development of IT can be likened to a race, where the competitors are rival IT companies, and the goal is nowhere to be seen. Sometimes the pressure to bring out new products has been so hard as to induce the competitors to sell their products before they are even finished or tested. On the other hand, this fierce competition can also be said to have led to the construction and implementation of information networks. These, in turn, formed the foundation for the development of information security.

Users of information systems have always appreciated usability more than security. Nevertheless, there is every indication that the importance of security needs to be reassessed. The question is, will there ever be a change of attitude? The lower value attached to security is reflected in the fact that security considerations tend to be placed on the backburner until something untoward happens, such as a malicious attack.

CAN WE TRUST TECHNOLOGY?

Each generation has suffered from security threats ranging from technological failures to software that is not up to scratch. As systems and networks become increasingly convoluted, the potential for security threats increases. In addition, the human users of information systems are re-emerging as the weakest point in the information processing chain. Staying abreast of all these developments requires constant change management.

In this process, new security solutions are an invaluable element. Currently, we were waiting to see whether the new Internet protocol, IPv6, with an impact on all protocol levels, introduces a better overall security level. The implementation of IPv6 can be regarded as a revolutionary process, because it requires that all software be written in a new way. Another keenly awaited solution was Public Key Infrastructure (PKI), but the emergence of a

standard took too long with the result that we now have a collection of dialects.

IPv6 involves a revolutionary step toward increased security at the application and system software level. In the long run, IPv6 will improve the situation, but in the short run, it requires investments of such magnitude that its comprehensive implementation seems unrealistic. As a result, other, incremental security solutions must be created to solve currently foreseeable problems. One possibility is to exploit biometric security solutions on a broader front.

In setting priorities for network and web technologies, computer generations have been largely ignored. To improve security standards, this trend has to be reversed for new technologies such as neural and quantum computers. Once they arrive, modern informatics will undergo a major upheaval on an unparalleled scale. The successful adoption of these technologies requires that security solutions are an integral part of their design.

SOFTWARE CHALLENGE

Producing software for network environments is a demanding task where particular emphasis has to be placed on security. Exposing and patching up security holes, although continuing to be a highly relevant and important activity, is insufficient as such. It is equally important to find working solutions to the processing of open and closed program code and to deal with ”grey” software areas. Part of this effort involves removing unnecessary modules from software packages.

Of paramount importance for the general development of IT is promoting security in the context of networked environments. This is definitely one of the main areas also from the viewpoint of e-learning.

Another major network-based development area is ubiquitous computing (ubicomp). This term refers to passive or ever-present computing. Although applications, such as ”under the skin” ubicomp for patient monitoring and smart-houses are at various stages of design and implementation, this technology will be available at a large-scale in the very near future - and the harbingers of that future, including certain types of smart clothes, are already here! It seems almost trite to say that the potential for misuse is unlimited. Examples include eavesdropping on devices monitoring the health status of dignitaries - or presidents. In the development of this technology, a major focal area must be security.

E-LEARNING AND SECURITY

What about e-learning, is it evolutionary or revolutionary in character? Before answering that question, we must understand changes in the modern educational environment. E-learning is rapidly developing into an important business area whose commercial potential attracts providers of education in hordes, among them a multitude of profit-seeking organizations. On the other hand, it is impossible to develop the structure and contents of on-line education without much less business-oriented organizations such as universities and polytechnics. (Epelboin, 2002).

Libraries will undoubtedly play a significant role in this development, as will, indeed, university computer centres that have already been assigned new responsibilities in view of the increased prominence of e-education. University departments and national virtual universities are in the process of designing and producing contents and e-learning materials. The scale of these changes that are gnawing at deeply rooted educational practices is such that e-learning must be considered as a revolutionary paradigm shift.

A change of this order inevitably prompts a range of questions, including security considerations. The question all educators and administrators must ask themselves is: What is the role of information security in online teaching and learning? Information security considerations usually start with the basic dimensions of security: confidentiality, integrity and availability (Parker, 1981), but in the e-learning community this discussion tends to begin with copyrights. Who owns the materials used in e-learning? Are they the property of the individual authors, or perhaps of the educational institutes, journal publishers, newspapers or software producers? These are important considerations, but they should not overshadow other aspects of the problem.

Secondly, as a working environment, universities are characterized by openness. However, the large-scale introduction of e-learning and the

concomitant move into environments such as the Internet necessitate strong protection methods. What do these changes involve for openness and the much touted academic freedom?

And what is the nature of the Internet as a working platform? Who has responsibility over it?

Other significant security considerations relate to security guidelines, access controls, personal identification numbers, rules for network users, end-user education and security management. And there are problems caused by viruses, spam mail, hackers, intruders, criminals and terrorists.

Previously, the focal question in information security concerned the balance between usability and security. Now the main problem is the balance between security and the protection of privacy.

Because e-learning is a global activity, we must adopt an international approach to the challenges it poses. One aspect of this involves harmonizing national legislations. Is it possible to find a common policy for the standardization of information security management at the global level? There is at least one possible solution to this question: the British Standard BS 7799 A Code of Practice for Information Security Management, which has become a de facto standard in a number of countries. In the US it is referred to as ISO-17799: BS.

Research exploring the relationship between e-learning and security is of paramount importance for a variety of reasons, including the following:

1. Crimes do not respect national borders.
2. Commerce and manufacture are networked and distributed.
3. E-learning is turning into a business activity.

The last point poses a host of challenges to universities and other institutes of higher education. One important point in this respect is legislation, another information security management (Kajava & Varonen, 2003 a).

CONCLUSION

Why are e-learning and information security dependent on each other? The answer is, of course, that without proper security solutions in place, e-learning loses much of its attraction and compromises its value. Thus, e-learning must be accompanied by appropriate security measures. Enhancing security requires that users of e-learning environments are aware of secure routines and follow given guidelines. Conversely, one emerging form of providing security education is through online programmes in e-learning environments.

Creating secure e-learning environments includes designing their structure and contents with a view on security aspects and necessitates a correct understanding of the character of information security. To avoid mishaps and misuse of learning environments, the security of the entire university environment must be enhanced and all users, particularly educators interested in online learning, must understand the importance of the secure use of the Internet (Kajava & Varonen, 2002 b). And they must act accordingly to minimize the potential use of their environment for other purposes, such as gaining unauthorised access to university networks.

We are still at a stage, where we have the option of selecting the evolutionary approach and of exploiting revolutionary periods in technological progress in a positive manner. Unless this opportunity is heeded, educators will soon face a new situation: a real revolution in which their environments are used for purposes other than educational.

REFERENCES

Epelboin, Y. : E-learning: putting documents on the web - Do and Don't. Workshop in the 8th Conference of European University Information Systems (EUNIS 2002). Proceedings. European University Information Systems (EUNIS) and University of Porto, Faculty of Engineering. 2002, June 19 - 22. 2002. Porto, Portugal.

Parker, Donn B.: Computer Security Management. Prentice Hall. 1981 Reston, USA.

A Code of Practice for Information Security Management. Department of Trade and Industry. DISC PD003. British Standard Institution. 1993. London, UK.

Information Technology - Code of Practice for Information Security Management. BSI ISO/IEC 17799:2000. BS 7799-1:2000. BSI. 2001. London, UK.

Kajava, J., Varonen, R. (a): Towards a Transparent University: the Role of Cryptography, Control Measures and the Human Users. In Eveline Riedling (ed.): VIEWDET'2002. Proceedings of Vienna International Working Conference - eLearning and eCulture. Austrian Computer Society (OCG) and Vienna University of Technology. 2003. Vienna, Austria.

Kajava, J., Varonen, R. (b): Internet Security and e-Teaching. In Eveline Riedling (ed.): VIEWDET'2002. Proceedings of Vienna International Working Conference - eLearning and eCulture. Austrian Computer Society (OCG) and Vienna University of Technology. 2003. Vienna, Austria.

A CONCEPTUAL FRAMEWORK FOR MONITORING INSIDER MISUSE

Aung Htike Phyo and Steven Furnell

Network Research Group, School of Computing, Communication and Electronic Engineering, University of Plymouth, Plymouth, United Kingdom
e-mail: nrg@plymouth.ac.uk
Web: http://www.plymouth.ac.uk/nrg

KEYWORDS

Intrusion Detection Systems, Insider Misuse, Role-based Monitoring.

ABSTRACT

Traditional Intrusion Detection Systems are ineffective in detecting users who abuse their legitimate privileges at the application level, because they do not have the knowledge of application level semantics, required separation of duties, and normal working scope. This paper outlines a novel framework for solving the problem of insider misuse monitoring. The approach argues that users with similar roles and responsibilities will exhibit similar behaviour within the system, enabling any activity that deviates from the normal profile to be flagged for further examination. The system utilises established role management principles for defining user roles, and the relationships between them, and proposes a misuse monitoring agent that will police application-level activities for signs of unauthorised behaviour.

INTRODUCTION

Many security incidents involve legitimate users who misuse their existing privileges, such that they have the system rights to perform an action, but not the moral right to do so. Current IDSs focus upon detecting problems such as network penetrations, access violations and privilege escalations. These tools are currently geared towards detecting attacks by outsiders, as well as insiders who employ the same methods to mount an attack. However, insiders may not need to exploit the systems because they already have legitimate access to it, and many incidents involve insiders only abusing their existing privileges (Audit Commission 1990), due to lack of separation of duties and application level control. Additionally current IDSs do not have knowledge of the normal working scope of a user for a relevant position and the separation of duties that should be enforced. Therefore, there is a need to provide the detection system with knowledge of organisation hierarchy and role-relationships in order to enable more effective monitoring. Role Based Access Controls (RBAC) (Ferraiolo and Kuhn 1992, Sandhu et al. 1996) utilises knowledge of role-hierarchy and role-relationships to make access decisions. This paper presents a novel framework that uses established role management principles used in RBAC to provide knowledge of organisation hierarchy and business process to the detection system. The next section briefly examines the nature of the insider misuse problem, leading into a discussion of the degree to which the detection strategies employed by traditional IDSs may be applicable. This section also introduces the potential for incorporating role based access controls, and the importance of role-relationship management. These ideas are then combined with the proposal for a novel framework for insider misuse detection.

THE PROBLEM OF INSIDER IT MISUSE

Insider misuse refers to users who have legitimate access to the IT systems and the data stored upon it, but abuse their privileges by using the resources in an inappropriate manner or for an unapproved purpose. Anderson (1980) classifies such users as 'misfeasors'. Computer crime surveys certainly suggest that one's own staff are a significant threat, with the results of recent surveys (Power 2002, Richardson 2003) by the Computer Security Institute (CSI) consistently suggesting that the dollar amount lost due to insider abuse is far greater than that of outsider attacks (e.g. the total losses over the last 6 years that were clearly attributable to outsiders were $46.5m, whereas the costs of insider misuse exceeded $220m).

Opportunities for insider misuse are many and varied (Phyo and Furnell, 2004), it is possible that appropriate use of traditional access controls could be used to prevent some of them. However, these will not be sufficient for all contexts (consider, for instance, the case in which the misfeasor has legitimate access to the payroll database, but modifies records to raise his own salary). One of the problems with insider abuse is that what users do with the system, or objects to which they are granted access rights, is neither monitored nor comprehensively logged most of the time. Different types of misuses can manifest themselves at varying levels of a system. Network access violations will show up at the network level, file access violations and application usage will be evident at the operating system (OS) level, whilst the user behaviour within the application environment will be most evident at the application level. Therefore it is important to collect the data for misfeasor analysis at the appropriate level in

order to increase the relevance of the collected data. The previous payroll example epitomises the case where data collected at the application level would provide more information about the user's intentions, when compared with the data collected at either the network level or the OS level.

Current IDSs are ineffective in detecting misuse of existing privileges. Access here might be just a simple read operation or modifying a database entry. Again, the users may access the resource in an unacceptable manner or for an unapproved purpose. Insider misuse is not only a technical problem, but also a managerial problem, because in some cases it is the improper segregation of duties that presented the opportunity to misuse (Audit Commission 1990). Therefore, in order to effectively monitor misfeasor activity, the monitoring system needs to have the knowledge of application level semantics, organization structure, separation of duties and user responsibilities. Coupled with this knowledge and monitoring at relevant levels of the system, a more effective system for detecting abuse of existing privileges may be designed.

APPLYING IDS TECHNIQUES TO INSIDER MISUSE

Traditional IDS employ two main strategies to identify attacks, namely misuse-based and anomaly-based detection (Amoroso 1999), and it is possible to see how each of these could be applied to the insider problem.

- *Misuse-based detection*: This approach relies upon knowing or predicting the intrusion that the system is to detect. Intrusions are specified as attack signatures, which can then be matched to current activity using a rule-based approach. A similar approach could potentially be incorporated for misfeasor incidents, based upon those methods that employees have been known to exploit in the past, or those that can be anticipated they would attempt based upon the privileges and resources available to them. For example, at a conceptual level, one such misuse signature might relate to a user who is identified as attempting to modify a record about him/her in a database (e.g. the payroll example indicated earlier). The rule here is that no one should modify their own records without someone else's authorisation. The problem with applying misuse-based detection to insider misuse is that the possible misuse scenarios for insiders are wide ranging and could be extremely organisation-specific.

- *Anomaly-based detection*: This approach relies upon watching out for things that do not look normal when compared to typical user activities within the system. In standard IDS, the principle is that any event that appears abnormal might be indicative of a security breach having occurred or being in progress. The assessment of abnormality is based upon a comparison of current activity against a historical profile of behaviour that has been established over time. One advantage insider misuse detection system has over outsider attacks is that it is possible to characterise normal activities of insiders according to their job position, as users with the same responsibilities should exhibit similar activities within the system and application environment to complete their daily tasks. The similarities may be profiled to represent normal behaviour for users with the same responsibilities, and different profiles for different job positions. If the user's behaviour deviates from the normal profile that represents his position, the activity should be flagged as suspicious. For example, a user who accesses a critical information system far more frequently than the other users within the same role may be browsing the database for personal gain.

Another problem associated in insider misuse detection is that current IDSs lack the necessary knowledge of business processes, organisation hierarchy, separation of duties, and the role of the users within the organisation structure. This knowledge needs to be expressed in the form that is understandable to the IDS, if effective misfeasor monitoring is to take place. Role management principles specified by Gavrila (Gavrila, and Barkley 1998) are utilised in Role-Based Access Control (RBAC) to support user role assignment, role relationships, constraints and assignable privileges. A role can be thought as a collection of operations required to complete the daily tasks of a user. In RBAC operations are associated with roles and the users are assigned to appropriate roles. This approach simplifies the task of assigning permissions to the user, as the roles for appropriate job functions are created with the least privileges required to complete the relevant tasks and the users are assigned to the role that reflects their responsibilities. Users can be assigned from one role to another, or assigned multiple roles, and permissions can be assigned at role-level to affect all users associated with the role. The type of operations and objects that can be controlled by RBAC is dependant upon the environment and the level at which it has been implemented. For example, at the OS level, RBAC may be able to control read, write, and execute; within database management systems controlled operations may include insert, delete, append, and update; within transaction management systems, operations would take the form that exhibit all properties of a transaction. The term transaction here

means a combination of operation and the data item affected by the operation. Therefore, a transaction can be thought of as an operation performed on a set of associated data items. The ability to control specific transactions, rather than restricting simple read and write operations are very important in database environments. For example, a clerk may be able to initiate a transaction and the supervisor may be able to correct the completed transactions, for which both users need read and write access to the same fields in the transaction file. However, the actual procedures for the operations and the values entered may be different Meanwhile, the clerk may not be allowed to correct the completed transactions and the supervisor may not be allowed to initiate the transactions. The problem is that determining whether the data has been modified in the authorised manner, for it can be as complex as the actual transaction that modified the data. Therefore, transactions need to be certified and classified before associating them with the roles. To characterise the required transactions for a role, duties and responsibilities of the users need to be specified first.

In RBAC, separation of duties can be applied by specifying mutually exclusive roles. In the RBAC framework administrators can regulate who can perform what actions, when, from where, in what order and sometimes under what circumstances. Access controls only allow or deny access to certain resources, however there is a need to monitor and analyse the user actions after the access has been gained and the operations had been carried out. In theory the idea of roles and role-management principles can be applied to misfeasor monitoring. Instead of allowing or denying operations to be performed, common user operations can be associated with roles, and the users can be assigned to appropriate roles. If the user's operations deviate from the common profile, a thorough investigation can be carried out to clarify if the user has misused the system in an inappropriate manner or for unapproved purpose.

MISFEASOR MONITORING SYSTEM: ARCHITECTURAL CONSIDERATION

It has been mentioned previously that anomaly detection is more suitable for insider misuse detection, because employees' normal behaviour can be profiled. For example, previous work in the DIDAFIT system (Low et al. 2002) has profiled database transactions by generating fingerprints for authorised SQL queries, along with variables that the users should not change, ensuring that the queries are executed in the expected order and only on the restricted range of records. It is assumed that the users with the same responsibilities within the organisation will exhibit similar activities within the system, and their working-scopes may be established. The idea of establishing working-scopes for users with same responsibilities has been tested in relational database environments by Chung et al (Chung et al. 1999). However, many of the insider misuse cases in Audit Commission (1990) surveys are a result of lack of separation of duties and application level controls. In order to be able to detect violation of separation of duties, the detection system needs to be provided with the knowledge of organisation hierarchy and relationships between roles. RBAC utilises role-relationship management principles to define role-hierarchy and separation of duties. The authors' proposed system aims to combine the ability of RBAC to provide knowledge of role-relationships with intrusion detection techniques to effectively detect users who abuse their existing privileges. Figure 1 presents the framework of the conceptual insider misuse detection system. Functional modules are explained in subsequent paragraphs.

Management Functions

All management functions, such as defining roles, characterisation of operations, association of operations to roles and user assignment to roles, are carried out from the Management Console. The working scope of a user is defined by the operations associated with the role(s) the user assumes. Once the separation of duties between roles has been defined, it is expressed in the Role-Relations Matrix, such as inheritance, static separation of duties, and dynamic separation of duties. Static separation of duties occurs at the role level by specifying mutually exclusive roles. When the two roles are in static separation of duties, a user may not be assigned both roles. Dynamic separation of duties occurs at the operations level and the conditions can be that operations within dynamically separated roles are:

- Mutually excluded
- Disallowed to execute concurrently
- Disallowed to perform both operations on the same set of data

When the two roles are in dynamic separation of duties, the user may not execute the operations that are mutually exclusive or on the same set of data. The relationships expressed in the Role-Relations Matrix are checked against the rules specified by (Gavrila, and Barkley 1998) for consistency.

Host

This is where the actual profiling of user(s) and the detection process takes place. Characteristics of each operation are stored in the *Operations DB* along with an appropriate name for each operation. The characteristics are dependent upon which level of the system they are being profiled at. Characteristics of the operations may be in the form of file access, sequence of system calls, SQL queries, API calls, User interactions, and Network access.

Recording the characteristics of each operation is controlled from the *Management Console*. The profiling should be done at all three levels of the system namely: network, system, and application level. At the network level, roles should be profiled based on the essential access to subnets in order for the users of the role to complete their daily tasks. At the system level, roles should be profiled on the use of applications required to complete the tasks. It should also be established which machines the users of the role can/cannot perform the task from. Again, at the system level, roles should be profiled based on what files need to be accessed in order to complete the task, along with the access mode and the application/process from which the files are accessed. Once the user has gained access to the file, and if the file is accessed from an application in which the file can be modified or manipulated (e.g. Databases), the application level monitoring should commence. At the database level, user queries and the associated values should be monitored. The problem is that determining whether the data has been modified in the authorised manner, for it can be as complex as the actual transaction that modified the data. Therefore, transactions need to be certified and classified before associating them with the roles. The *Detection Engine* then checks the roles available to the active user, and next checks the *RoleOperations* table for the names of the operations available to the user. After which the characteristics of the available operations from the *Operations DB* are compared to the current user actions. If current user actions do not match the characteristics of operations available to the user, the administrator is alerted. This alert may indicate the user performing a totally new operation, or performing a valid operation in the *Operation DB* but is violating separation of duties because the operation is not listed under any roles the user may assume.

The envisaged detection flow is as follows:

1. Detection Engine gets the name of the user from the Client. Looks for the roles the user's name is associated with, in the Role-User table.
2. After acquiring the list of roles for the user, the Detection Engine looks for the names of the operations associated with each role in the Operations DB. (Note: only names of the operations are associated with the Roles.)
3. After acquiring the names of operations available to the user, the Detection Engine reads the characteristics of available operations from the Operations DB and they are compared against current user actions.
4. If the current user action matches with the characteristics of operations available to the user, then the user is not in breach of static separation of duties.
5. If OpA belongs to RoleA, OpB belongs to RoleB, and RoleA and RoleB are in dynamic separation of duties. Condition of the separation is checked to clarify whether the operations are:

   - mutually excluded
   - disallowed to execute concurrently
   - disallowed to perform both operations on the same set of data

If the user violated the specified condition, the system security officer is alerted. In addition, the misuse rules employed in expert systems within traditional IDSs can also be included. These rules may then be associated with an operation, such as modifying the payroll database to increase one's own wages. In this case, the process is as follows: If modification is performed on the payroll database, check that the employee ID of the user is not the same as that of the record being modified.

Client

This is where the actual data is collected and transferred to the Host for analysis. The *Clients* can be network server systems or end-user workstations. The nature of the data collected may vary depending on the type of the *Client*. For example mail logs can be collected from the mail server, user queries from the database server, and application logs from user workstations. The data to be collected is specified by the system administrator from the *Management Console*. The collected data can then be refined to a standard format by the *Communicator* module before sending the data to the *Host*, so that data from heterogeneous *Client* systems is in a standard format. The *Client* may also have a *Responder* module to respond to detected incidents, and the appropriate response for each incident can be specified from the *Management Console*. For example, when a misuse is detected, the *Responder* may be configured to terminate the user session, revoke privileges, deny further access, alert the security officer, or terminate the anomalous process (Papadaki et al. 2003).

Implementation Issues

In order to be able to implement the system successfully, separation of duties would first need to be defined at the organisation level. Before doing this, the responsibilities of the users need to be defined. Then it needs to be checked that the operations a user is allowed to perform would not lead to a successful misuse. All of these are more of a managerial (rather than technical) issue. However, these are not trivial and could require considerable amount of time and labour. Again, at a technical level, monitoring of user behaviour at application level may require modification of the software package if appropriate APIs are not included.

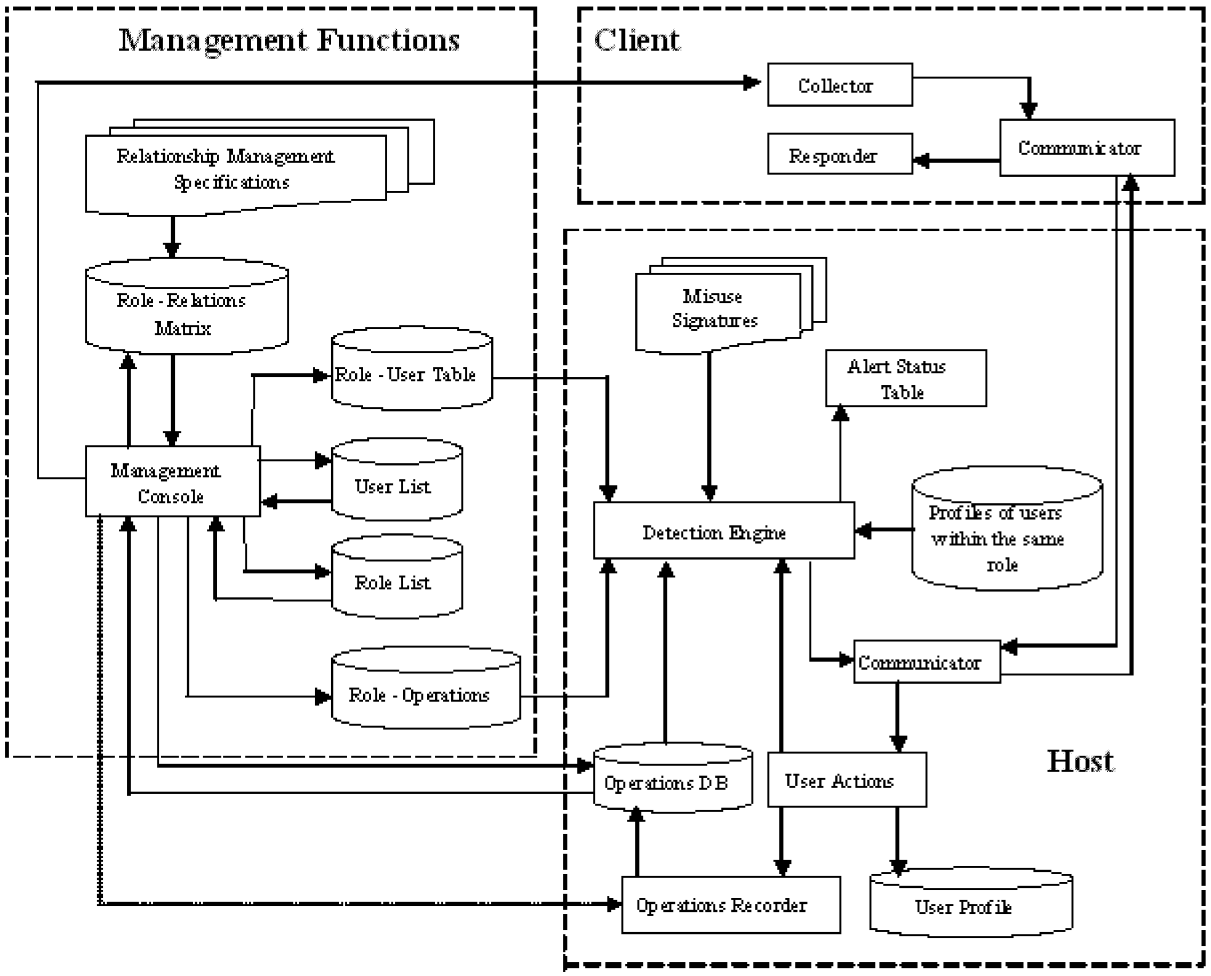


Fig. 1. Conceptual Framework of Misfeasor Monitoring System

CONCLUSIONS

Insiders pose a considerable threat and organisations need to give equal priority in detecting insider abuse as well as outsider attacks. Access controls only allow or deny access; however there is a need to monitor what the user does after gaining access to the system and objects. In order to effectively monitor privilege abuse, IDS require the knowledge of organisation hierarchy, managerial controls, responsibilities and working scopes of each user. The methods employed in RBAC to express knowledge of roles, organisation hierarchy, and separation of duties can be coupled with intrusion detection techniques to detect users who abuse their existing privileges. This paper presented a framework for monitoring users who abuse their existing privileges. The authors' future research will focus on developing the proposed system and testing it against a variety of simulated insider misuses, such as data theft, fraud, net abuse, sabotage, and breach of privacy.

REFERENCES

Amoroso, E. 'Intrusion Detection: An Introduction to Internet, Surveillance, Correlation, Traceback, Traps and Response', First Edition, Intrusion.Net books, NJ, ISBN: 0966670078, (1999)

Anderson, J.P. 'Computer Security Threat Monitoring and Surveillance', Technical Report, James P. Anderson Company, Fort
Washington, Pennsylvania, April (1980)

Audit Commission, 'Survey of Computer Fraud & Abuse: Supplement.
' Audit Commission, (1990)

Chung, C.Y. Gertz, M. Levitt, K. 'DEMIDS: A Misuse Detection System
for Database Systems', in the Proceedings of the 3rd International
Working Conference on Integrity and Internal control in

Information Systems, pp. 159-178. (1999)

Ferraiolo, D. Kuhn, R. 'Role-Based Access Control', In the Proceedings of the 15th National Computer Security Conference, pp. 554-563, Baltimore, MD, October 13-16 (1992)

Gavrila, S.I. Barkley, J.F. 'Formal Specification for Role Based Access Control User/Role and Role/Role Relationship Management', Third ACM workshop on Role Based Access Control, pp. 81-90, Fairfax, Virginia, October 22-23 (1998)

Low, W. L. Lee, J. Teoh, P. (2002) 'DIDAFIT: Detecting Intrusions in Databases Through Fingerprinting Transactions'. In the Proceedings of the 4th Internationsal Conference o Enterprise Information Systems, Ciudal Real, Spain, April 2-6, 2002,

Phyo, A.H., Furnell, S.M. (2004), 'A Detection-Oriented Classification of Insider IT Misuse', to appear in Proceedings of the 3rd Security Conference, Las Vegas, USA,

Power, R. '2002 CSI/FBI Computer Crime and Security Survey', Computer Security Issues & Trends, Vol. VIII, No. 1. Computer Security Institute. Spring (2002)

Richardson, R. '2003 CSI/FBI Computer Crime and Security Survey', Computer Security Institute. http://www.gocsi.com, Spring (2003)

Sandhu, R.S. Coyne, E.J Feinstein, H.L and Youman, C.E. 'Role-based access control models'. IEEE Computer, 29(2):38-47, February (1996)

Papadaki, M. Furnell, S.M. Lines, B.M and Reynolds, P.L. 'A Flexible Architecture for Automated Intrusion Response', Proceedings of Seventh IFIP TC-6 TC-11 Conference on Communications and Multimedia Security, pp.65-75, Turin, Italy, October 2-3 (2003)

CHALLENGE-RESPONSE PARADIGM IN ELECTRONIC MAIL

Pawel Gburzynski
University of Alberta
Department of Computing Science
Edmonton, Alberta, CANADA T6G 2E8
E-mail: pawelg@sfm.cs.ualberta.ca

KEYWORDS

Electronic mail, privacy, abuse, spam.

ABSTRACT

We point out the weaknesses of many popular approaches to eliminating E-Mail abuse (spam) and argue in favor of the challenge-response paradigm that, in our opinion, is the only viable way to address the problem at its present stage. We also present an E-Mail server (available under GPL) and a publicly accessible free demo service (see http://sfm.cs.ualberta.ca). Our server exemplifies desirable features of an E-Mail handling system that completely eliminates spam while providing for reliable legitimate human contacts including acceptable e-commerce.

INTRODUCTION

Proposed solutions to the spam problem, range from drastic legislative measures to revolutionary changes in the infrastructure of electronic mail. In this paper, we discuss the nature of spam and explain what *exactly* makes it different from "legitimate" E-Mail. In contrast to many other voices, we claim that spam is easy to spot and eliminate in a way that will put reliability, decency, and respectability back into electronic mail. This will require a relatively simple change in the paradigm of electronic correspondence which, as we demonstrate, can be accomplished without modifying any critical elements of the existing infrastructure. To get our point across, we present a ready publicly available software package that can be immediately deployed at the MTA (Mail Transport Agent) level. Our tool has been in operation for over a year now and has evolved into a reliable, friendly, and spam-proof E-Mail handling system.

The cause of spam

Electronic mail has brought about the first truly free and egalitarian tool for probing the applicability of the famous maxim of H.L. Mencken.[1] This is because, for all practical purposes, the cost of spamming is zero. This makes spamming quite different from other tools used for mass marketing, and this is also what turns it into a plague. Even a token or imaginary return from the free marketing of a scam or a semi-legitimate product makes the venture worthwhile. Breaking even is not a problem. But the spammers do better than that. A sales log intercepted on the network (McWilliams 2003) revealed the magnitude of income from a blatantly phony merchandise sold through moderately massive spamming. During a four-week period, the number of orders for a $50 bottle of penis enlargement pills reached 6000 (most people ordered more than one bottle). Considering that the cost per bottle to the merchant was about $15 (including the materials and the spammer's fee), the profit was hardly insignificant. Thus, the plague will not go away on its own. It does bring revenue to its champions.

Techniques for fighting spam

Generic solutions aimed at eliminating spam can be put into the following three categories: 1) anti spam legislation, 2) filtering, 3) reorganizing the E-Mail transport system. The most popular proactive approach is filtering, which occurs in two basic variants: text categorization, and collaborative filtering based on shared databases of various fingerprints of spam. Owing to its simple logistics, text categorization receives most attention in many practical implementations as well as in academic research involving AI techniques (for numerous examples see http://www.spamconference.org/), especially Bayesian filters (Gee 2003; Hindle 2003; Robinson 2003; Androutsopoulos et al. 2000; Massey et al. 2003), which many people see as a remedy for spam (see http://www.paulgraham.com/spam.html). Some newer variations on the theme include case-based filtering (Cunningham et al. 2003) and learning systems (Oda and White 2003).

The collaborative approach is exemplified by Vipul's Razor (see http://razor.sourceforge.net) and SpamNet (see http://www.cloudmark.com). Some commercial systems, e.g., Brightmail (http://www.brightmail.com), deploy bogus E-Mail accounts intentionally exposed for harvesting by spambots. A message arriving at such an account is guaranteed to be a spam.

More drastic proposals call for a revision of the present paradigm of electronic mail. Among them is the idea of implementing a payment scheme for the right to send an E-Mail message (Carnor and La Macchia 1998; Fahlman 2002), which would bring E-Mail marketing at least up to par with the traditional (paper) mass mailing. The Tripoli project (Weinstein 2003), described at http://www.pfir.org/tripoli-overview, outlines a comprehensive solution based on public-key encryption and certified tokens used for granting sending rights and authenticating senders.

Owing to the fact that the most radical proposals are incompatible with the present infrastructure, the practical solutions being deployed today are less revolutionary. They can be jointly categorized as sender-confinement schemes, whereby to be considered legitimate a message must arrive from a demonstrably trusted source, with the trust established through some kind of sender authentication. The simplest commercial solutions, e.g., Spamex (http://www.spamex.com), allow the subscriber to create multiple aliases to be given away to different senders. Some other services, e.g., Mailblocks (http://www.mailblocks.com), maintain a single address of the subscriber, but associate with it a list of legitimate senders allowed to send E-Mail to that address. The first message from a new contact is bounced with a challenge intended to verify

1 No one in this world, so far as I know ... has ever lost money by underestimating the intelligence of the great masses of the plain people (Mencken 1926).

that the sender's address is legitimate.

Two non-commercial solutions of this kind, in addition to our system discussed further in this paper, are TMDA (http://www.tmda.net) and ASK (Paganini 2003) (http://www.paganini.net/ask). TMDA is implemented at the delivery point. Different senders are allocated different aliases of the recipient, similar to the idea presented in (Ioannidis 2003). An alias may expire on a given date or be restricted to a particular sender. The system also defines so-called *keyword addresses* (similar to addresses assigned by Spamex), which are not restricted a priori, but can be easily revoked when abused or no more needed. Unknown senders are challenged with a bounce and instructed to send a message to a dynamic confirmation address.

ASK guards the single address of the recipient with a whitelist and a blacklist. The sender's address is added to the whitelist if the sender responds to the bounced message (assuming that spambots do not reply to bounces). Owing to its simplicity, ASK can be implemented as a *procmail* script and causes little hassle to the subscriber.
A complex system functionally similar to TMDA is outlined in (Tompkins and Handley 2003). Its improvement over TMDA consists in postulating cryptographic signatures to authenticate senders (which TMDA achieves in a sense by signing the confinement attributes of its dynamic addresses) and insisting that the challenge be insurmountable to spambots.

Futility of anti-spam legislation

To people familiar with the technical aspects of the Internet, it is obvious that anti-spam legislation (Butler 2003; Weiss 2003), (also see http://www.spamlaws.com), is going to help little, if at all. First, even if declared illegal in the United States (or in any particular country), spam will continue to arrive from abroad. With the present convenience of acquiring disposable Internet domains and temporary IP addresses, whose jurisdiction is at best unclear, it is impossible to enforce a law that blocks messages with a certain content from arriving to subscribers within a given country. Many of the scams presently circulating in the network are provably illegal and punishable by law (e.g., the numerous pyramid schemes or derivatives of the notorious "Nigerian" money transfer scam), and have been so for many years with little negative consequences to the perpetrators.

Second, the trend with the anti-spam legislation in the United States is not to eliminate bulk E-Mail marketing but rather to define the framework of its legitimacy (Butler 2003). This attitude may in fact increase the level of junk mail in the network by legitimizing the kind of spam that complies with the rules. Following the Senate approval of the Can-Spam Act (http://www.spamlaws.com/federal/108s877.html), we immediately see a proliferation of new service providers specializing in laundering spam to make it conform to the law.

Futility of spam filtering

Spam filtering via text categorization is, in our opinion, little more than an academic exercise. People involved in this work assume that "spam employs a distinct tone and language that can be used to identify it" (Gee 2003). We claim that this is incidental and reflects the current intermittent stage of spam evolution rather than a fundamental property of abusive E-Mail marketing. For illustration, consider the following message:

Dear Son:

We enjoyed our visit very much, and I will shortly send you the pictures that we took on our way back home. The Shmodak 500 camera that you gave us is terrific: the pictures came out unbelievably clear and sharp.

Take good care of yourself,

Mother

and suppose that you have to decide whether it is spam or not. There seems to be something fishy about this message – it mentions the (bogus) brand name of a product – so, considering that our discussion is about spam, you may be inclined to put your bets on the latter. But the decision is not easy, even for a human being. Many TV commercials are not clearly distinguishable from the shows they interrupt, and one can argue that the best among them are the subliminal ones, i.e., least aggressive and least "commercial" in content.

Even if we agree that a spam message must sound ostensibly commercial, the spammer can always resort to encoding the commercial content in an image attachment. With this approach, the spammer need not worry about making the message itself subliminal. Moreover, it is easy to randomly disturb the image without affecting the encoded message. Such simple tricks, in addition to completely circumventing all filters based on text categorization, will additionally trick the collaborative filters driven by databases of sighted spam.

These simple ideas have not yet become overwhelmingly popular among spammers, but they will when the sophisticated (e.g., Bayesian) filters are deployed on any significant scale. The spammers will easily and quickly learn to circumvent those filters because, as we have argued, fooling them is far from posing a moderately challenging problem. They will accept the increased cost of doing their business because there will be no other way to spam. Note that the "cost" we have in mind is solely the amount of time spent by a program.

The correct definition of spam

The inescapable conclusion is that spam filtering is futile. This is because the whole concept of filtering is based on the wrong definition of spam. The definition assumed by the categorization-based filters is:

Spam is a message whose textual component includes words or phrases indicative of a commercial advertisement or offer and fitting certain patterns determined by a reasonably large corpora of messages collectively categorized as unsolicited bulk E-Mail by human recipients.

whereas the definition assumed by the collaborative filters is:

Spam is a message that has been sent in (nearly) identical copies to a significantly large number of different users.

Spam need not fit any of the two definitions, and the fact that most of it does fit them at present should be viewed as incidental. Thus, the above definitions do not cover the whole of spam. Moreover, they do not apply exclusively to spam. There is nothing wrong with people being genuinely

interested in Viagra®, refilling inkjet cartridges, or stuffing envelopes, and willing to exchange E-Mail on those topics. Also, one can think of legitimate (or even important) messages being sent in identical copies to multiple recipients, e.g., alerts, memos, bona-fide newsletters. For example, in a certain hospital in Toronto, an indiscriminately deployed categorization filter created a havoc by blocking, among others, all E-Mail that included the words "penis" and "prescription."

In our opinion, the only definition that captures the essence of spam is this:

Spam is a message with no human contact at the sending end who would be interested in the fate of its individual instances.

It accounts for the critical premise that makes spamming profitable: the sender of spam is not interested in actually contacting any single recipient, unless the recipient responds to the offer. If the sender were forced to personally (manually) send the message to every single recipient on the huge list, the whole procedure would suddenly become truly costly, and spamming would cease to make sense. Consequently, to prevent spam from entering your mailbox, you have to make sure that only human beings actually interested in contacting YOU in person can ever make it through the software guarding your privacy. Note that this may also apply to a program sending you E-Mail, as long as there is a human being behind that program that actually wants to get in touch with YOU.

How to eliminate spam

Many people believe that the key to eliminating spam is to enforce some form of sender authentication or certification, e.g., to verify the authenticity and validity of the message headers (Paulson 2003). The implicit assumption is that if the spammer is forced to reveal his/her "true" identity and operate "in full daylight," then 1) few people will be willing to put up with the shame, 2) it will be easy to track down spammers and enforce the anti-spam laws, 3) no respectable agency will want to certify a spammer's identity. We believe that this line of thought is naive and shortsighted, as all ideas relying on the decency of human race. First, there will never be a shortage of people ready to sell their reputation for not so big money. Second, as we mentioned above, the spam laws are unlikely to make a (positive) difference. Third, the "respectable" certifying agencies care little about moral issues related to the activities of their customers (or even themselves, e.g., try a Google search using the keywords "VeriSign abuse"). The spam problem is not a consequence of some minor deficiencies of SMTP (like the fact the message headers can be faked), but results from the openness of the underlying paradigm of electronic mail. Spam naturally exploits those deficiencies, but it can live and proliferate without them.

We claim that the only effective way to eradicate spam is to implement the kind of validation scheme that would put the human factor back into the operation of dispatching a message. We have to conclude that the only promising avenue is in the direction of challenge-response systems along the lines of TDMA, ASK, and Mailblocks. Their role is not to authenticate the sender or verify message headers but to make sure that the sender is a person rather than a program (spambot). A "person" can be formally defined as an entity capable of passing the Turing test (Turing 1950), although, due to the notorious incompetence of programs in certain areas (Ahn et al. 2003), the actual challenge presented to the sender can be quite trivial.

SFM: an outline

Our system, accessible at *http://sfm.cs.ualberta.ca*, is dubbed SFM for *Spam-Free Mail*. It operates as an extension to a standard (E-Mail transport agent) MTA that fully conforms to SMTP (Klensin 2001; Postel 1982). The main function of SFM is easy (typically automatic) generation of limited-accessibility, alternative E-Mail addresses pointing to the subscriber's *permanent* or *fixed* address which can belong to any E-Mail domain.

In principle, the permanent address of the subscriber need not be known outside the SFM server. This, however, is irrelevant from the viewpoint of spam elimination. In contrast to some other aliasing schemes (Hall 1998), the reliability of SFM does not depend on address secrecy. The permanent address, in addition to providing the forwarding target for legitimate E-Mail addressed to the subscriber, also plays the role of the user Id identifying the subscriber to the SFM server.

There are two main reasons why an alias created by SFM is immune to spam. First, it isn't published (exposed) but presented to a single contact (which can be a group of people). Second, it is restricted to the specific contact (a narrow population of senders). In the context of spam classification, this operation has been traditionally viewed as a tricky and unreliable component of any sender restriction scheme. What if a legitimate sender uses an alternative address? What if a legitimate sender passes the alias to another sender in a bona-fide attempt to forward your request to a more interested or competent person? What if the identity of a legitimate responder cannot be known at the time your are making the contact? To account for these issues, the confinement procedure for an alias is carried out as follows.

An alias is created *open*, and it remains open for a predetermined amount of time, e.g., two weeks. During that time, it will accept messages from everybody adding their senders to its *personalization list*, i.e., the list of authorized contacts. Then, when the open time expires, the alias becomes *closed*. From this point on, it will only accept E-Mail from the registered senders.

With this approach, the new contact is given ample time to identify himself/herself to the alias and shape its *personalization*. When that time is over, the personalization of the alias solidifies. If the alias is subsequently compromised, it cannot be used for mass mailing because it will only accept messages from a select group of senders (that cannot be known by the spammer). If those senders themselves decide to abuse the alias, the subscriber can easily delete the single alias without affecting other contacts.

A special type of address created by SFM is a *master*. The primary role of a master is to be published and exposed as a publicly available point of contact with the subscriber. A message arriving at a master is never forwarded to the

subscriber but instead treated as a *query*, i.e., request for an alias of the subscriber personalized to the sender. In response to this request, SFM sets up a new alias and bounces the message with simple instructions explaining that it should be re-sent to the alias. To eliminate automatic acquisition of aliases by spambots, the alias is shown in a CAPTCHA image (Ahn et al. 2003), as shown in Fig. 1 (also see http://www.captcha.net).

To get the message through, the sender has to reply to the received bounce substituting the presented alias for the recipient address. All the sender has to do is to copy the first segment of the address (the string *vathigof* in Fig. 1), as all the remaining components are already present in the sender address of the bounce.

Figure 1: An Alias Presented as a CAPTCHA Image

Another role for a master is to serve as a template for creating aliases. One idea behind having several different masters, which provide multiple publishable identities to the subscriber, is to associate different degrees of trust with those different identities. A typical setup involves two masters: one for contacts of a permanent nature (aliases created from this master never expire), and the other for intermittent and casual contacts (e.g., with aliases expiring after one month).

Organization of SFM

The interaction of all components in a complete SFM server is shown in Fig. 2. The three boxes circumscribed by the dashed line represent the SFM-specific components.

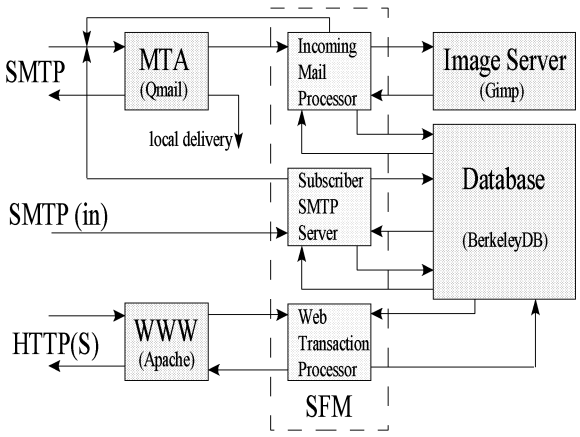


Figure 2: Organization of the SFM Server

Dynamic Web forms provide a secure authenticated interface to the subscriber's record. Through this interface, you can set up your personal attributes, create, view and edit your masters, look up your aliases and, in exceptional circumstances, create/edit them by hand.

Typically, once your SFM environment becomes set up, there is little need to access it via the Web. One exception is a quick (single-click) acquisition of a new alias, e.g., to be inserted into the Web form of an electronic merchant .

The sole task of the image server is to turn simple ASCII texts (aliased addresses of subscribers) into JPEG images. This way of presenting addresses to human contacts renders them resistant to automatic harvesting.

The alternative SFM-specific SMTP server implements only one direction of the SMTP protocol. It is available only to SFM subscribers and provides a means of dispatching E-Mail in a way that consistently and transparently presents their aliased identities to all contacts.

Reliability of contacts and e-commerce

The complete aliased address of a subscriber presented to the other party (see Fig. 1) consists of the proper alias name (the part before the first dot), followed by the master name, followed in turn by the mail domain name of the SFM server. The master name (the so-called *spice*) is not needed to reach the subscriber, but it provides a fallback measure in a situation when the alias has expired, has been removed, or is unavailable to the sender. In such a case, the message is treated as if it were addressed to the master: the sender is assigned a personal alias and informed about it via a CAPTCHA message.

This way of handling rejected messages provides for a graceful, reliable, and secure renewal of expired contacts. For example, you can create a short-lived alias and insert it into the Web form of an electronic merchant without having to worry about its possible abuse in the future. When the merchant decides to contact you after the alias has expired, SFM will make sure that the message arrives from a human being before renewing the contact.

SFM provides for an easy (one-click) alias acquisition for exactly this purpose: inserting an address into the Web form of an electronic merchant. The alias's name pops up in a new window as a complete address, which can be conveniently copied and pasted into a Web form or another document.

Mail processing

To relieve the subscriber of the responsibility for managing a potentially large and dynamic database of aliases, SFM comes equipped with an SMTP server which accepts outgoing E-Mail from its subscribers and translates their identities in accordance with the personalization of their aliases. To access this service, you must identify and authenticate yourself to SFM, which is accomplished by embedding a PIN code in your (permanent) sender address of the original outgoing message.

The authenticated sender address can be specified in one of two forms. In the simpler case, it includes the PIN code following the user name component separated from it by the + sign, e.g., *pawel+2357@dejunk.com.* Having received an SMTP request, the SFM server parses the sender address and verifies that the address identifies an existing subscriber whose PIN code stored in the database matches the PIN code included in the address. Only if this verification is successful does the server proceed with the request.

In the next step, the server collects all the destination addresses of the message and attempts to locate an existing

alias whose personalization list covers all the recipients. If such an alias is found the sender identity of the message is set to the that alias and spiced with its parent master. If no alias matching the destination list is readily available, SFM generates a new alias, adds all the recipients to its personalization list, and uses the new alias as the sender identity. Then, the message is forwarded to all its recipients.

In the second case, SFM must select a master for creating the new alias. By default the first master on the subscriber's list is used for this purpose. It is possible to explicitly indicate a different master by putting its identifier after the PIN code in the authenticated sender address, e.g., *pawel+2357+pgburzyn@dejunk.com*.

The simple PIN-based authentication scheme has been chosen as being compatible with all popular E-Mail clients, as well as the standard (unauthenticated) variant of SMTP being widely in use. It requires absolutely no effort on the subscriber's part, except for an initial configuration of the E-Mail client (MUA), and is quite secure within the framework of its application.

When the SMTP server of SFM receives an outgoing message from one of its subscribers, it looks up an existing alias whose personalization list includes the destination address. If the message is addressed to several recipients, all those recipients must be known by the alias before it is deemed suitable. If several aliases appear suitable, the one whose personalization list gives the tightest match to the list of recipients is selected. Then, the SFM server replaces the original sender address of the message with his/her aliased address (including the spice) and forwards the message to the destination.

If no alias fitting the recipient list is readily available, SFM will create one on-the-fly and initialize its personalization list with the list of recipient addresses of the message. The server avoids generating superfluous aliases; however, one should remember that each outgoing message presents a single aliased identity of its sender to all recipients. Consequently, the alias used for this identity must be personalized to all recipients. It is not uncommon that the same contact of yours will see several alternative aliases of yourself, depending on the configuration of group recipients of received messages. This poses no problem: any of those aliases can be used by your contact to reach you reliably and safely.

An incoming message intended for a subscriber may arrive on a master or on an alias. A message addressed to a master is bounced to the sender with a pertinent explanation. Two cases are possible. If an alias personalized to the sender already exists (the sender may have lost of forgotten it), the sender is reminded about the existing point of contact with the subscriber. If there is no ready alias personalized to the sender, a new alias is set up, and the sender is notified about it. In both cases, the address sent to the new contact is encoded in an image (Fig.1). The sender will have to resend the original message to the new address.

The processing of a message arriving on an existing and non-expired alias depends on whether the alias is open or closed. In the first case, the message is unconditionally forwarded to the subscriber and its sender address is added to the alias's personalization list. If the alias is closed, the sender address is verified against the personalization list. If the verification succeeds, the message is forwarded to the subscriber, otherwise, it is rejected. A rejected message whose recipient address is spiced is treated as if it were sent to the spice master. Thus, SFM generates a new alias (from the spice master) and presents it to the sender along with the bounced message.

Deployment

The primary concern of a site contemplating a transition to the new E-Mail paradigm represented by SFM will be the fate of the existing infrastructure of E-Mail addresses that have been heavily harvested and compromised. The following solutions to this problem are possible:

If SFM is installed within the E-Mail domain of an existing address infrastructure, the old addresses can be re-declared as masters. This will let them retain their official and traditional publishable status, while freeing them of spam, no matter how heavily they have been abused in the past. Owing to the fact that the MTA servicing the SFM address domain need not give up its traditional duties, this solution can be adopted gradually, as the users become convinced that they really want to switch to the new type of service.

Even if the server is installed in a different E-Mail domain, the old addresses can still be used as permanent addresses for SFM subscription. To de-spam them, the subscribers can deploy trivial and aggressive filters, e.g., blocking all incoming messages except for the ones arriving from the SFM server. This property is easily and reliably asserted via *filter cookies*, i.e., user-defined signatures inserted by SFM into message headers.

Most E-Mail clients can take advantage of the *refiltering* feature of SFM to completely de-spam an old E-Mail address while retaining its traditional and official status as a publicly known point of contact with the subscriber. This solution works with the SFM server installed in any domain, not necessarily within the domain of the old address. The necessary prerequisite on the client's side is the ability to filter messages based on keywords detected in headers and, conditionally, forward them to a special E-Mail address in a way that preserves the essential information from the original headers.

Usage

The present version of SFM has evolved significantly from its early prototype (Gburzynski and Maitan 2003). The most painful problem with the old version was the lack of SFM-specific SMTP service for processing outgoing E-Mail from SFM subscribers. To make sure that his/her identity in outgoing E-Mail was consistent with the personalization of aliases, the subscriber had to send messages to a single address within the SFM domain while passing the recipient information via special sequences in subject lines. That wasn't very friendly – the system, although useful to experts, was criticized as being cumbersome to an average user.

Another inconvenience with the old system was a large collection of arcane attributes associated with aliases and masters. The apparent need for those attributes, including complex patterns matched to the subject line and message body, was dictated by our (not always successful) efforts to account for different types of casual contacts in a way that

would make them as reliable as possible. The simple and amazingly powerful idea of keeping a new alias widely open for a limited time eliminated with a single stroke a large number of nasty problems and, at the same time, made all legitimate contacts 100% reliable. The configuration of alias attributes was reduced to a trivial number of easily understood items. As viewed by a non-expert user, the system became transparent and maintenance-free.

Summary

The general idea of a challenge-response protocol for establishing the first contact with an E-Mail recipient is sometimes criticized as cumbersome, unfriendly, unreliable, or impolite. Interestingly, we have had a chance to observe how the attitudes of people towards filter challenges evolve with time, or rather with the amount of spam that those people are forced to dig through every day. A few years ago, a message bounced with a challenge would occasionally meet with an objection from a mildly upset sender. These days, instead of objections, we are receiving words of appreciation and inquiries about our spam prevention tools. To put it in the right perspective, there is nothing wrong about a politely worded challenge after which the correspondence becomes noiseless, spam-less, and smooth.

The present author, together with many colleagues in his department, uses SFM for all E-Mail contacts, professional and business alike. We can recall no single complaint about a missing message, nor a single case of spam having sneaked through the system.

The proliferation of spam on the Internet has brought us a challenge, which we originally interpreted as a need to devise better filters in response to new spamming tricks. This is difficult and unfair. Ultimately, to carry out its duties without a mistake, a spam filter must be able to understand not only the message, but (as we argued in this paper) also the intentions of its sender. We cannot beat the creativity of spammers with mechanical filters, but we can easily reverse the problem and challenge the spambots instead. This will put the reliability, privacy, and respect back into our E-Mail contacts: the contents of our mailboxes will be shaped by people rather than programs.

REFERENCES

Ahn, L., Blum, M., Hopper, N.J., and Langford, J., "CAPTCHA: Using Hard AI Problems for Security," *Proceedings of EUROCRYPT'03,* pp. 294-311, Warsaw, Poland, 2003.

Androutsopoulos, I., Koutsias, J., Chandrinos, K.V., and Spyropoulos, C.D., "An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-Mail Messages," *Proceedings of ACM SIGIR,* pp. 160-167, Athens, Greece, 2000.

Butler, M., "Spam - the Meat of the Problem," *Computer Law & Security Report,* Elsevier, vol. 19, no. 5, pp. 388-391, 2003.

Cranor, L. and LaMacchia, B., "Spam!" *Communications of the ACM,* vol. 41, no. 8, pp. 74-83, 1998.

Cunningham, P., Nowlan, N., Delany, S.J., and Haahr, M., "A Case-Based Approach to Spam Filtering that Can Track Concept Drift," *Technical Report TCD-CS-2003-16,* Trinity College, Dublin, 2003.

Fahlman, S., "Selling Interrupt Rights: A Way to Control Unwanted E-Mail and Telephone Calls," *IBM Systems Journal,* vol. 41, no. 4, pp. 759-766, 2002.

Gburzynski, P. and Maitan, J. "A Comprehensive Approach to Eliminating Spam," *Proceedings of EUROMEDIA'03,* Plymouth, UK, April, 2003.

Gee, K.R., "Using Latent Semantic Indexing to Filter Spam," *Proceedings of the 2003 ACM Symposium on Applied Computing,* pp. 460-464, Melbourne, Florida, 2003.

Hall, R., "How to Avoid Unwanted E-Mail," *Communications of the ACM,* vol. 41, no. 3, pp. 88-95, 1998.

Hindle, R., "An Introduction to the Spambayes Project," *Linux Journal,* no. 107, 2003.

Ioannidis, J., "Fighting Spam by Encapsulating Policy in E-Mail Addresses," *Proceedings of NDSS'03,* San Diego, CA, 2003.

Klensin, J., "Simple Mail Transfer Protocol," *Request for Comments 2821,* Internet Engineering Task Force, 2001.

Massey, B. et. al, "Learning Spam: Simple Techniques for Freely-Available Software," *Proceedings of the USENIX Annual Technical Conference (FREENIX Track),* pp. 63-76, San Antonio, TX, 2003.

McWilliams, B., "Swollen Orders Show Spam's Allure," *Wired News,* Internet publication: http://www.wired.com/, August, 2003.

Mencken, H.L., "Notes on Journalism," Chicago Tribune, September 19, 1926.

Oda, T. and White, T., "Developing an Immunity to Spam," *Lecture Notes in Computer Science,* Springer-Verlag, vol. 2723, pp. 231-242, 2003.

Paganini, M., "ASK: Active Spam Killer," *Proceedings of the USENIX Annual Technical Conference (FREENIX Track),* pp. 51-62, San Antonio, TX, 2003.

Paulson, L.D., "Group Considers Drastic Measures to Stop Spam," *IEEE Computer,* vol. 36, no. 7, pp. 20-22, News Briefs, 2003.

Postel, J., "Simple Mail Transfer Protocol," *Request for Comments 821,* Internet Engineering Task Force, 1982.

Robinson, G., "A Statistical Approach to the Spam Problem," *Linux Journal,* no. 107, 2003.

Tompkins, T. and Handley, D., "Giving E-Mail Back to the Users: Using Digital Signatures to Solve the Spam Problem," *First Monday,* vol. 8, no. 9, http://www.firstmonday.dk/, 2003.

Turing, A.M., "Computing Machinery and Intelligence," *Mind,* vol. 49, 433-460, 1950.

Weinstein, L., "Spam Wars," *Communications of the ACM,* vol. 46, no. 8, p. 136, 2003.

Weiss, A., "Ending Spam's Free Ride," *netWorker,* vol. 7, no. 2, pp. 18-24, 2003.

SECURING CLIENT–SERVER COMMUNICATIONS OF AN INTERNET AUCTION SERVICE

Alessandro Amoroso
Massimo Nanni
Università di Bologna
Dipartimento di Scienze dell'Informazione
Mura Anteo Zamboni 7, 40127 Bologna, Italy
E-mail: {amoroso, mnanni}@cs.unibo.it

KEYWORDS

E-Commerce, Internet auction, Security protocol.

ABSTRACT

We present the results of an extensive research on securing the client-server communications of an auction service.

The aim of our work was to identify COTS components that offer security properties, that are lightweight, and that are as much transparent to the client as possible.

We identified and intensively tested the TLS suite that slowdowns the server performance of about five times with respect to not secure communications.

SCENARIO

Current auction services over the Internet rely, in general, on a central auction server. Such a centralized approach represents a limitation with respect to the scalability of the system. Typically the increasing number of customers of Internet auctions suggests that scalability is a crucial issue that must be addressed in order to prevent overloading of an auction server. Moreover, since Internet was designed to provide a best–effort service, most of today's e–commerce systems can provide only soft or no real–time services [Arlit at al. 2001, Bapna et al. 2001, Kumar and Feldman 1998, Maxemchukc and Shurc 2001, Wrigley 1997].

In contrast, our architecture is both scalable and responsive. Specifically, for scalability purposes, our architecture supports the implementation of auction services based on geographically distributed servers, and allows load distribution among those servers. Thus, this architecture accommodates an arbitrary number of users (*i.e.* auction participants), and prevent the server overloading problem. In order to meet responsiveness requirements, our architecture supports the implementation of client–server binding policies, that allow clients to be bound to the servers exhibiting the shortest response time.

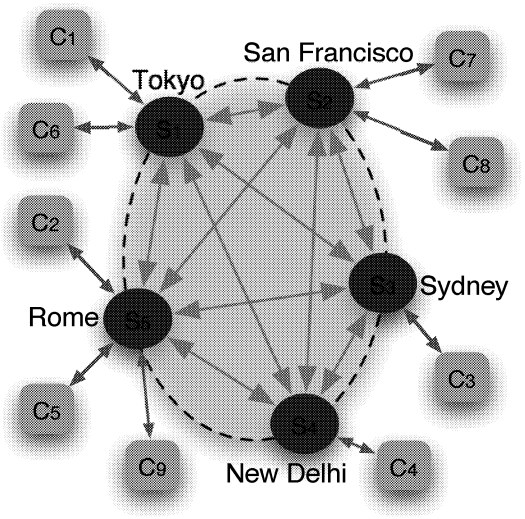


Figure 1: General Architecture of the Auction Service

The main idea of our auction system is that a set of servers cooperate as in a *periodic soft real–time system* in order to provide an auction service to the customers. During the auction progress the servers periodically synchronize to define the "best" bid they received in the current round, and then they asynchronously continue the collection of bids until the next synchronization phase. The servers do not need access to any kind of global timing service; we showed that the local physical clock is sufficiently accurate to maintain the periodic synchrony [Amoroso and Panzieri 2003].

Figure 1 illustrates an example of our architecture. In this figure, five auction servers are distributed in five geographically distant locations. Each client is connected

to its most responsive server. The group of servers is grayed because it provides a service. A detailed discussion of the communications within the group of servers is outside the scope of this paper.

The solid arrows between the servers in Figure 1 represent the communications within the servers' group. These communications are based on a reliable multicast protocol that provides the server with timely deliver of the messages, in the absence of failures.

The communications between the clients and the auction service, are represented by the red arrows in Figure 1. We wish to point out that, in order to guarantee consistency of the shared auction state, the client-server interactions need to be structured as atomic actions; *i.e.*, it must be guaranteed that the bid submission operation invoked by a client either terminates by correctly delivering the client bid to an auction server, or it has no effect (in either cases, the invoker of the operation is notified of the termination of the operation).

Our implementation has been developed in the Java programming language. This language offers platform independence; hence, it is possible to test the system using a wide variety of hardware resources. Even though Java has not been developed for soft–real–time distributed programming, in the auction distributed system we developed, the performance bottleneck is the network; the latency of the communication layer prevails against the overhead of the local programs; thus, Java has proved sufficiently adequate for our purposes.

The client–server protocol

Our aim in the client–server communication system is that the client side should be as simple as possible. Therefore, the solely requirement for a user that would like to participate to an auction is the availability of a Web browser with Java and JavaScript capabilities activated.

The client–server interaction can be summarized as follows:

begin of interaction: the client reaches a well known URL, such as the auction house's home page, and receives an active JavaScript page;

server selection: the JavaScript page, transparently to the user, scans a set of servers and select the one that exhibits the lowest response time;

enter the auction: the JavaScript page downloads a Java Applet from the selected server, the Applet is in charge of managing the client–server interaction;

observed state: periodically the servers broadcast the current best bid of the auction to their clients, those display that value to the users;

submit a bid: the user can submit a bid any time it likes;

leave the auction: when the auction terminates it notifies all the users; a user can also leave an auction any time;

fault tolerance: in case the server does not communicate the auction state before a timeout, the client assumes that the server is faulty and connects another server by means of the above described mechanism.

Security threats

In the following analysis we discuss the traditional security threats in the specific context of our application. We denote the client as the *participant* to the auction, and we denote the server, selected as shown in the previous section, as the *auctioneer*. According with [Stallings 2000], the kind of possible attacks to the *participant–auctioneer* communications can be summarized as follows:

Interruption: an attacker prevents the participant and the auctioneer to communicate each other; while the auctioneer can suppose that the participant left the auction, the client software will detect the communication disruption as a server fault, and will automatically select another server; non solvable problems arise when the attacker can interrupt the client communications with any of the servers. In the latter case the client cannot participate to the auction, or she/he is forced out from it. The worst case for the auctioneers would be a Denial of Service attack, DOS, but this case would involve millions of contemporary attackers, due to the high scalability of the auction system. We decided to accept the risk of interruption attacks.

Interception: an attacker can read the exchanged messages, but cannot alter the messages. In this case the attacker may know the winning bids in advance with respect to the participants.

Modification: an attacker can modify the contents of the exchanged messages. The participant may observe an auction state that is different from the state observed by the others, while the auctioneer may receive bids of different value with respect to the bids submitted by the participant. This kind of attack needs a real exchange of "original" messages to modify them; the auctioneer periodically sends the state of the auction to the participant, and the latter asynchronously submit bids to the auctioneer.

Fabrication: an attacker can made messages pretending that they are originated by the partner of the communication. As in the previous case, the participant may observe an inconsistent auction state, and the auctioneer may receive forged bids.

Replay: an attacher can send again a sent message. This kind of attack is useless in our system because any message is marked with an unique ID.

Repudiation: one of the two partners repudiates a message. It could be the cases that a participant disclaims a winning bid, or that the auctioneer deny the sending of an auction state.

The main countermeasure against the majority of those threats is the cryptography of the messages. The effectiveness of the countermeasure strictly depends on the kind of cipher method. At first cut we can assume that our system should be resistant to: eavesdropping, modification, forgery, and repudiation. The security method in our system has to be transparent to the user, *i.e.* it does not require specific software to be installed into the user system, both for simplicity of usage and for platform independence. The cryptographic mechanism should not involve any third party during its usage, to increase the responsiveness of the system. Moreover, the cryptographic mechanism should be computationally as lightweight as possible on both client and server side, to increase responsiveness and to avoid bottlenecks. Finally the cryptographic mechanism should be usable by means of the most popular Web browsers, specifically by means of the Java Applet that implements the participant–auctioneer communication.

CHOOSING OUR SOLUTION

In our system we decided to use COTS components to ensure security, as recommended in [Schneier 2000]. The main advantages of the COTS components are that: they implement the "state of the art" of security, they apply deeply studied algorithms, their implementation is well tested, and it is easy to plug them in a system.

We focussed on the following cryptographic mechanisms:

Java Cryptography Extension (JCE) [Sun Microsystem 2003] This is a programming suite of cryptographic algorithms included in the *Java Virtual Machine (JVM)* since version 1.4, released in 2002. This solution seems to be the most natural in our application because it is a Java standard feature, however it presents two order of problems. Firstly, the JVM 1.4 and subsequent are not yet supported by all of the Web browser, and several users do not systematically update their JVM to the last version (that is tenth of megabytes to download). Secondly, the suite provides the programmer with the algorithms, but the programmer has to implement "from scratch" all the procedures and protocols that make usage of those algorithms.

IPsec [Kent and Atkinson 1998] This is an Internet security mechanism that operates at the same layer of the IP protocol, therefore its usage is completely transparent to the applications that communicate through that layer. This protocol is part of the IPv6 standard, but actually it is not included in several operating systems, and its installation requires the skill of a power–user. Moreover, a correct configuration of the IPsec may be difficult. Several critics to the mechanism have been made in [Ferguson and Schneier].

Transport Layer Security (TLS) [Dierks and Allen 1999, Blake-Wilson et al. 2003] this is the standardization of the Netscape's SSL protocol. TSL operates at the *transport layer*, on top of the TCP/IP layer, and then does not require any manipulation of the communication stack. The applications are aware of TLS, then they need to be designed to take advantage of it. TLS is made by a two layers stack of protocols. In the lower layer there is the *TLS record protocol* that encrypt the communication. In the upper layer there are the *service* protocols that are functional to the record protocol. In the set of service there is the handshake protocol that establish an agreement between the two communication partners on the set of algorithm to apply.

The degree of security supplied by TLS has been proven in the years to be robust and without flaws. TSL offers the same security property as IPsec, such as *confidentiality* and *integrity* of the communication, while they both lack of*non repudiation* of the communications. Moreover, IPsec requires the authentication of the partners, while TLS usually does not make a client authentication. Therefore our participant–auctioneer communication protocol requires an *a priori* registration of the client to the auction house; the usual *username–password* couple identifies the user.

We adopted the TLS suite, that can be easily embedded in a Java Applet, resulting completely transparent to the user, and that supplies deeply tested implementations.

EXPERIMENTAL ANALYSIS

We would like to measure the decrement of performances brought by the introduction of TLS in the

participant–auctioneer communication. Our experiments were aimed to measure the number of requests that an auctioneer can answer in a second when TLS is in use. This parameter is crucial to assess the scalability of the whole auction system. As shown in Web usability literature, a client abandons a Web site if that is responsive above an eight seconds threshold.

We tested the following set of algorithm of the TLS suite:

RSA [Rivest at al. 1978] to exchange the *session key* between the participant and the auctioneer. This is a well known public key algorithm, whose copyright has recently expired, and that offer a high level of security. We adopted a 1024 bit key, that actually is considered robust enough for e-commerce purposes;

AES [Chown 2002] to cipher the communication by means of symmetrical cryptography, computed with the session key;

SHA-1 [Schneier 1996] to compute the cryptographic hash of any message.

To measure the performance of the auction server we largely adapted an Hewlett–Packard (HP) benchmark tool [Hewlett–Packard 2001]. The basic idea of the tool is that the server under test keeps a log file of the answered requests, and that several clients continuously request ciphered WEB pages. The HP tool explicitly demand that each request form the clients has to follow the complete algorithm: handshake, ciphered request end answer. This scenario consider that the request are independent each other, such as in the case of different clients.

In the scenario of our auction system we can reasonably suppose that it is not necessary to repeat the handshake phase at each request. Each client can make several requests in a limited interval of time, that is not large enough to compromise the session key. Moreover, the server periodically sends the auction state to the clients. Therefore we can reasonably suppose that the clients and the server can reuse the session keys. This simplification considerably reduces the computational time of the ciphering.

While the auction system is designed to be highly scalable with respect to the number of users, *i.e.* to support several thousands of users, it wold be impossible to emulate a use case. In our experiments we simulated a probable scenario. The auctioneer is simulated by a Java program that runs on a server, it replies to the requests of the participants, while it simulates the interaction with the other auctioneers of the system. Each participant is simulated by a Java thread that runs parallel in a bunch of Java programs; those programs run in parallel on different computers. The Java threads repeatedly submit bids to the auctioneer, wait for the answer from the server, and then submit a new bid. This is a very unlikely situation, but it represents an upper-bound to the worst case.

The computers running the participants and the auctioneer programs belong to the same $100Mb$ LAN, that supplies a network latency that is rarely measurable, and it is several orders of magnitude smaller than the measured response times of the programs. The computers that we used in the simulation share the same architecture, that is: CPU Intel Pentium 4, running at $2.4GHz$, $500MB$ of RAM, Linux Debian 2.4 operating system, and $100Mb$ ethernet card.

We tested the set of algorithms in the TLS suite that offers the higher level of security (RSA, AES, SHA–1). We compared the performance of that set with respect to those of a less secure set of algorithms (RSA, DES, SHA–1), the performance are strictly comparable. The performance of the set of the most secure algorithms are about 10% lower with respect to the ones of the less secure set, while the detriment in security of the latter is very high [Blaze et al. 1995]. Based on that observations, we used the most robust set of algorithms in all the experiments.

The first experiment compared the performance of the secured communications with respect to those of an un–secure communications. The latter case has been implemented by means of Java socket connections. The experiment has been conducted with five computers, one running the server, and the other four computer running 40 client threads each one. Figure 2 shows the results of the experiment. The X axis of the graph represents the time of the experiment, in seconds, while in the Y axis the number of requests that the server can satisfy per second at that time. The upper curve of the figure represents the socket connection, while the lower curve represents the TLS connection. Since we do not demand an handshake at each request, the performance of the TLS suite are roughly five times lower than the socket connections. The two curves increase at the beginning of the experiment, and then stabilize, because either the JVM needs to initialize the pseudo–random number generator, and the client threads does not start simultaneously. The gap between the two performance depends mainly on the computational cost of the cipher algorithm.

The variance of the TLS data of this first experiment is shown in Figure 3 second by second. Again, the results stabilize after the first seconds of the experiment. The analogous graph for the socket connection has the same trend.

In a second experiment we suppressed the server session cache, *i.e.* it was not possible to reuse the session keys, and each communication required the complete se-

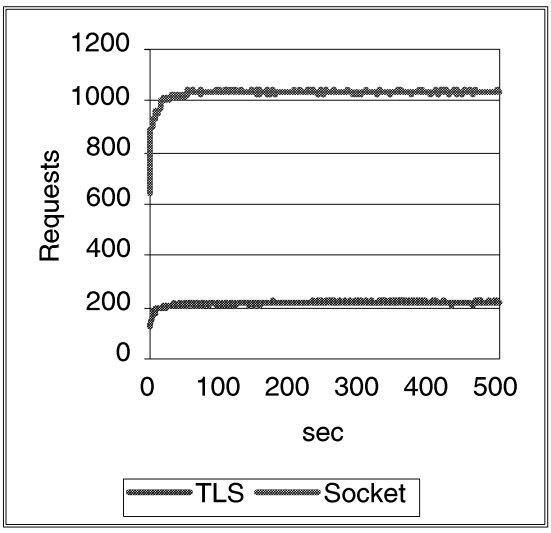


Figure 2: The Number of Requests that a Server Can Satisfy per Second

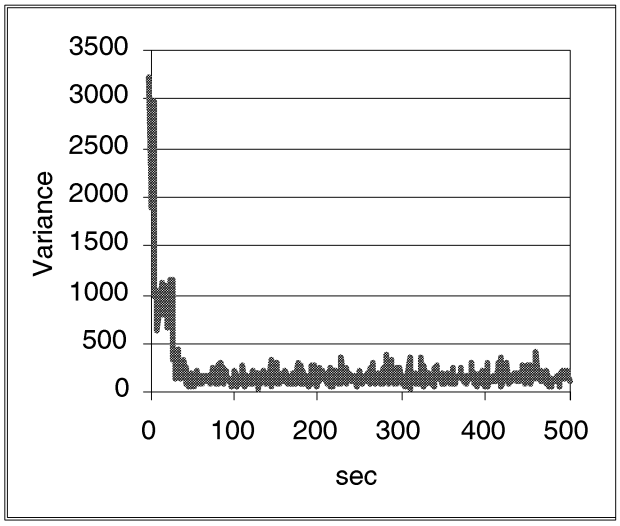


Figure 3: The Variance of the Data Shown in the Previous Figure

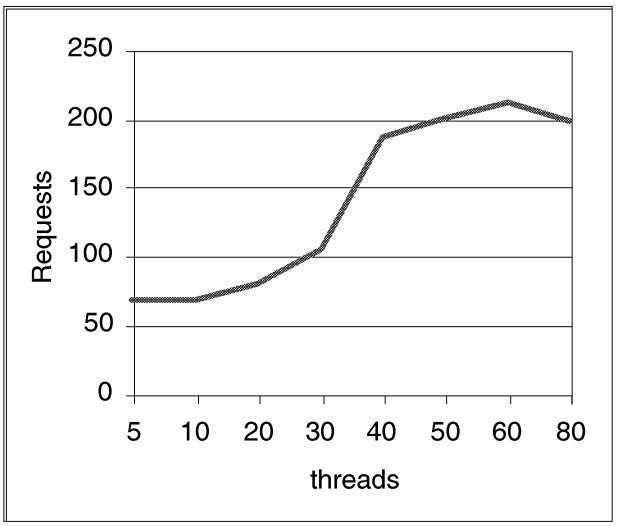


Figure 4: The Number of Served Requests per Second in Function of the Number of Clients

cure protocol. The results have the same trend shown in Figure 2 but the performance are about 25% lower. The variance of the data in this experiment is very high with respect to the one shown in Figure 3. It appears that it is not possible to fully control the parameters of the JVM, such as the session cache duration, and that some expensive processes, such as the generation of new keys, are "randomly" executed by the JVM.

In the last experiment shown in this paper we studied the transient phase of the previous experiments. Figure 4 shows the number of requests that the server can answer in a second by augmenting the number of client threads. The trend of the curve is monotonically increasing until the server get saturated, at 60 threads. Increasing the number of client threads beyond 60 causes the request to wait in the incoming queue.

CONCLUSIONS

The implementation in our auction system of the "state of the art" secure client–server communication protocols leads to a decrease of performance that is about five times with respect to un–secure protocols. We consider that this is a necessary price to pay to make a really usable system.

It is worthwhile to mention that choosing an high level of security, does not deteriorates the performance with respect to a lower level of security.

ACKNOWLEDGMENT

This work has been partially funded by the project *FIRB/"WebMinds"* of the Italian Ministry of Educa-

tion, University and Research.

BIBLIOGRAPHY

[Amoroso and Panzieri 2003] Amoroso A., Panzieri F.. "A Scalable Architecture for Responsive Auction Services Over the Internet", *Technical Report UBLCS*, University of Bologna, `cs.unibo.it/techreports/2003/2003-10.ps.gz`, 2003.

[Arlit at al. 2001] Arlit M., Krishnamurthy and Rolia. "Characterizing the Scalability of a Large Web–Based Shopping System", *ACM Transaction on Internet Technology*, vol. 1, August 2001, pp. 44-69.

[Bapna et al. 2001] Bapna R., Goes P. and Gupta A.. "Insights and Analyses of On-line Auctions", *Comm. of the ACM*, vol. 44(11), November 2001, pp.42–50.

[Blake-Wilson et al. 2003] Blake-Wilson S., Nystrom M., Hopwood D., Mikkelsen J. and Wright T.. *Transport Layer Security (TLS) Extensions*, IETF, RFC 3546, June 2003.

[Blaze et al. 1995] Blaze M., Diffie W., Rivest R. L., Schneier B., Shimomura T., Thompson E., and Wiener M.. *Minimal Key Lengths for Symmetric Ciphers to Provide Adequate Commercial Security*, Business Software Alliance, Chicago (IL - U.S.), November 1995.

[Chown 2002] Chown P.. *Advanced Encryption Standard (AES) Cipersuites for Transport Layer Security (TLS)*, IETF, RFC 3546, June 2002.

[Dierks and Allen 1999] Dierks T. and Allen C.. *The TLS Protocol Version 1.0*, IETF, RFC 2246, January 1999.

[Ferguson and Schneier] Ferguson N. and Schneier B.. *A Cryptographic Evaluation of IPsec*, Counterpane Internet Security Inc., 3031 Tisch Way, Suite 100PE, San Jose (CA - U.S.), February 1999.

[Hewlett–Packard 2001] Hewlett–Packard. *The New HP Benchmark for Measuring Encrypted Transactions*, `www.zeus.com/library/technical/hp_bench.pdf`, 2001.

[Kent and Atkinson 1998] Kent S. and Atkinson R.. *Security Architecture for the Internet Protocol*, IETF, RFC 2401, November 1998.

[Kumar and Feldman 1998] Kumar M. and Feldman S.J.. "Internet Auctions", proc. *3rd USENIX Workshop on Electronic Commerce*, Boston (MA - U.S.), Aug. 1998, pp. 49–60.

[Maxemchukc and Shurc 2001] Maxemchuk N. F. and Shur D. H.. "An Internet Multicast System for the Stock Market", *ACM Trans. on Comp. Sys.*, vol 19(3), August 2001, pp. 384–412.

[Rivest at al. 1978] Rivest R., Shamir A., and Adleman L. M.. "A method for Obtaining Digital Signatures and Public–Key Cryptosystems", *Comm. of ACM*, vol. 21(2), February 1978, pp.120–6.

[Schneier 1996] Schneier B.. *Applied Cryptography, 2nd ed.*, John Wiley and Sons Inc., New York (NE - U.S.), 1996.

[Schneier 2000] Schneier B.. *Secrets and Lies - Digital Security in a Networked World*, John Wiley and Sons Inc., New York (NE - U.S.), 2000.

[Stallings 2000] Stallings W.. *Network Security Essentials - Applications and Standards*, Prentice–Hall, Upper Saddle River (NJ - U.S.), 2000.

[Sun Microsystem 2003] Sun Microsystem. *Java Cryptography Extension*, `http://java.sun.com/products/jce/`, 2003.

[Wrigley 1997] Wrigley C.. "Design Criteria for Electronic Market Servers", *EM–Electronic Markets*, vol.7(4), 1997, pp.12–16.

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 area of Amoroso is the distributed systems. He participated to several scientific projects of National Research Council (CNR), National Energy Board (ENEA) and University of California at San Diego.

MASSIMO NANNI 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.

A MULTIMODAL WORKBENCH FOR AUTOMATIC SURVEILLANCE

Dragos Datcu, L.J.M. Rothkrantz
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
E-mail: { D.Datcu, L.J.M.Rothkrantz }@cs.tudelft.nl

KEYWORDS

Multimodal Fusion, Video and Speech Processing, Remote Processing.

ABSTRACT

Multimodal applications stand for the missing chain to overcome the limitations of classical multimedia processing tools currently used. Therefore data fusion is seen as a very active research field and is also set to grow in importance in the coming years. At Delft University of Technology there is a project running on the development of a software workbench with native capabilities for signal and information processing and for fusion of data acquired from hardware equipments such as microphones and video cameras. A first prototype of the system has been already developed. At the moment, one additional project aims to develop an automatic surveillance system by using only the resources of the workbench.

INTRODUCTION

Multimodal systems represent and manipulate information from different human communication channels at multiple levels of abstraction. More than that, they automatically extract meaning from multimodal, raw input data, and conversely produce perceivable information from symbolic abstract representation [Benoit 1998]. Current approaches to audio-visual data fusion commonly attempt to fuse the multi-sensory data at one abstraction level. Hence, up to date, issues of context-awareness and internationality, requiring the accomplishment of multimodal data fusion at different abstraction levels, have been largely avoided. Moreover, virtually all multimodal-data-fusion applications, including multimodal human-computer interaction and surveillance, apply context-independent interpretation-level data fusion.

To accomplish a multimodal analysis of multi-sensory data, which resembles human processing of such information, input signals cannot be considered mutually- and context-independent and cannot be combined only at the end of the intended analysis [D.L. Hall, J. Llinas 2001].

As applications, smart environment and ambient intelligent information systems are widely thought to be the coming of "fourth generation" computing and information technology [H.J.W.Spoelder et al. 2002]. Such systems are expected to appear in public places, work floors and home environments and to provide us with services ranging from information processing and entertainment to surveillance and safety.

Successful design approaches for data fusion architectures denote integration of high level features such as smart data management, parallelization and system scalability. The system that is being developed in the Department of Knowledge Based Systems at T.U. Delft is called **A.I.D.P.T.** and stands for Artificial Intelligence aided Digital Processing Toolkit. It descends from previous **I.S.F.E.R.** (Integrated System for Facial Expression Recognition) [M.Pantic, L.J.M.Rothkrantz 2000] project that has been backed up during time by a long-term research.

Since 1992 various studies and researches related to the problem of visual face analysis have been conducted. The initial idea was to create a modular system to fulfill the requirements of a highly configurable software toolbox for facial expression analysis. The core part of the system was made in such a way as to rely on a set of external routine packages for specific tasks. That included low and high level signal processing algorithms, holistic pixel based, deterministic feature based, reductionistic and hybrid methods and as well as neural network and fuzzy logic computing.

The major problem addressed by this project is the fusion of data from different modalities: audio and video. Solving this problem also requires a solution to the problem of ambiguity, redundancy and a-synchronicity of multimodal data streams.

RELATED RESEARCH

[J.C.Martin et al. 1995] describes a basic language for cooperation of modalities in a multimodal application. The event queue mechanisms and the communication between tasks and execution modules and also analysed.
[M.T.Vo, A.Waibel 1997] presents a toolkit for multimodal application development. It includes graphic tools that enhance multimodal component and integration modules with context free grammar support. Krahnstoever [N.Krahnstoever 2001] describes a multimodal framework targeted specifically at fusing speech and gesture with output being done on large screen displays. Several applications are described that have been implemented using this framework.
The W3C has set up a multimodal framework specifically for the web [J.A.Larson, T.V.Raman 2002]. This does not appear to be something that has actually been implemented by the W3C. It proposes Extensible Multimodal Annotation Markup Language (EMMA) as a set of properties and standards that a multimodal architecture should adhere to.
The field of multi-sensory data fusion has witnessed a number of advances in the past couple of years, spurred on by advances in signal processing, computational architectures and hardware. Nevertheless, these advances pertain mainly to the development of different techniques (e.g., Kalman fusion, ANN-based fusion, HMM-based fusion) for multimodal data fusion at a single abstraction level.
Existing approaches to multi-modal data fusion are usually limited to multi-sensory data fusion at a single level. Multi-modal fusion in literature has a strong focus on improving speech recognition and person identification. In the past couple of years our research group was already involved in several multi-modal fusion projects [M.Pantic, L.J.M.Rothkrantz 2000,2003].

MODEL

The performance of ambient intelligent information systems is not only influenced by the different types of modalities to be integrated; the abstraction level at which these modalities are to be integrated/fused and the technique which is to be applied to carry out multi-sensory data fusion are clearly of the utmost importance as well [H.J.W. Spoelder et al. 2002]. In contrast to the existing approaches to multimodal data fusion, which are usually limited to multi-sensory data fusion at a single level, the main goal of this project is to achieve audio-visual data fusion at different abstraction levels - from the numeric level of raw multimedia data streams up to the high semantic level.
Accordingly, two research aims are pursued in this project:

- (Low-level Fusion) The development of models and methods for sensing audio-visual data, coding/representing the sensed data. The format would allow further fusion stages of the pertinent observations of different types and also the realization of probabilistic fusion on the observed information.
- (High-level Fusion) The development of both models for behavioral patterns and appropriate methods for fusing and interpreting the pertinent audio-visual information conform the developed models. The models for behavioral patterns would represent individuals and/or groups of people in certain environments (home, public places).

As an application, a surveillance system can be seen as an intelligent system that is assumed to have an awareness of the spatial properties of the environment (locate objects and people, build maps), changes of those properties over time (tracking objects and people) and to have an awareness of the behavioral aspects (recognize intentions and emotions).
One software component for signal acquisition is the media client that is aimed to work as a bridge between the central system and each hardware peripheral attached. The software component for user interaction allows the user to access the central system to develop and execute programs by using a C like programming language.
The structure of the framework is based on the model of an initial project that aimed to develop an automated system for the analysis of human non-verbal communicative signals [R.J. van Vark et al. 1995].
The user has remote control of the centralized data processor. The concept of Terminal Server in A.I.D.P.T. context states that a Client Component is seen as a window through which everything from one side is taken to the other. There is no task to be executed on the client component. Instead all the computations are done centralized by the server.
The Data Components consist of a management mechanism to handle data in a distributed manner. There is no direct link between any of the clients, no matter the type (Figure 1). Only the Server Component is able to control data traffic to and from each client. There can be also connected one or more server systems to the existing one in order to improve the running performance.
In that case the servers perform synchronizations for tuning the general working parameters and task parallelization is also possible.
An important characteristic of the all system is that it is scalable in the way it acts differently in typical situations. As part of the program management module, A.I.D.P.T. central component includes a Garbage Collection Mechanism (G.C.) for efficiently handling data in the memory. In addition to that, there are also protection and security mechanisms included.

Data storage components have the purpose to safely handle data in a distributed manner. An adapted replication method is used to fulfill the requirements of further applications that might need such support [L.Gao, A.Iyengar et al. 2003].

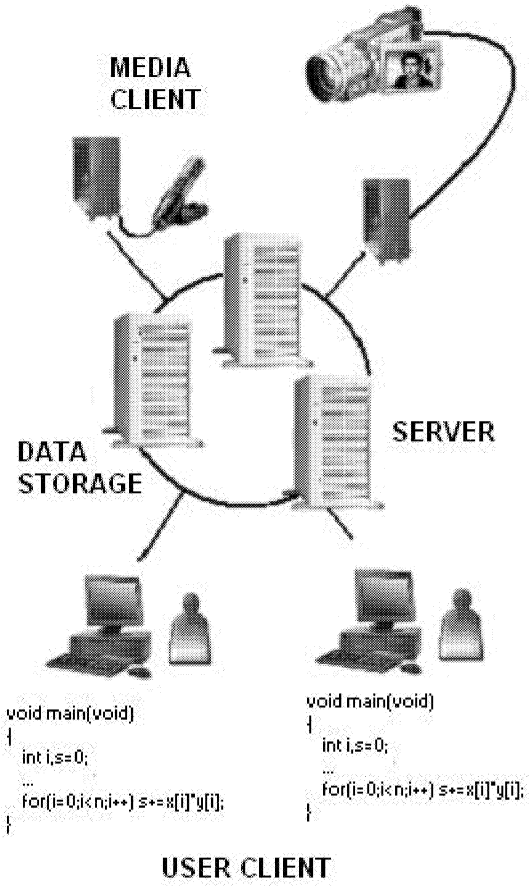


Figure 1: A.I.D.P.T. component diagram

AUTOMATIC SURVEILLANCE

Automated context awareness of where people are, what they are doing or plan to do enables application to support, protect, to take care of and assist people on a large scale in private or public environments. This approach enables new applications. However this technology also enables to transmit or store data (video, sound) only when a potential offence is detected instead of storing and transmitting all available data. In this way it leads to increased privacy compared to the current approaches. In this way automated surveillance in the private, home environment can be less obtrusive or embarrassing than human observation. Video surveillance is increasingly accepted by society in the trade-off between privacy and safety.
To show the relevance some cases will be considered, as possible applications of the technologies developed in the fusion project.

Surveillance: Observing individuals and groups of people using cameras and microphones is nowadays seen as a solution to increasing violence and costs of destruction. But observing people can be boring; a human can screen only a maximal number of monitors in a time consuming and costly way. Automated surveillance systems can be used on a large scale enabling more events to be detected and prevented.

Calamities: In case of accidents in tunnels or fire in buildings, visual and audio signals are used to rescue people. But sight and hearing of people can be serious disturbed in those situations. Smart devices can be used to guide people to the safety area. Such devices have to fuse audio and visual signals to detect and localise rescue signals and to find an appropriate route using visual/tactile cues.

Smart home environment: Context aware systems will be applied in future intelligent houses. Because of the multimodal sensing system, the intelligent house is aware of locations, identities and intentions of users and adapts its services (information services or physical services) more optimal to the users. In an optimal way, such systems will have an increasing importance in future care systems, since they allow elderly people to live alone in their own home environment for a longer period of time, leading to lower cost for elderly care.

IMPLEMENTATION

At the implementation stage of the project there have been used a few tools. The system consists of many software applications that are communicating with each other. The client components give users access to the available resources that are handled by the central component of the system running on Windows machines. For WEB configuration, a Windows version of the Apache server has been adopted and runs on the same machine with the system server. For DBS support, MySQL is server.
The lexical analyzer and the parser according to the defined grammar were generated using two common tools freely available, namely **flex** and **bison**.For writing the server component and the clients, the C++ language has been chosen. The PHP programming language was used to develop the modules for the scripting part related to the report and WEB configuration of the server.The A.I.D.P.T. server has a console interface and does not send messages to the screen. Instead, the initial configuration is read from a file on the disk and while in a running state, it writes all the notices in some existing log files on the disk. Besides there is an inverse connection from a WEB server's scripting modules to the server. The WEB interface is useful for generating various reports concerning the user statistics and the resource load states such as for the CPU and system memory. It also includes a module for changing the server's running configuration.

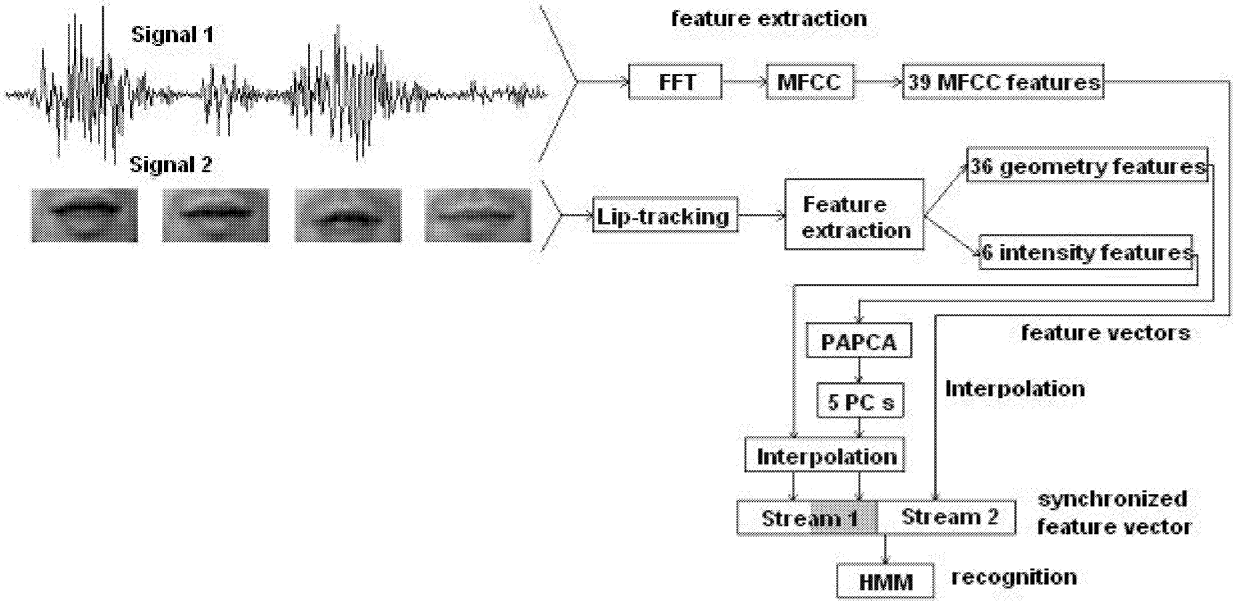


Figure 2: The feature fusion for a multimodal speech recognition application

There exists an internal server mechanism for choosing the mood the server acts for each working context. The decision is made taken into account the values of the parameters related to the given configuration and the resources load status. The server handles in the same time multiple user activity through an internal multithreading component. Each user is assigned with a virtually separated memory block for running specific tasks. When the CPU load is too high, the server can pass tasks to other similar servers in order to have them executed. To make such a situation possible, the servers must be initially properly configured.

Because of the fact that the system framework was designed to serve as a base for an automatic surveillance system, some security mechanisms has been added. These features relate to security and data protection mechanisms. Privilege-based access implies that each block in the internal memory or on any external disk has a user identification associated.

The other security feature relates to data encryption. Data protection aims to set hardware fault tolerance mechanisms and the server complies with that by integrating a module for data replication among different machines on the network.

The system processing routines are grouped around libraries and are accessible to the programming environment through Component Object Model (COM) interfaces that resides in Dynamic Link Libraries (DLL) files.

In addition to the routines that formed the I.S.F.E.R. signal and data processing library, there is new specific routine support for Bayesian Belief Networks and speech and video control. The new framework includes Bayesian Belief Networks routine packages because many new models for high level data processing are mainly based on this kind of computing. The support for B.B.N. is based on **S.M.I.L.E.** (Structural Modeling, Inference, and Learning Engine), a platform independent library of C++ classes for reasoning in probabilistic models [M.J.Druzdzel 1999]. S.M.I.L.E. is freely available to the community and is developed at the Decision Systems Laboratory, University of Pittsburgh.

APPLICATION

As an application, a multimodal speech recognition system has been implemented using A.I.D.P.T. That is part of a major project that concerns specific automatic surveillance systems. The model behind the processing tasks designed for the application is presented in Figure 2 [J.C.Wojdel 2003]. It has been used 13 MFCC features together with their deltas and acceleration values (a total of 39 features for an audio stream). For the lipreading part of the system Lip Geometry Estimation (LGE) extraction procedure augmented with intensity features has been used. The geometry data has been processed with Person Adaptive PCA (PAPCA) so that the resulting geometry features were transformed into the Person-Independent Feature Space. The projection parameters for PAPCA were estimated off-line beforehand for each of the persons for both the training and the validation sets. The lipreading-related data stream was sampled with a rate twice lower than that of the audio stream, so interpolation occurred every second observation in order to bring both streams in synchrony. Each state in each of the models was extended with the single Gaussian observation with its parameters set to the mean and variance of the whole lipreading data training set. The algorithm was more

advantageous because the already trained auditory part of the models would allow for better initial segmentation of the data and improve the speed of training the visual part. Such a fused system was retrained on the bimodal input.
Following the experiments, it has been proved that adding visual cues to a continuous speech recognizer results in better performance. The effect is much more accentuated in noisy environments. The best results so far were obtained by using a multi-stream Hidden Markov model with phoneme units for the speech stream and viseme units for the video stream. In the case of clean audio, the speech stream dominates the performance of the system. In the case of noisy audio, the relative performance of the system gets better as the weights of the video stream are gradually increased according to the noise level [P. Wiggers, J.C. Wojdel 2002].

CONCLUSION

The systems with multimodal capabilities present great advantages when comparing with previous unimodal approaches. Information is derived from sensors measuring different modalities of sound and vision. The purpose of the project is to develop a system to process data from those sensors that can be fused in order to obtain a consistent and robust model of the world.
In our project we defined an architecture for a complex framework that would provide the modality for developing multimodal applications. A particular case is the implementation of a surveillance system. We provided the functional framework with different efficient algorithms for data protection (by replication), security (by cryptography and privilege-oriented access mechanisms) and parallelization for distributing the computing the power over the network. The reasoning is done at different layers and by taking into account different features. The first results are promising, confirming the expectation that the sum of all the processing modules will be more than just adding the parts.

REFERENCES

C.Benoit et al.: 'Audio-Visual and Multimodal Speech Systems', 1998.

M.J.Druzdzel. 'GeNIe: A development environment for graphical decision-analytic models'. In Proceedings of the 1999 Annual Symposium of the American Medical Informatics Association (AMIA-1999), page 1206, Washington, D.C., November 6-10, 1999.

L.Gao, A.Iyengar et al.: 'Application Specific Data Replication for Edge Services', ACM. Hungary 2003.

D.L.Hall, J.Llinas: 'Handbook of multisensor data fusion' CRC Press, 2001.

N.Krahnstoever, S.Kettebekov, M.Yeasin, and R.Sharma: 'A real-time framework for natural multimodal interaction with large screen displays' In Proc. of Fourth Intl. Conference on Multimodal Interfaces (ICMI 2002), Pittsburgh, PA, USA,October 2002.

J.A.Larson, T.V.Raman: 'W3C multimodal interaction framework' http://www.w3.org/TR/mmi-framework, December 2002. W3C Note.

J.C.Martin et al.: 'Towards adequate representation technologies for multimodal interfaces' In international Conference on Cooperative Multimodal Communication', 1995.

M.Pantic, L.J.M.Rothkrantz: 'Automatic analysis of facial expressions: The State of the art' , IEEE transactions on Pattern Analysis and Machine Intelligence 22(12): 1424-1445, 2000.

M.Pantic, L.J.M.Rothkrantz: 'Expert system for automatic analysis of facial expression' Image and Vision Computing 18/2000, 881-905.

M.Pantic and L.J.M.Rothkrantz: 'Towards an Affect-sensitive Multimodal HCI', Proceedings of the IEEE, Special Issue on Multimodal Human-Computer Interaction (HCI), vol 91, no. 4, 1370-1390, 2003.

H.J.W.Spoelder, D.M. Germans, L. Renambot, H.E. Bal, P.J. de Waal, and F.C.A. Groen: 'A framework for interaction of distributed autonomous systems and human supervisors', IEEE Transactions on Instrumentation and Measurement, *51*(4):798-803, August 2002.

R.J.van Vark, L.J.M.Rothkrantz, E.J.H. Kerckhoffs 'Prototypes of Multimedia Stress Assessment' In: Proceedings of the MediaComm, pp108-112 SCS International. Ghent, Belgium 1995.

M.T.Vo, A.Waibel: 'Modeling and interpreting multimodal inputs: A semantic integration approach' Technical Report CMU-CS-97-192, Carnegie Mellon University, 1997.

P.Wiggers, J.C.Wojdel: 'Medium vocabulary continuous audio-visual speech recognition' In Proc. of (international conference on spoken language processing (ICSLP 2002), Denver, USA 2002.

J.C.Wojdel: 'Automatic lipreading in the Dutch language', PhD thesis Delft University of Technology 83-89003-62-7, 2003.

ANIMATION TECHNIQUES

FED – an online Facial Expression Dictionary

E.J. de Jongh, L.J.M. Rothkrantz
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
E-mail: L.J.M.Rothkrantz @cs.tudelft.nl

KEYWORDS
Multimodal databases, Image retrieval, Affective computing, Automatic recognition of facial expressions

ABSTRACT

People communicate with each other through spoken words and nonverbal behavior. Verbal communication is used to convey objective information, whereas nonverbal communication is used to convey subjective and affective information. Due to a number of reasons, confusion and misunderstandings can arise when people communicate with each other. A verbal dictionary can be used to look up the meaning and spelling of a certain word and can help people communicate verbally. The goal of this research was to develop an online Facial Expression Dictionary as a first step in the creation of an online Nonverbal Dictionary. A Nonverbal Dictionary would have the same functionality as a verbal dictionary and could help people communicate nonverbally.

1. INTRODUCTION

Human communication is based on verbal and nonverbal behavior. It is commonly assumed that natural language is used to communicate objective information to other people and nonverbal behavior is used to convey subjective and affective information [M. Pantic, L.J.M. Rothkrantz 2000, 2003]. Speech has a verbal and a nonverbal aspect. It is more appropriate to speak about the denotative and connotative aspect of multi-modal communication. The denotative aspect is based on the grammar and the connotative aspect is based on the rules of communication.
In [P.Ekman and W.F.Friesen 1975] P.Ekman introduced the concepts emblems and emotional emblems. The last ones are expressed by employing parts of the corresponding affects they refer to, while the first ones are used to replace and repeat verbal elements. Most of the time both are intentional, deliberate actions used to communicate. In general, they are produced consciously and are driven by the semantics of the utterance. They are conventionalised. Since they are discourse driven, the user enters their appearance. What is needed is a library of possible emblems. Efron gave a large list of them [J.R. Averill 1975] and Ekman proposes a set of words, which have a corresponding emblem. Nevertheless the user can build his/her own emblem and add them to the library. This paper describes the development of such a library: an online facial expression dictionary or FED for short.
In [P.Ekman and W.F.Friesen 1975] P.Ekman introduced 7 basic emotions, labeled as surprise, joy, sadness, disgust, fear, anger and contempt. He claimed that these emotions are universal, used by human people all over the world, with similar semantic interpretation. There are many other words, which have an emotional loading [J.R. Averill 1975].

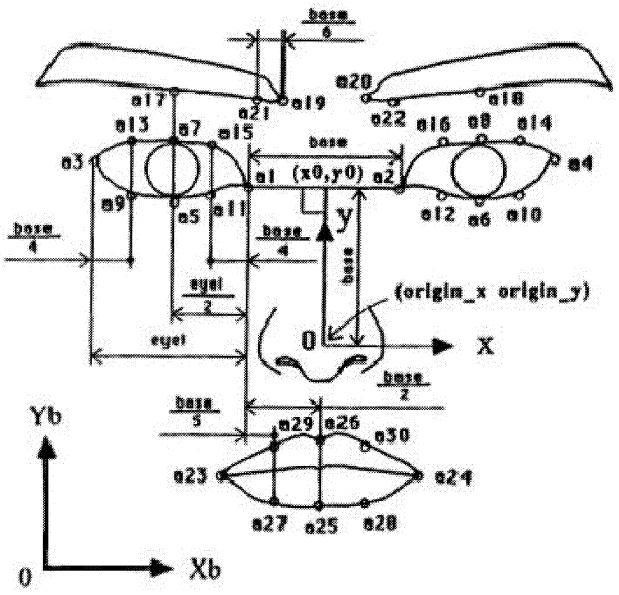


Figure 1: The Face Model developed by Kobayashi and Hara

About 300 of these words can be expressed by facial expressions. Usually these expressions are context and person dependent and here can be more that one way to express the word by a facial expression.

In a verbal dictionary the semantic meaning of words, spelling etc are described. Our facial expression dictionary provides the same facility for facial expressions. The "words" corresponds to facial expressions, "letters" to "Action Units" or "facial" muscles" and sentences to video sequences.
FED could be extended to a complete Nonverbal Dictionary [P.Ekman and W.F.Friesen 1975, C.Pelachaud et al. 1998], with an appropriate syntax and grammar, which would contain information about all the ways people communicate with each other nonverbally.

2. RELATED WORK

In 1970, P.Ekman and F.W.Friesen [P.Ekman and W.F.Friesen 1975] developed a universal Facial Action Coding System (FACS) based on Action Units (AU's). Each facial expression can be described in terms of active AU's. An AU is said to be active if the corresponding face muscle is flexed. FACS is used in FED to define a facial expression analogous to how a word in a verbal dictionary is defined by using characters. Also, an essential feature of a nonverbal dictionary is that it is possible to issue a *nonverbal query*. For FED, this means that it has to be a possible to issue a query using a facial expression itself. This could be accomplished by either letting users supply a picture containing a facial expression, or by letting users sketch a facial expression online by using a facial expression generation tool. Subsequently, FED would then determine the label of the facial expression. Implementing this type of query requires techniques from the fields of automatic facial expression recognition and multi-modal information retrieval. Each facial expression stored in FED will have to be represented in a way that enables the implementation of automatic facial expression recognition. The face model developed by Kobayashi and Hara [H.Kobayashi and F.Hara 1972] was used to represent a facial expression in FED. This model is based on 30 so-called *facial characteristic points* (FCP's), which describe the shape of the eyebrows, eyes and mouth (see figure 1).

3. FED – an online Facial Expression Dictionary

A corpus-based approach was chosen to create FED. Ideally, FED would contain a 24- hour view of the facial expressions of all the people on the planet. Since this clearly is unattainable, a simpler approach had to be taken. The basis for all entries in FED is a picture of a facial expression. This picture is either a picture of a facial expression taken from the real world or a picture generated by a facial expression generation tool. All FED entries will be stored in a database. Analogous to an online verbal dictionary, users will have the possibility to issue various queries into FED online. Furthermore, a FED administrator will have the possibility to manage the entries in FED. These management facilities can be used to create the FED corpus. Finally, FED will be set up in such a way that enables easy adaptations and extensions of FED in the future. There are several problems that need to be considered when trying to implement a facial expression dictionary:

- As mentioned in the previous section, not all facial expressions are universal. The meaning of a facial expression can differ per culture and context.
- Not every possible facial expression can be labeled / has a meaning.
- People can display a mixture of several facial expressions.
- Facial expressions are not always displayed to the same degree.
- Certain facial expressions can be displayed in a number of different ways. For example, the facial expression indicating 'happiness' can be displayed in a number of different ways.

4. Design

FED has been implemented as a website that handles query requests via a client –server architecture. Figure 2 shows the global design of the FED system. The individual components of the FED system can be described as follows:

- The *FED main website* enables users to issue queries into FED. It consists of static HTML pages and Java applets. The applets implement the GUI for issuing a query into FED. For each query, there exists an applet that implements the GUI for that query.
- The *communication layer* of the FED system resides on the server and handles all data traffic between the client and the server. The communication layer consists of a collection of Java servlets.
- The *Query Processing Module* or *QPM* of the FED system also resides on the server and consists of several modules, each of which has the ability to process a specific type of query. Each module is implemented through one or more static Java classes.
- The *FED admin website* provides the GUI for the management part of the FED system. Like the main website, it consists of static HTML pages and Java applets. This website is only accessible through user authentication.
- The *Admin Processing Module* or *APM* implements the functionality needed to manage the FED system. Like the QPM, the APM consists of several modules that in turn consist of one or more static Java classes.
- The *FED database* contains all the entries in the dictionary, admin user information, and log info.

The PostGreSQL database management system is used to implement the database. There are several advantages to this setup. Firstly, it ensures that the time to process a query is independent of the capacity of the computer of the user, because all calculations involved in processing a query are performed at the server, which is a high performance computer. Secondly, this setup increases the security of the website because all database access takes place at the server.

A third advantage is that by separating the GUI, communication and processing modules, the FED system becomes more easily adaptable and extendible. If all query processing, communication and GUI code is located in one applet, the code can become very complex and thus difficult to adapt and/or extend. Finally, by using Java servlets in the communication layer, it becomes easy to upgrade the capacity of FED, i.e. the number of simultaneous users the website can handle. This is possible through increasing the amount of resources available to the servlets.

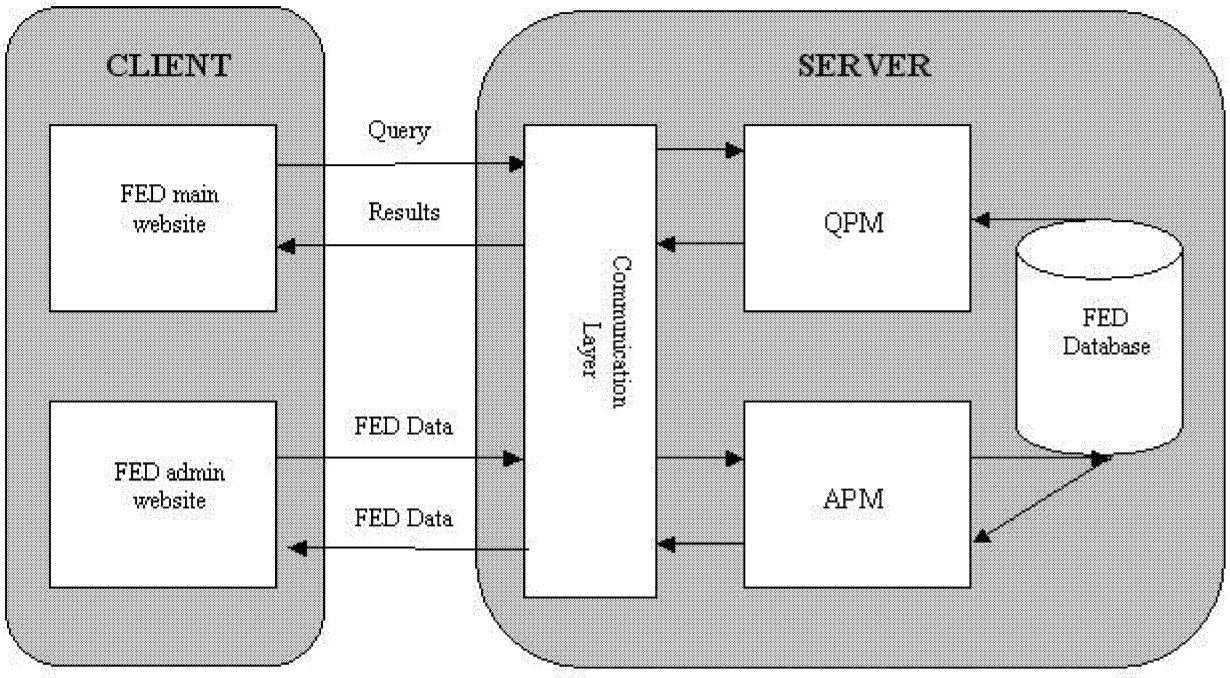


Figure 2: FED global design

5. FED entry implementation: Clusters vs. Templates

With FED, it is possible to let the system determine the label of an unknown facial expression shown in a picture or the label of a facial expression sketched online. This is accomplished by comparing the FCP coordinates of the unknown facial expression to the FCP coordinates of the FED entries. The closest match determines the label of the unknown facial expression. The performance of these queries depends on the way a facial expression is represented in FED. One way of representing a facial expression entry is through a cluster of different realizations of that facial expression. Each realization has it's own FCP coordinates and active AU's. Another possibility is to represent a facial expression entry by just one template expression. The latter approach can be described as a first-order solution, whereas the first approach can be seen as an extension of this method or a second-order solution.

As a first approach, the choice was made to represent each facial expression by just one template. Given the fact that this will not cover all possible realizations of that facial expression, this is likely to affect the performance of FED queries to a certain extend.

Implementing each entry through a cluster however would mean that there would be no time to implements all the basic functionality of FED.

6. DATA ACQUISITION

Research has shown that there exist approximately 7000 different facial expressions [P.Ekman and W.F.Friesen 1975]. The number of possible combinations of 44 action units is of course much larger, but not every combination constitutes an expression. Some action units are mutually exclusive, and other combinations form expressions that do not have any meaning. Ideally, FED would contain all 7000 possible facial expressions, each with a scientifically valid interpretation. The primary goal of this project however is to develop a prototype and to show FED is a viable concept of a facial expression dictionary. For each entry in FED, the coordinates of the FCP's have to be determined. This could be accomplished manually, but this can be a tedious and time-consuming process.Instead, FED makes use of a facial expression generation tool called FaceShop.

This tool enables a user to sketch a facial expression by changing the position of a number of slide bars. FaceShop calculates the coordinates of the FCP's automatically.
Other information associated with an entry in FED is the facial expression label, label synonyms, description, and example picture. Also associated with each entry are the active Action Units of the facial expression.

7. SEARCH MODALITIES

Analogous to a verbal dictionary, it is also possible to issue certain queries into FED. As can be expected here is a need for nonverbal search modalities. Assume the user has a picture of a facial expression or is looking at a face and wonders about the semantic meaning of the facial expression. So there is a need to fed in a picture to the FED system or a sketch of a facial expression.
The representation of FED entries through FCP's enables the implementation of a number of additional queries, not present with a verbal dictionary.

Label Query
As with a verbal dictionary, it is possible to issue a query on a keyword. In the case of FED, the entered keyword is matched against the facial expression label, label synonyms and words in the description of entries in FED.

Query on active AU's
With this query, the user selects the action units that are active in the facial expression(s) he is looking for from a list of AU's.

Query on geometrical features
The coordinates of the FCP's can be used to let users issue a query for facial expressions with certain geometrical features of the eyebrows, eyes and/or mouth present. Examples of valid queries are 'mouth open', 'eyebrows raised' and 'eyes slanting upwards'. Using the logical AND and OR operators, it is also possible to issue an aggregate query. Further restrictions can be imposed on a query by using the bracket operators '(' and ')'. An example of an aggregate query is 'eyes slanting upwards OR (mouth open AND eyebrows raised)'.

Incremental Query
It is possible to look up a certain facial expression incrementally. With this query, the user is presented with four facial expressions as representatives of 4 clusters covering the whole dictionary. The user then selects the best fitting picture. The selected cluster is again divided into 4 clusters and corresponding representative facial expressions that are again presented to the user, who again has to select the best match.

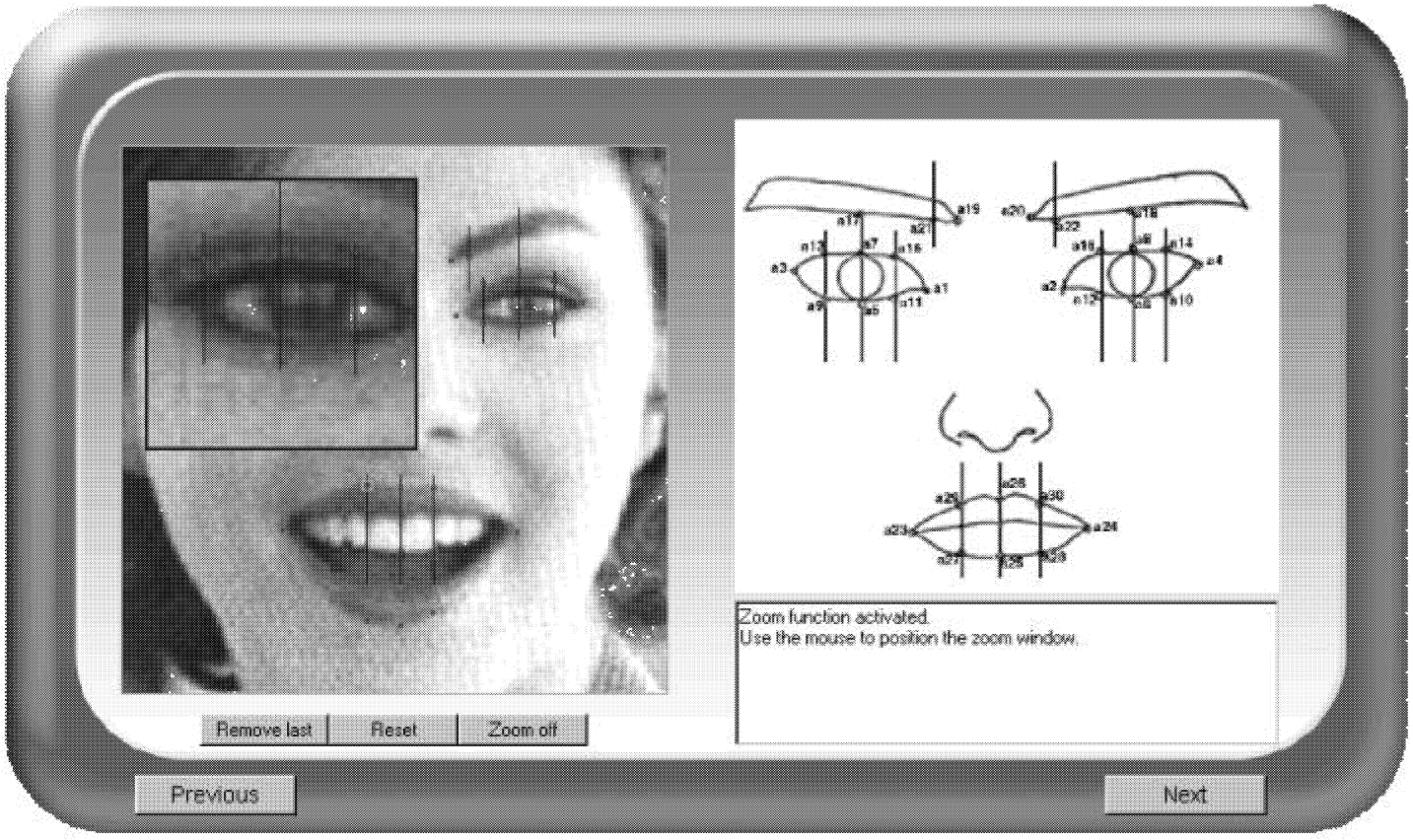


Figure 3: Determining the FCP coordinates semi-automatically

Labeling of an unknown facial expression

Users can determine the label of an unknown facial expression in two ways. First of all, it is possible to determine the facial expression shown in a picture supplied by the user. With this type of query, the user determines the positions of the FCP's semi-automatically. In [M. Pantic, L.J.M. Rothkrantz 2000] we developed a workbench, called Integrated System for Facial Expression Recognition (ISFER), which forms a framework for hybrid facial feature detection. To label an unknown facial expression, we have to attach to the picture facial template (Fig 1). Adapting the distance between the inner corners of the eyes can normalize this template. So we have to drag the FCP's of the template to the appropriate places on the contours of the eyes, mouth and eyebrows. Using the automatic mode with manual correction or setting of special points. Using known relations between the coordinates of the FCP's, the user is restricted to place each FCP in a for that FCP valid area. Furthermore, the x-coordinates of certain FCP's are fixed depending on the coordinates of other FCP's. Also, a zooming function is available. When all FCP's have been determined, FED will determine the closest matching entries. In our database a 60-dimensional vector of x,y coordinates of the FCP's represents all the labeled facial expressions. We compute the shortest distance from the vector of the unknown face to one of the labeled faces.
The second method of determining the label of an unknown facial expression is by sketching it with FaceShop. As mentioned earlier, in that case the FCP coordinates are determined automatically.

CONCLUSION

At this moment our database is filled with about 100 facial expressions. We focused on emotional facial expressions. During one of the lab works we asked 120 students to design facial expressions using the FaceShop tool corresponding to one of the 300 labels. There are about 300 words with emotional meaning. A problem is that almost every label can be expressed in different ways. Next facial expressions can be mixtures of other facial expressions. Finally facial expressions can vary in intensities. So for every label we have a cluster of facial expressions. In our first prototype we computed the barycentre and used this as a prototype of the facial expression corresponding to the specific label. In the next future we will compute the distribution of points in the cluster and compute the probability of a point belonging to a cluster. And we will add non emotional facial expressions to our database.
From the other side we have many facial expressions, which can be labeled in different ways depending of the context. This is not unusual. In a verbal dictionary one word can have different meanings and typical expressions are presented to show the different meanings.

It proves that the Euclidean distance is a good measure to research the similarities of facial expression from the geometrical viewpoint but not from the semantic viewpoint. Many researchers studied the dimension of the space of all facial expression [M. Pantic, L.J.M. Rothkrantz 2000, C.Pelachaud et al. 1998, M. Turk, A. Pentland 1991].
To conclude , we designed and implemented a first prototype of a facial expression dictionary. The prototype is available via our website. The last two years it is tested by students Technical Informatics during courses and their labwork.

REFERENCES

J.R. Averill, A semantic atlas of emotions concept. JSAS Catalog of selected documents in Psychology, 5, 330, 1975.

D.Efron. Gesture, Race and Culture. The Hague, Mouton & Company, 1972.

P.Ekman and W.F.Friesen. Unmasking the Face. Englewood Cliffs, New Jersey, USA, Prentice-Hall, 1975.

I.Essa, A.Pent.land, Coding, analysis interpretation, recognition of facial expressions, IEEE Trans. Pattern Analysis and machine Intelligence, vol. 19, no 7, pp. 757-763, July 1997.

H.Kobayashi and F.Hara. Recognition of Mixed Facial Expressions by Neural Network. IEEE International workshop on Robot and Human Communication, 381-386, 1972.

M.Pantic, L.J.M. Rothkrantz, Expert system for automatic analysis of facial expressions, Image and Vision Computing 18 (2000) 881-905.

M.Pantic, L.J.M. Rothkrantz, Automatic Analysis of facial expressions: the state of the art IEEE Transaction on pattern analysis and Machine Intelligence, vol 22, no 12, December 2000.

M.Pantic, L.J.M. Rothkrantz, Towards an affect-sensitive Multimodal Human-Computer Interaction, Proceedings of the IEEE, vol 91, no 9, September 2003.

C.Pelachaud, N.I.Badler, and M.Steedman. Generating facial expressions for speech, Cognitive Science, vol 20, no.1:1-46, 1998.

M. Turk, A. Pentland, Eigenfaces for recognition, J,Cognitive Neuroscience, vol. 3, no. 1, pp. 77-86, 1991.

A.Wojdel and L.J.M.Rothkrantz, A Text Based Talking Face. Proc. of the third Int. Workshop on Text, Speech and Dialogue, Brno, Czech Republic, 2000.

INCREMENTAL RULES FOR GROWING PLANTS

Andrew Davison
Department of Computer Engineering
Prince of Songkla University
Hat Yai, Songkhla 90112, Thailand
E-mail: dandrew@ratree.psu.ac.th

KEYWORDS

3D worlds, rules-based programming, animation tools, 3D authoring tools, toolboxes for moving graphics.

ABSTRACT

L-systems are widely used for plant modeling and simulation, with remarkable results. However, we argue that the mathematical formalism underpinning L-systems encourages inefficient rendering of plants which grow and change over time. We propose new types of rules which emphasize the incremental nature of change in a plant's elements, and highlight an element's relationships with other components (e.g. a plant limb has a parent, children, and occurs at a certain level in the overall structure). We have implemented a Java 3D prototype using this approach, and compare it with code using a standard L-system.

INTRODUCTION

Lindenmayer systems (L-systems), consisting of rewrite rules, have been widely used for plant modeling and simulation, due to the direct mapping between the string expansions of the rule system and a visual representation of a plant (Prusinkiewicz and Lindenmayer, 1990; Prusinkiewicz and Hanan 1990; Prusinkiewicz et al. 1993; Prusinkiewicz et al. 1999). For example, the following bracketed L-system contains a single start string 'F', and the rewrite rule:

$$F \rightarrow F\ [-F]\ F\ [+F]\ F$$

The visual characterization is obtained by thinking of each 'F' symbol as a *limb* (or *module*) of the plant. The bracket notation can be viewed as a branch-creation operator; the '-' as a rotation to the right for the branch, '+' a left rotation.

Since each limb is an 'F' symbol, rewriting can continue, creating longer strings, and more complex plant-like shapes.

Each rewrite causes all the limbs (modules) of the current tree to be replaced in a parallel derivation step, reflecting the simultaneous passage of time in all parts of the tree. Time is viewed as discrete, represented by the sequence of derivation steps.

GROWTH AND L-SYSTEMS

Our application domain is a networked 3D virtual world, which changes over time – day turns to night, and plants/trees grow. The scenery is not photo-realistic; more emphasis is placed on the fast rendering of a large number of relatively simple shapes representing different kinds of trees. At certain times, these trees need to grow extremely rapidly.

The application is implemented in Java 3D, a high-level 3D graphics library for Java (see http://java.sun.com/products/java-media/3D/). In the first version of the application, a L-system generated strings which were passed to a rendering component to be turned into trees made from groups of coloured cylinders and other objects. Growth was represented by having the L-system pass the current strings to the renderer at the end of each derivation step. The renderer would dispose of the old 3D trees, generate new trees using the strings, and add them to the scene.

This approach proved to be very slow, and quickly ran out of memory when more than about ten medium-size trees were placed in the scene. The slowness was partly caused by Java 3D's slow run-time removal and addition of scene objects, and the subsequent garbage collection of the discarded trees. Also, as the L-system strings became larger, the renderer required increasingly large amounts of memory to recursively parse the strings and create the 3D shapes.

Part of the problem is due to Java 3D, but we also identified four problems with the L-system: two fundamental ones present in the rule formalism, and two minor ones that are language-related. We discuss the two serious issues first, then the lesser two.

Rewriting = Replacement

Each rewrite of a L-system string creates a more complex tree, but it is hard to see how the new tree has 'grown' out of the simpler one. For instance, what part of the current tree is new wood, which is old wood that has grown a little?

A L-system represents growth as a new structure completely replacing the old one. That is unimportant when the structure is a mathematical abstraction, but has serious consequences when implementing a growth algorithm. The

natural approach, and the most disastrous from an efficiency point of view, is to discard the current structure at the start of a rewrite and generate a new one matching the new string expansion. There is no simple alternative to this since the L-system does not distinguish between old elements (either changed or unchanged) in the structure and the new parts.

No Tree Relationships

Another drawback of the L-system notation is its lack of tree nomenclature. For example, it is not possible to talk about the parent of a node, its children, or its level in the tree. To be fair, some of these capabilities can be programmed by using parameterized L-system rules. However, we believe that a production system for plant modeling should contain intrinsic ways of talking about the branching structure that it represents.

Locating Limb (Module) State

Limb state includes information such as the present length of a limb, its current colour, and its age. Parameterized L-systems handle state by adding additional parameters to the rules, which tends to lead to large rules. Arguably, this solution places the data in the wrong place: state details for each limb should be located *inside* the particular limb rather than in the rules which are applied to all the limbs. This is really an argument for an object-based view of limbs, rather than a procedural one centered around the rules. Benefits include the ability to hide state, improved modularization, and cleaner abstractions.

Rule Reuse

Large groups of L-systems rules often contain very similar rules for the different plant elements. For example, most types of limbs will grow for a period of time (represented by a recursive, parameterized rule), followed by the appearance of child limbs as branches sprout (this is often called a decomposition rule).

Once an object view of limbs is utilized, it follows that plant node types should be represented by classes with their own data, methods, and rule behavior. Commonalties between the classes, whether in their state or rules, can be dealt with by subclassing.

INCREMENTAL RULES

Our principal change to the L-system formalism is to utilize rules which specify rewrites as *incremental changes* to existing limbs, such as a gradual increase in length or a deepening colour. Child branches can be spawned, but are defined in terms of how they are added to their parent limb. This requires the introduction of a tree notation so that the parent-child relationships can be stated.

We implemented our ideas in Java, so gaining the advantages of OOP. In our prototype system, a limb is represented by a `TreeLimb` class, which has over 20 public methods, roughly classified into five groups:

- scaling of the cylinder's radius or length;
- colour adjustment;
- parent and children methods;
- leaves-related;
- various others (e.g. accessing the limb's current age).

The system is activated every 100ms (the time interval between rewrites), and applies its rules to all the `TreeLimb` objects, affecting a parallel rewrite analogous to the L-system model. The difference lies in the incremental nature of the rules.

The rules have an if-then form, where the action is only carried out if the conditions evaluate true for the current limb.

The rest of this section contains descriptions of the simple 'length' and 'thickness' rules, and the slightly more complex 'child limbs spawning' rule.

The 'length' rule incrementally increases the length of a limb up to a maximum of about 1 unit:

```
if ((limb.getLength() < 1.0f) &&
     !limb.hasLeaves())
  limb.scaleLength(1.1f);
```

`limb` is the current `TreeLimb` object under consideration. The `hasLeaves()` part of the condition stops branches from growing any longer once they have leaves.

The 'thickness' rule mandates how a limb's thickness should change:

```
if ((limb.getRadius() <=
    (-0.05f*limb.getLevel()+0.25f))
                  && !limb.hasLeaves())
  limb.scaleRadius(1.05f);
```

The equation `-0.05*limb.getLevel()+0.25` relates the maximum radius to the limb's level. For example, a limb growing directly out of the ground (level == 1) can have a larger maximum radius than a branch higher up the tree. This means that branches will get less thick the higher up the tree they appear, as in nature.

The 'child limbs spawning' rule creates at most two child limbs:

```
if ((limb.getAge() == 5) &&
    (treeLimbs.size() <= 256) &&
    !limb.hasLeaves() &&
    (limb.getLevel() < 10)) {
  if (Math.random() < 0.85)
    makeChild(randomRange(10,30), limb);
  if (Math.random() < 0.85)
    makeChild(randomRange(-30,-10), limb);
}
```

The four conditions only permit child limbs to appear if the parent is at least 5 time intervals old, the total number of

limbs in the scene is less or equal to 256, the parent has no leaves, and the branch isn't too far up the tree.

`Math.random()` is employed to make it less certain that two children will be spawned. `randomRange()` returns a random number (in this case, an angle) in the specified range.

`makeChild()`'s definition:

```
private void makeChild(double angle,
                       TreeLimb par)
{ TransformGroup startLimbTG =
              par.getEndLimbTG();
  int axis = (Math.random() < 0.5) ?
               Z_AXIS : X_AXIS;
  TreeLimb child = new TreeLimb(axis,
            angle, 0.05f, 0.5f,
            startLimbTG, par);
  treeLimbs.add(child);
   // add new limb to tree limbs list
}
```

The first line gets the 'end point' of the parent limb, which becomes the place where the child is connected. `Math.random()` is used to randomize the child's orientation axis.

Figure 1 shows a sequence of screen shots of the application. Five trees grow from saplings, young green shoots turn brown, leaves sprout, all taking place over a period of a few seconds.

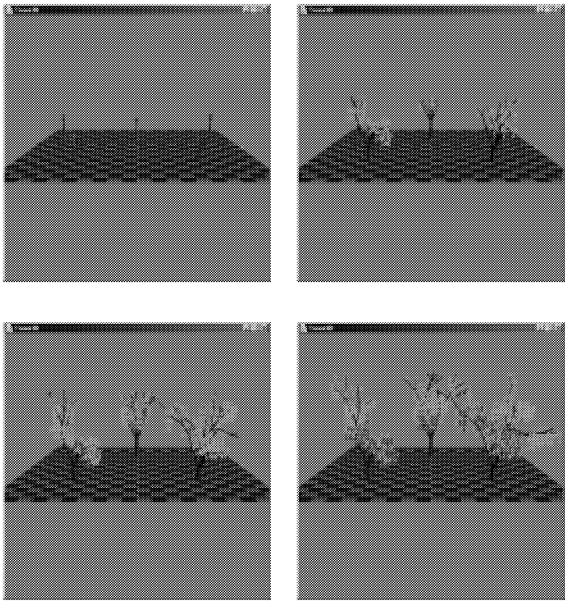


Figure 1: Growing Trees

DISCUSSION

The application carries out very little garbage collection, in contrast to the original L-system-based code, because the trees are not being repeatedly regenerated. In fact, no tree limbs are discarded at all.

The application's main control structure is a time-triggered loop through all the limb objects, applying the rules to each one. The original code uses recursion to generate all the limbs in each tree from scratch in every derivation step. It is hardly surprising that the new application has a much faster rendering time (sometimes twice as fast).

Over 4000 limbs can be created before the application needs additional heap space, compared to tens of limbs in the original code. This is due to the reduced garbage collection needs and the use of looping rather than recursion in the rendering.

Time is represented discretely, and growth is defined in incremental steps rather than as differential equations (e.g., as in dL-systems). The primary reasons for this choice was to make rule definition simpler: most users find the specification of differentials rather difficult. This approach also avoids the need for a run-time solver for the equations.

Java offers the advantages of OOP, and each limb is represented by its own object. The principal limb class is `TreeLimb`, which can be subclassed easily.

A drawback of our application is the complexity of the rules which refer to graphical elements of the 3D models. For example, `makeChild()` utilizes a Java 3D `TransformGroup` node to connect a child to its parent. This suggests the need for a higher-level notation which hides connection details.

The application shown in Figure 1, together with a more detailed explanation of its implementation can be found at `http://fivedots.coe.psu.ac.th/~ad/jg/ch178`.

REFERENCES

Prusinkiewicz, P. and Hanan, J. 1990. "Visualization of Botanical Structures and Processes using Parametric L-Systems", In *Scientific Visualization and Graphics Simulation*, D. Thalmann (ed.), pp.183-201, John Wiley & Sons.

Prusinkiewicz, P., Hanan, J., and Mech, R. 1999. "An L-System Plant Modeling Language", In *Proc. of the Int. Workshop AGTIVE'99*, M. Nagl, A. Schuerr and M. Muench (eds), Kerkrade, The Netherlands, September, LNCS 1779, Springer, pp.395-410.

Prusinkiewicz, P., Hammel, M.S., and Mjolsness, E. 1993. "Animation of Plant Development", *Computer Graphics*, Vol. 27, No. 3, pp. 351-360.

Prusinkiewicz, P. and Lindenmayer, A. 1990. *The Algorithmic Beauty of Plants*, Springer-Verlag, NY.

MEDICAL APPLICATIONS

USABILITY OF THERAPIST'S USER INTERFACE IN VIRTUAL REALITY EXPOSURE THERAPY FOR FEAR OF FLYING

Lucy T. Gunawan &
Charles van der Mast
Delft University of Technology
Mekelweg 4 2628 CD Delft
the Netherlands
Email:
c.a.p.g.vandermast@ewi.tudelft.nl

Mark A. Neerincx
Delft University of
Technology, Mekelweg 4,
2628 CD, Delft, & TNO
Human Factors, Soesterberg,
the Netherlands
Email: neerincx@tm.tno.nl

Paul Emmelkamp &
Merel Krijn
University of Amsterdam,
Roetersstraat 15,
1018 WB Amsterdam
the Netherlands
Email: p.m.g.emmelkamp@uva.nl

KEYWORDS
Virtual reality, human-computer interaction, usability, phobia treatment.

ABSTRACT

Delft University of Technology has implemented an experimental virtual reality system for treating fear of flying. The patient sits in an airplane chair equipped with vibration devices. By wearing a head-mounted device, the patient is immersed in the virtual world (patient's user interface), sitting in the computer-generated cabin of a virtual airplane and experiencing various aspects of flying. The therapist takes control of what will be experienced by the patient, by using the therapist's user interface. After being used in a large scale comparitive study, the usability of the system emerged as an important consideration. Therefore, the system's usability was our main concern in this study. After analyzing the current user interface and its deficiencies, some improvements were carried out. These improvements include a new design of therapist's user interface, including database support and reporting tools, and adding some important features in the patients user interface: lightning simulation, time of the day background, cabin density control, flying destinations, attendant announcements, airplane rolling,. The improvements were evaluated with a usability analysis. Interviews and laboratory experiments were done to measure the usability. And as we expected, the results gave positive indicators leading to improvements of the system usability – as seen by the therapist.

INTRODUCTION

Recent research (e.g. Emmelkamp et al. 2002) has proved that virtual reality technology (VR) can be implemented in clinical therapy.
One of the virtual reality projects at Delft University of Technology is virtual reality exposure therapy (VRET) for treating fear of flying (Schuemie 2003). This project is a collaboration between two disciplines, Psychology, brought in by University of Amsterdam and VALK foundation in Leiden, and Human-Computer Interaction (HCI), brought in by Delft University of Technology. Delft University of Technology is in charge in the technical part for this project, developing the VRET system. The system has two main user interfaces, one for the therapist to control the session and one for the patient to experience flying. This experimental system was deployed successfully during experimental therapy sessions in comparative and other studies for acrophobia, claustrophobia and agoraphobia (Schuemie 2003).
Our goal is not only to implement this system but also to support clinical therapy professionally. Therefore, the intention of this research is to improve the usability of the system, both for the therapist's and the patient's user interface. Schuemie (2003) worked mainly on the usability of the patient's interface by parameters of presence and locomotion. In this paper, we only discuss the therapist's user interface improvements. The complete reference to patient's user interface is described by Schuemie (2003).
Our first step was to collect feedback by the few therapists who used the system during the comparative studies on treatment in vivo and in VR. By using the information gathered, we developed some improvements. The therapists participated in every step of system development.

THE SYSTEM

The patient sits in a real airplane seat equipped with a bass loudspeaker in the seat to simulate vibration in the airplane. By wearing the head-mounted device (HMD), the patient is immersed in a virtual world, enters the computer-generated cabin of the virtual airplane and will experience various aspects of flying controlled by the therapist who is nearby in the same room. The patient is exposed to flying situations such as: sitting in a standing still airplane, taxiing on the runaway, taking off, flying in good weather, flying in bad weather and landing. During therapy, therapist can see and hear patient's experience during the virtual flight. The therapist works most of the time using the therapist's computer where he/she can control the VR world, monitor the patient regularly and check the level of fear experienced by the patient, see Appendix A.
The first version of this system was built in 1999 by Schuemie, see (Schuemie & Van der Mast 2001). In this paper are described the details of hardware and software specifications.

REQUIREMENTS

From the interviews with the therapists was found that they need to have more control over the aircraft and what is happening during the flight. In the old therapist user

interface (see Appendix B) an overview of all the functionality is shown. Input was given via keyboard, mouse and joystick. The new requirments should be important to be able to give a more flexible treatment reacting to individual characteristics of the patients. The therapsts asked e.g. for more destinations (in the old user interface only Milan), flying during day or night, more bad weather control, control of flap wings and rolling.

METHOD

The adaptation of the usability evaluation model proposed by (Scriven 1967) quoted by (Rosson & Carroll 2002), is used as our research methodology (Figure 1). The model distinguishes between formative and summative evaluations. The goals of formative evaluation are to identify the design aspects that can be improved, to set priorities, and to provide guidance in how to make changes to a design. This evaluation is conducted during the design and development process. The summative evaluation goals are to measure quality; to evaluate a design result whether the system has met its usability objectives and it is conducted at the end of development process. This model can be seen as an iterative process that the current system can be evaluated as if it still is in the design process, although it was finished before.

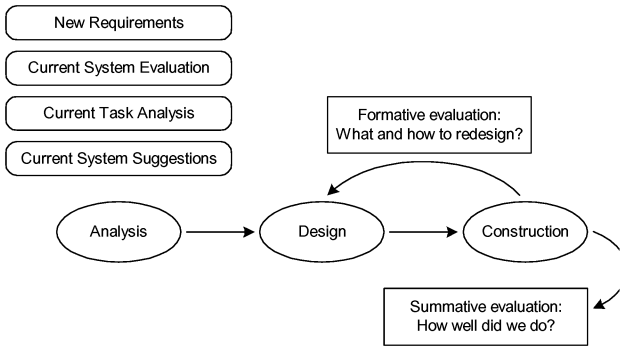


Figure 1: Research methodology

The analysis process had some inputs such as task analysis of the current system, new requirements gathered by interviewing the therapists, evaluation of the current system and some suggestions from the research partners at University of Amsterdam. In the design phase, we designed two kinds of user interface (UI), the virtual world for the patient and the user interface control for the therapist. Our main concern in this study was to improve the usability of the therapist's user interface, but sometimes we have to enhance the patient's UI in order to extend the therapist's UI. New features were added, improvements were carried out, and a new therapist user interface was implemented. The evaluation of the therapist's UI is measured in terms of usability, i.e.effectivity, efficiency and satisfaction (Rosson and Carroll 2002). We used a mediated evaluation which is a mix between analytical and empirical method. The analytic evaluation is done early and during the design process. The result of this analysis is used to motivate and develop materials for empirical evaluations. Heuristic evaluation as an inspection method was done and also the ten general heuristic guidelines by Nielsen (1993) were taken into account. The empirical method is done by a user evaluation experiment. Because we have only a very limited number of professional therapists for treating fear of flying we added other persons to do the same treatment/job. From Neerincx at al. (2001) we know that this may deliver valid results. The therapists tested the system directly with a user/patient in the virtual airplane. Some specific tasks were given to the therapist to complete. Information was gathered, such as the observation protocol, performance time, errors and subjective evaluation. The subjective evaluation was acquainted by using usability questionnaires and interviews.

DESIGN AND DEVELOPMENT OF THE NEW USER INTERFACE

Based on the task analysis by Schuemie (2003), the task model of current in vivo therapy for phobia treatment was formed. The main goal of each therapy is to cure the patient. During the exposure, the therapist determines patient's fear by exposing manipulating stimuli to patients, and changing it when needed to adjust patient's fear. The therapist responds to each question the patients might have. This solves any ambiguity patients might have. Responding or answering the patient's question might not have contribution to cure the patient directly, but at least it will facilitate patients in performing their tasks. Patients believe that by following the therapist's instruction, they can get rid of their fear. People with phobias have strong tendency to avoid fearful situations. This conflicts with therapist's instruction. To resolve ambiguity in therapy, the patient sometimes need to inquire about certain matters.

As the other input to the analysis, there were also some suggestions in the several areas where usability can be improved such as providing the therapist with cognitive artifacts representing the historical patient's score over the therapy and increase the learnability and memorability of using the system.

Though all new requirements urged to be added, some consideration was taken, and we could not implement them all. A new UI for the therapist was designed, some features to the world were added such as: lightning, possibility of flying during different time of the day (morning, day, afternoon, and night), possibility to change the cabin's passenger density, possibility to fly to another destination, possibility to choose the voice of pilot and purser, possibility to roll the airplane during the flight, possibility to dim the cabin's light and the most important one is the feature of database, the possibility to save and print historical data of the patient with its Subjective Unit Discomfort (SUD)s artifacts.

The overview of the old therapist's user interface and the new one can be seen in appendix B and C.

RESULTS

The evaluation phase took place in the end of system development. The evaluation goal was to evaluate the usability of the therapist UI whether the new features added and changed showed significant improvement. Thus we formulated our hypothesis as follow: *The "improvements" in Therapist UI are increasing the usability of the system.*

We did one experiment with sixteen pairs of patients and therapists (32 participants), five of them were real therapists, and the rest were students. There were two therapy sessions for each therapist and patient pair, one using the old system (System A) and one using the new improved system (System B). The order was at random. Eleven students were trained for this experiment as therapists (most of them never used our system before) and five real therapists were asked to do therapy sessions. To give more objective judgment of the two systems, none of them was informed which the old system was and which the new improved system was. Each session took about twenty minutes and there were a small break between the sessions. Detailed therapy session tasks was given to the therapist. It included instruction to load the correspondence world, fill patient and session information, gradually expose the patient to the flying sequences in virtual world and end the simulation. Time elapsed was recorded during each task and what the therapist done was monitored and noted such as mistakes done by therapist, questions, and assistance needed. Each therapist had to fill in the usability questionnaire after each session. After finishing the therapy sessions, the therapist were asked about general remarks, comments, suggestion and general comparison about two systems. Another extended subjective evaluation with real therapist also was done for gathering information that is more authentic.

Usability Questionnaire

The reliability analysis for the usability questionnaire was performed. Cronbach alpha was 0.9254 (N of cases=32, N of items=27), showed that responses have a really good internal consistency.

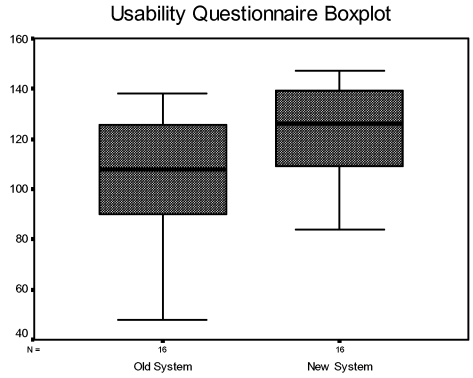


Figure 2: The box plot of the usability questionnaire for the old and the new system

We analyzed our data using ANOVA for repeated measures, because in our experiment the same patient took part in two sessions (the old and new system). The ANOVA for repeated measures show a significant difference in the total score between two systems: F=12.321, p=0.004. This significant difference should be verified; which one had the higher or lower usability? By calculating the means of questionnaire for old and new system, we drew a box plot as seen in *Figure 2*. This figure shows us that the means of the new system (122.75) is higher than those of the old system (105.44). Thus, by this result we accepted our hypothesis that the improvements made in therapist UI increase the usability of the system. There were no significant differences between the groups who tried the systems in a different order: the old system for this first session followed by the new system or the new system for this first session followed by the old system. (F=1.310, p=0.275). There were no significant different results between the real therapist and student as therapist when they fill in the questionnaire. (F=0.207, p=0.658).

Table 1: Average scores (and standard deviation) of the additional usability questions (n=16) regarding new features of the therapist UI, scale from 1 to 7.

Element	***Usefulness***	***Ease of use***
Flight Plan Control		4.9375(1.3401)
Cabin Control		4.8750(1.0247)
Roll Control		4.8125(1.1087)
Flight View		5.3125(1.0145)
Print Function	5.1250(1.2583)	4.8750(1.5000)
Timer Feature	4.5000(1.3166)	
Simulation Control		4.9375(1.5262)

Additional usability questions filled only for the new system to evaluate the new features had a reliability alpha of 0.8216 (N of cases=16, N of items=8). It means that these eight additional questions had a good internal consistency. The results of the additional questionnaire are displayed in *Table 1*, which shows that all the new features were evaluated positively. They were very useful and/or easy to use. Thus, by these results we add our hypothesis to include proof that the new added features are useful and easy to use.
We found also a significant correlations between usability of the old and the new system (Pearson Correlations=0.727, p=0.01). A higher score in the usability questionnaire of the old system tend to paired with ahigher score in the usability questionnaire in the new system.

Performance Time and Error

Performance time and error were measured during the experiment, in every task given. The first task (task 1 in Figures 3) was to load the virtual environment. The second task (task 2 in Figures 3) was to fill the session information, such as: patient, therapist and session number. The third task (task 3 in Figures 3) was to gradually expose the patient in the virtual world.

Table 2: Means of tasks completion time (in seconds)

Participant	***Old System***	***New System***
Real Therapist	1090.64	1344.50
Student as Therapist	969.06	1209.86

We could not really compare the time completion between the old and new system, because they are different worlds with different added features. The new system always takes longer time to complete everything, because of new added features. What we can see here is the comparison of time completion between the real therapists and students as therapist. The student as therapist tends to complete the task

faster than the real therapist, as you can see in *Table 2*. This is maybe because students don't really know the pace of therapy session.

Errors in our system were defined as errors made by the therapist during the therapy sessions, and when assistance was needed. The comparison of error rate for both systems can be seen in *Figure 3a, 3b, 3c*. From the graphics we can see that the error rate for the new system is better than that of the old system. We can also compare the means of error rate between real therapist and student as therapist. We can see here the difference, that students make more errors than the therapists.

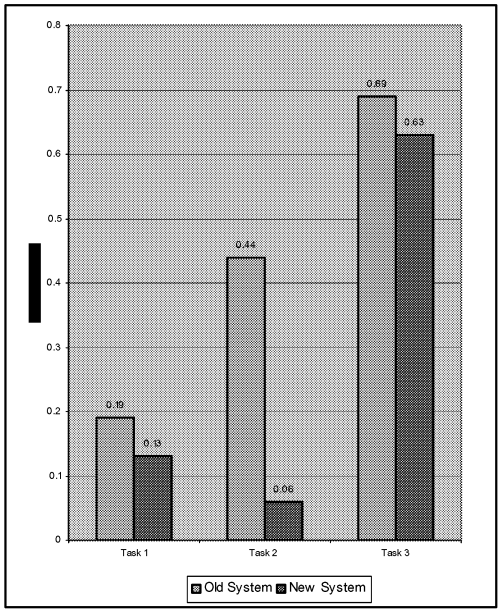


Figure 3a: Bar chart comparison of average error rate for old and new system

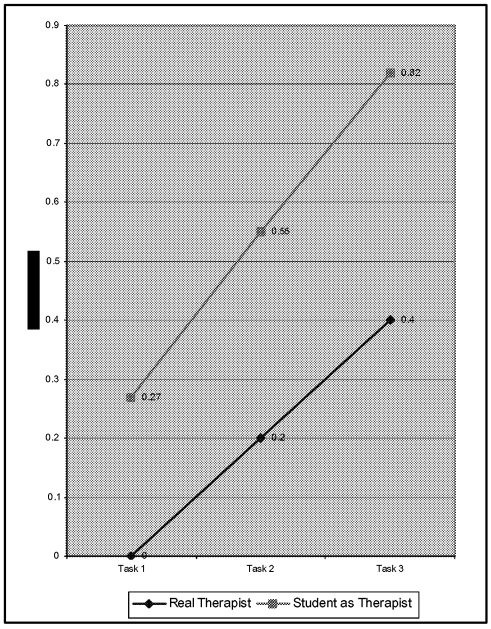


Figure 3b: Old system comparison of average error by real therapist and students as therapist

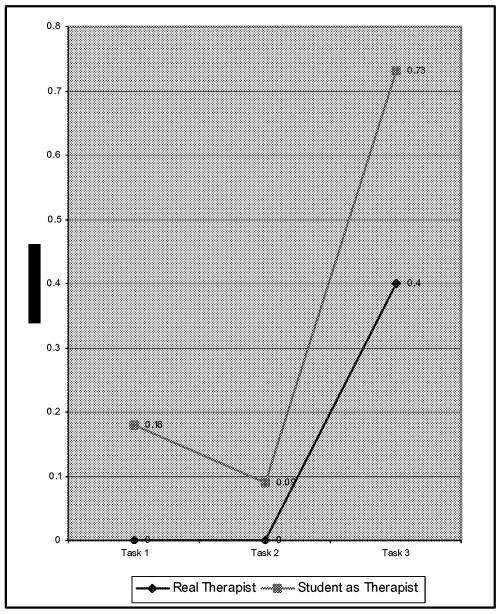


Figure 3c: New system comparison of average error by real therapist and students as therapist

Therapist Subjective Evaluation

Five therapists were given more questionnaires after conducting two sessions. The results of the first five questions are summarized as in *Table 3*. The roll control was not used often, and it was not too easy to use either. The roll control is used during flying stage. It rolls the airplane so the horizon will slightly leaning in the patient's view. It was unclear in the therapist's UI when the roll control can be used. The roll control should represent a continuous control process rather with a discrete button.

Table 3: Average scores (Standard Deviation) of the therapist subjective evaluation, scale from 1 to 5.

Element	Frequency of use	Ease of use	Usefulness
Roll Control	2.20(1.3038)	3.60(1.9494)	-
Bad Weather Control	3.60(1.6733)	4.80(0.4472)	-
Flight View	4.00(1.7321)	-	4.80(0.4472)
Timer feature	-	-	4.80(0.4472)
Print Function	-	-	4.40(0.8944)

The therapists did not too often use the bad weather control but the easiness of this control was evaluated very positively. The new feature of flight view was evaluated positively both for the frequent use and usefulness, this allowed the therapist to see an overview of the sessions. The timer feature was also found to be very useful. It gave information when one stage was about to finish so the

therapist can plan the next action to be carried out in the therapy session. The print function was also discovered to be very useful. Most therapists supported their answer by stating that the report will be used in the future, to know the overview what the patient did during the sessions and also to give feedback to the client.
Four therapists stated their preferences to fill in the patient and session information in the same user interface with the world control. One therapist stated that it actually does not matter, as long as this feature exists. Five therapists agreed that the same form for patient and session information entry was easier to use than the separated ones. Therapists also liked the flight control subjectively. It helped the therapist in planning a session.
Three therapists liked the idea of given restrictions in controlling the VE, but two of them stated these limitations were very annoying and did not give them enough freedom, especially in controlling the voice announcements.
All real therapists agreed the new improved system was easier to use than the old system. One therapist said that the new system was more difficult to learn due to the more complicated features, but it has a structured user interface that makes it convenient to use. Three therapists stated the new system is easier to learn and one therapist said it did not a matter. All therapists also agreed that they liked all the added features. Most of them like the lightning and thunder, because they gave a surprise feedback. The most useful feature of all was the flight view that combined options from flight control and voice control. The therapists also liked real voice announcements from pilot and purser, and the sound of flap wings and landing gear. One therapist stated that the report feature would be very useful. All therapists preferred to use the new system to treat a patient who has fear of flying. They said that the new system was more organized than the old system. One therapist who gave initial requirements states that we had almost everything fulfilled, except for the amount of the avatars and the unreal look of the clouds. One therapist suggested that we should have separated approaching, touch down and taxiing stages during aircraft landing.

DISCUSSION

Overall, all subjects gave positive feedbacks to the improvement of the therapist's UI. From sixteen therapists, ten of them state their preferences in using the new system instead of the old system. One therapist preferred the old system to the new one, and five therapists did not given their preferences. Most of them preferred the new system to the old system because of the language used, more controllable features, its ease of use, easily learnable, and it provides clearer instructions. One therapist preferred the old system because it was less complicated due to less number of buttons that needed to be pressed to operate the system.
The possibility to compose scenarios and to simply run them afterwards was coined by one of the therapists as his suggestion. We referred to this function as autopilot. We thought about this function in the beginning [of what???], but from initial interviews, the therapist wanted to have complete control during therapy session. Thus, this feature was not implemented. Other useful suggestions were the introduction of cabin sound (people talking, baby crying, etc.) and alert sound for alarm.
Some feedbacks were also gathered and therapists were asked to list three things they liked most and least in using the system. Lightning and thunder became favorite features in the new system, followed by feature of information overview during therapy sessions with linked option and limitation. The possibility to see what the patient's sees in VE by the therapist also evaluated very well. The system gave therapists a feeling of full control during the therapy session. The overall sound effects in the old system were louder than the new system. It can clearly be heard during landing stage. We think it would be nicer if the new system could use the same quality of sound as in the old system. Most of the therapist did not like the alarm reminder that was not functioning very well in the new system. It did not produce a reminder alert. The note feature was also not too useful either.

CONCLUSION

We can conclude that our formulated hypothesis for usability is accepted. It was significantly better than the old system, i.e. it increases the usability for the therapist.
VRET is slowly becoming the daily practice of therapists. During this transition, the usability issues play an important role in the acceptance of such an advanced technology. In the coming years, more therapists will work with this VRET "fear of flying" systems. It is hoped that the usability improvement of this and other systems could make their work much easier and could possibly increase the effectiveness of their therapy. However, to convince therapists to start using VRET systems an important issue is to show the profits on the level of the general health system and its costs. It is already shown that VRET for acrophobia works (Emmelkamp et al. 2002) and achieves the same results as treatment in vivo. This means that all "other" profits such as flexibility, better control of the conditions, increase the scale of use, etc. are to be exploited. But the health system including insurance companies must get interested and involved.
The new system including the new therapist's user interface as evaluated here offers a considerable better function for the therapist to treat fear of flying using VRET. For the usability engineering method, it is interesting to note that the usability profits of the new system show the same pattern for students as for real therapists. It is often hard to get a large number of specialists involved, and adding non-specialists can help to collect sufficient data on user behaviour (cf. user interfaces for astronauts in the space domain; Neerincx et al, 2001). By a better user interface for the therapist we think that the therapy for this kind of phobia can be done more efficient, more effective and with more satisfaction. The interface could be further improved by offering some support to the therapist in the form of a built-in agent advising the therapist on the next steps to take.

ACKNOWLEDGEMENTS

The authors like to thank Lucas van Gerwen from Stichting VALK in Leiden for his comments and for his support as director of a team of fear of flying therapists.

REFERENCES

Emmelkamp, P.M.G.; M. Krijn; A.M. Hulsbosch; S. de Vries; M.J. Schuemie; and C.A.P.G. van der Mast. 2002. *Virtual Reality Treatment versus exposure in vivo: A Comparative Evaluation in Acrophobia Behaviour Research & Therapy*, Vol 40(5), pp.509-516.

Neerincx, M.A.; M. Ruijsendaal; and M. Wolff. 2001. Usability Engineering Guide for Integrated Operation Support in Space Station Payloads, *International Journal of Cognitive Ergonomics,* 5(3), 187-198.

Nielsen, Jacob. 1993. *Usability Engineering*. Boston: AP Professional.

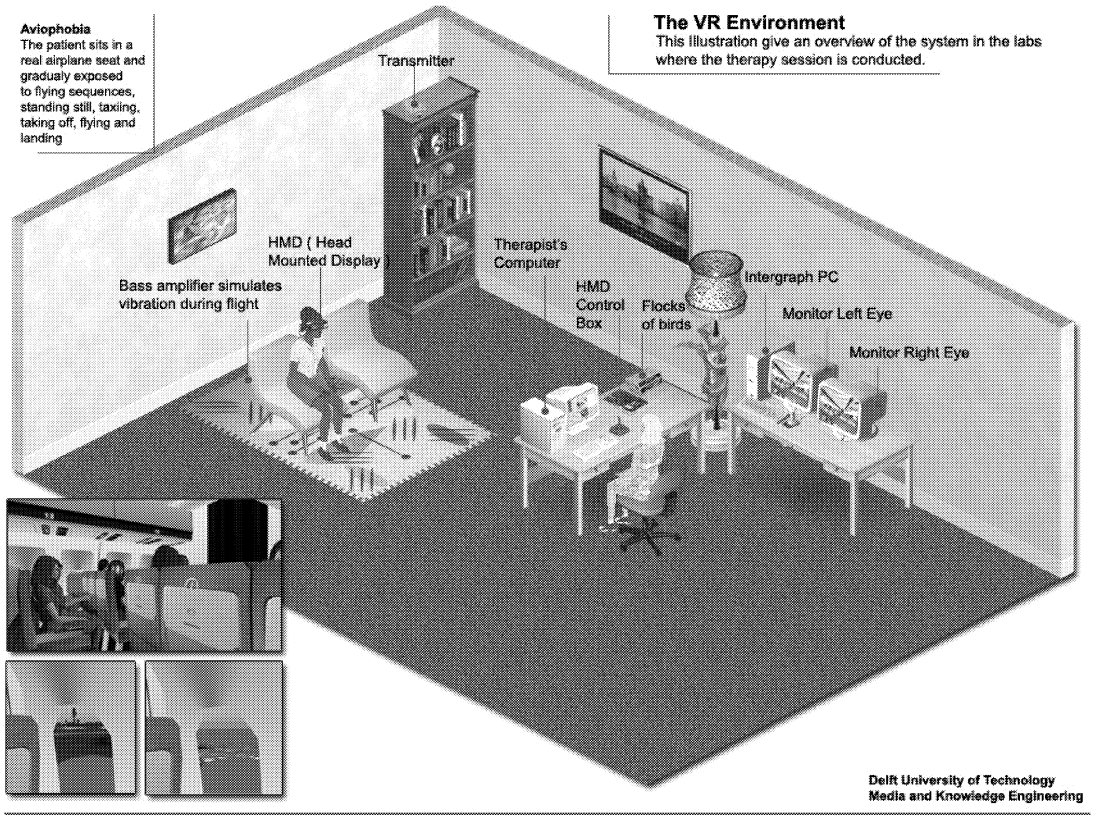
Rosson, Mary Beth; and John M Carroll. 2002. *Usability engineering scenario-based development of human-computer interaction.* 1st ed. Morgan Kaufmann Series in Interactive Technologies. San Fransisco: Academic Press.

Schuemie, Martijn; and Charles van der Mast. 2001. *VR Testbed Configuration for Phobia Treatment Research*. Proceedings of the Euromedia 2001 Conference, April 18-20 2001, Valencia, Spain, pp.200-204.

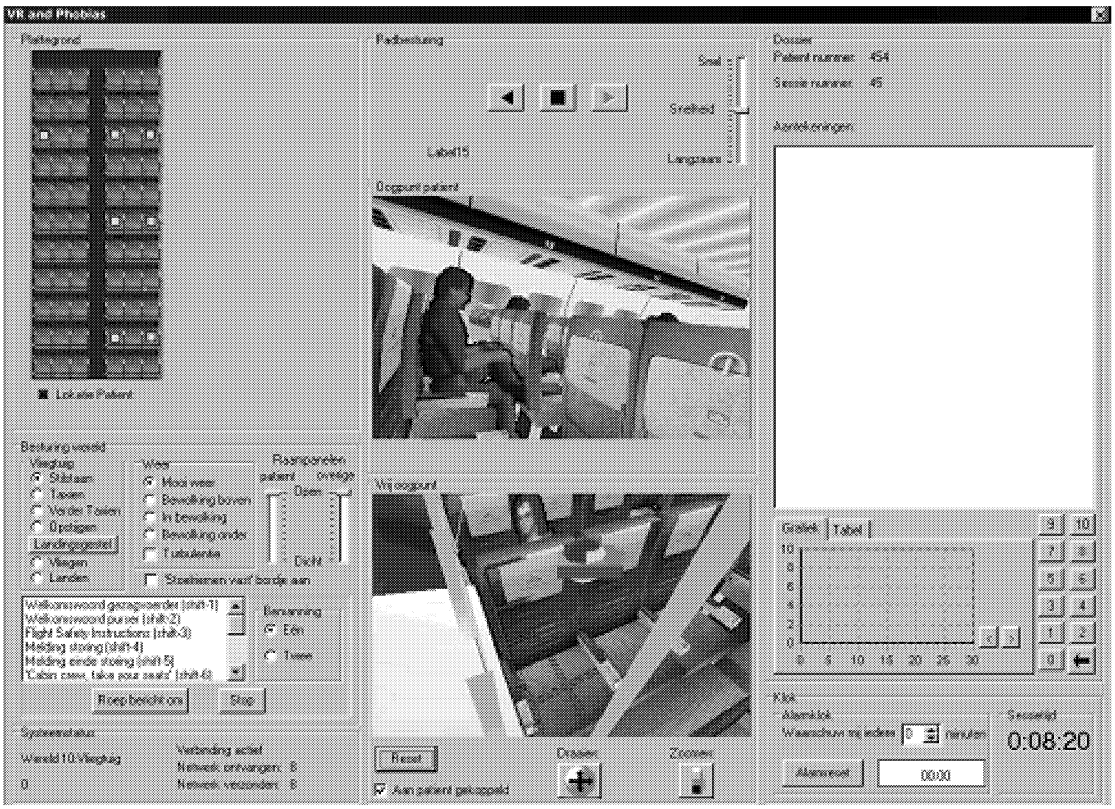
Schuemie, M.J.; C.A.P.G. van der Mast; M. Krijn; and P.M.G. Emmelkamp. 2002. *Exploratory Design and Evaluation of a User Interface for Virtual Reality Exposure Therapy*, in: J.D. Westwood, H.M. Hoffman, R.A. Robb, D. Stredney (Eds.), Medicine Meets Virtual Reality 02/10, IOS Press, pp.468-474

Schuemie, Martijn. 2003. *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*. Ph.D. Thesis, Delft University of Technology, available via http://graphics.tudelft.nl/~vrphobia.

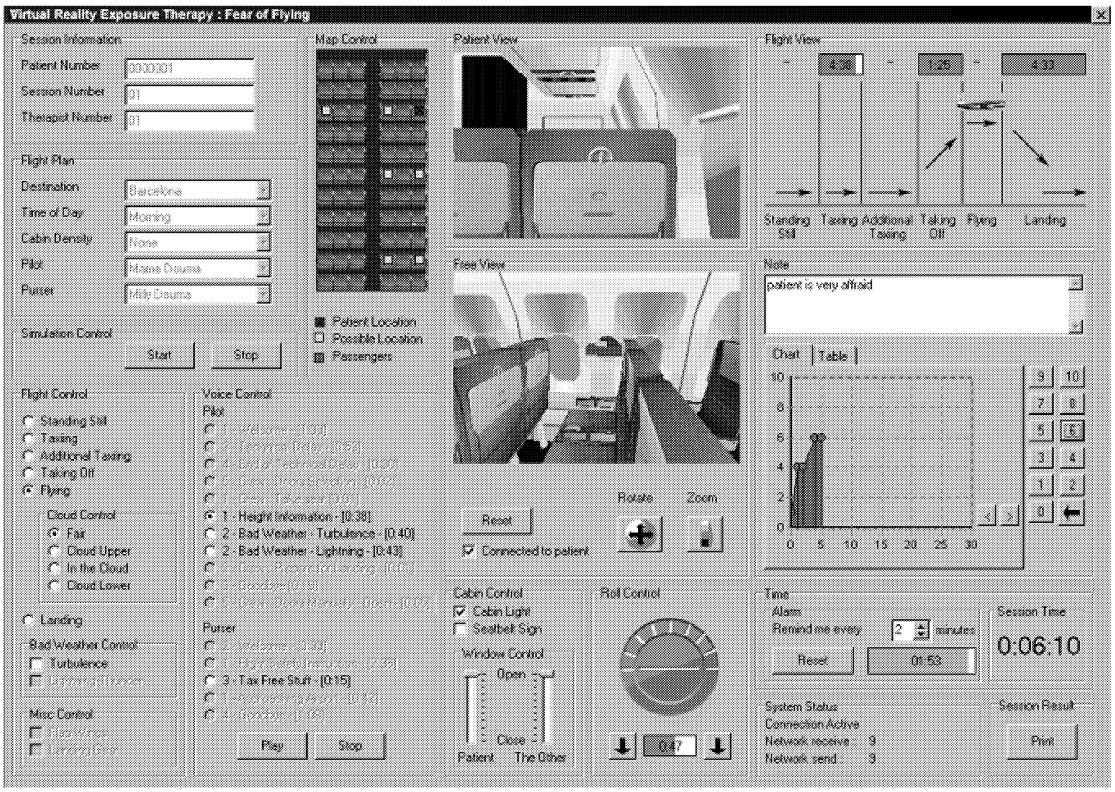
Appendix A. *The VR environment for treating phobias at Delft University of Technology*



Appendix B. The old user interface for the therapist to control the sessions.



Appendix C. The new user interface for the therapist to control the sessions.



PATIENT CONSENT ADVISORY SYSTEM

Titis Karweni and Steven M Furnell
Network Research Group
University of Plymouth
A304 Portland Square
Plymouth UK

Nick Gaunt
Plymouth Hospital NHS Trust
ITTC Building, Derriford
Plymouth UK

KEYWORDS

Electronic Health Record (EHR), Patient consent, Access control

ABSTRACT

To enable personal electronic health records (EHR) to be lawfully shared between healthcare providers, patients require a means to control the dissemination of their data. However, without appropriate advice, patients may withhold consent in an uninformed manner, resulting in the unavailability of their data to certain parties in critical situations. The authors propose an automated advisory tool that assist patients in determining whether to grant or withhold consent to share elements of their record. The proposed model will inform the patients of the consequences of their chosen action based on the patient's circumstances. The patients will be able to use the system independently with only minimal intervention from their primary care doctor. It is considered that a full implementation could be seamlessly integrated into the EHR system as a significant feature of its access control mechanism. This will promote a greater degree of freedom for the patient in controlling the access to their records without putting their, and other's, safety at greater risk or breaking the law.

INTRODUCTION

All health care organizations have patient records. Traditionally these records are in the form of paper folders (Singleton 2002). With the recent movements towards patient centred healthcare services, and the ever wider use of the Internet worldwide, web-based electronic health records are being used in many such organizations. Currently, more than 60 web-based personal health records are available on the Internet (Carr 2003).

The last two decades have also seen the increasing number of laws and regulations being enforced for the protection of privacy. Examples of these are the Data Protection Acts 1998 [DPA 1998] (HMSO[a] 1998) and the Human Rights Acts 1998 (HMSO[b] 1998) in the UK, the EU Directives that applies to all EU member countries (European 1995), and the HIPAA that is applicable in the USA (HSS 2003). Such laws require all organizations to protect personal information of their clients. In the context of healthcare services, organizations should, among other things, obtain patient's fully informed consent before any information can be processed. They should also allow and provide means for patients to easily access their information, as well as to exercise their right to object to its processing, provided that it is justified by national legislation (European 1995).

BACKGROUND AND REQUIREMENTS

The UK National Health Service (NHS) is aiming to develop an Internet accessible personal health record as a response to these requirements. A total of 19 health communities were chosen to be part of the Electronic Records Development and Implementation Programme (ERDIP), with the authors being part of one such project. This program *was established to provide the opportunity for in-service development and demonstration of best practice and progress towards shared Electronic Health Records* (NHSIA[a] 2003). It is envisaged that the EHR will be accessible via the Internet by every citizen in the UK, as well as their clinicians and carers, which often involves multiple parties or organisations, on a need-to-know basis.

EHR security and patient confidentiality

An EHR system potentially contains information that is classified by law as personal sensitive information (e.g. mental or sexual health information), and therefore patient confidentiality must be protected. The EHR system should be highly secured and yet still allow for easy access for their carer. Maintaining patient confidentiality and the security of the records are even more crucial when parts of the records are to be shared with other domains such as the social care team (Gaunt 2003).

Access control and patient safety

Due to the above requirement, the EHR access control mechanism needs to be one that:

- operates strictly on a need-to-know basis;
- follows the dynamic nature of the working patterns of the carers;
- respects patient's wishes for privacy.

On the privacy issue, one could argue that people have different views regarding the privacy of their personal information. For example, many women are likely to be very happy to share information about their pregnancy, but it may not be so for one that happens outside marriage. On the other hand, information about HIV, which is usually considered as highly sensitive and requires high protection, may not be considered so for certain individuals. It is therefore understandable that it would be very difficult, if not impossible, to classify health information such that it would be efficiently suited to every individual. On that ground, it makes sense to let the patients themselves to express their preferences on permitted access to their health information. That is to say that the patients will determine the level of security required for each data element in their records, based upon their familiarity with those people requesting access to their information (Schoenberg and Safran 2000).

In doing this, however, it is clearly important that confidentiality does not impede the provision of prompt and effective patient care (Caldicott 1997). It must be recognised that patients may not always appreciate the implications of denying access, and so in some circumstances, they may wish to withhold critical information in a manner that is potentially unnecessary, unlawful, or that may jeopardise their care. For example, a report from one of the ERDIP projects revealed that healthcare professionals may be at risk if a violent patient were allowed to prevent the release of such information to their current or future carers (NHSIA [b] 2003). To prevent such thing from happening, it is necessary for the care provider to give advice to patient on what information is worth risking disclosure to different parties, such that patients are able to make an educated choice without putting themselves or other people in danger, or breaking the law.

CONSENT ADVISORY SYSTEM

Ensuring that patient will make the right choices in deciding which parts of their record are to be shared or withheld will ideally involve asking their primary care doctor (who is usually in the best position to know a patient's health status) to spend time with them explaining all the issues surrounding the releasing or withholding of patient's health information. One can imagine, however, that this process will take a great deal of the primary care doctor's valuable time (Schoenberg and Safran 2000). For this reason, a patient advisory system is proposed to reduce the burden upon the clinician. It should be noted in advance that the system will not completely replace the role of the clinician. There are two reasons for this: firstly some complicated cases may still best be tackled by a clinician, and secondly it is recognized that some people cannot read well (e.g. due to a non-English speaking background) or have limited ability to understand things. In such cases, a clinician (who may be specifically hired for this purpose) will be needed to ensure that the patient's understanding of the basic issues is met.

The system will need to provide two main functions. The first is to explain to the patient the ways in which their personal information will be used and processed. This is to ensure that the patient has understood what his consent to sharing information entails. A consent recording mechanism can also be provided as a sub-function, which can later be used in litigation cases. The second function is to give advice or guidance to the patient as he is making choices about what information is to be shared or withheld.

The proposed model – a prototype

Figure 1 proposes a simple model of a patient advisory system, the components of which are described in the paragraphs that follow. The patient will mainly use the advisory system independently, although he may be accompanied by an appointed person in special circumstances.

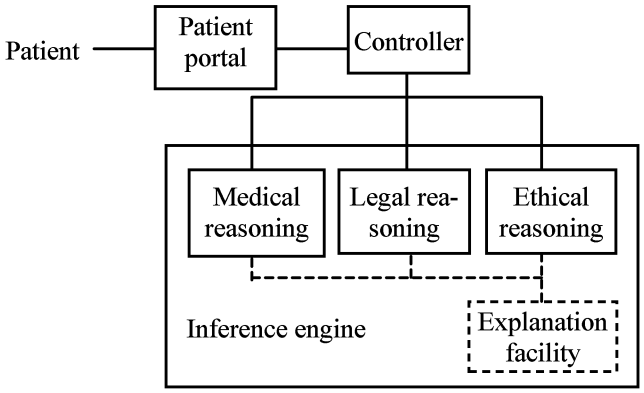


Figure 1: Conceptual Model

The patient portal

The patient portal is the user's interface to the system. The first thing to be established here is to find out the patient's intended action. To do this, it will do the following: firstly the system will give a general explanation regarding consent for information sharing

- what it means to give consent
- data uses of the healthcare establishment (HCE)
- an overview of who can read which part of the record (i.e. the HCE will have a default setting for access control, which can be changed by the patient)
- the patient's rights

At the end of this explanation, the patient will be asked whether he fully understands the explanation. If he is unsure or does not fully understand, he will be advised to see his doctor, or other appointed staff, for more explanation.

If the patient states that he fully understands the implications of giving consent, he will be given three choices:

- to keep the default setting for access control, in which case the advisory session will end;
- to withhold all data in the EHR from being shared with anyone, in which case the system will give an explanation of the possible consequences in general terms;
- to withhold only parts of the data in EHR, in which case it will trigger the next function of the advisory system.

If the next function of the system is triggered, it is necessary to obtain more detailed information regarding the patient's intentions. To do this, the system will ask a series of questions with possible answers listed as a drop down box. This approach is necessary to avoid free form inputs, which will be difficult to codify.

The first question to be asked is "Which part of the EHR?". To allow greater flexibility and encourage more information sharing, the answer to this question should be very granular in nature. For example the patient could allow access to his latest electrocardiography results, but could keep the results of a CD4 cell count (and the fact that it was performed) in a deeper, more secure layer of data (Schoenberg and Safran 200). However, this also requires categorising the medical terminology in a very granular way, and later to interpret it in suitable layperson-level language. Whilst this is not impossible to achieve, it is not practically feasible at present due to the lack of available resources and time within the project. Therefore for the prototype system, the answer to this question will be broad in category, and the patient has to choose the closest matching to his intents. At the time of writing, the actual category of answers is still to be decided, thus the following examples are neither exhaustive nor rigid:

- Coronary Hearth Disease
- Venereal disease
- Mental health
- History of violence
- Substance use disorder
- Others

For some of the categories above, it may be necessary to have sub-categories when the disease type carries substantially different risks for its unavailability, as well as from the legal and ethical view points. For example within the venereal disease, there could be further sub-categories of HIV and other STD. Note that there is also an option of "others". What this means is that if the patient wants to withhold information that is not listed here, he will be prompted to ask for guidance from his doctor or other healthcare professional.

The next question is "Who should the data be withheld from?" The example of answers for this could be:

- Withhold information from HCP (e.g. doctors)
- Withhold information from other carer (e.g. social services)
- Withhold information from other third party (a person or organisation, including governmental agencies)
- Withhold information from anyone apart from those originally involved in direct care
- Withhold information from family member
- Withhold information from neighbours/relations
- Withhold information from employers or insurance company

The above list is derived from the compendium of scenarios for security developed by Dr. Anthony Griew for the NHSIA (Griew 2000). From these scenarios, it was possible to extract the different types of individuals whom the patient may want to withhold their information from. Again, as the system will still need to undergo review process, this list may change.

A third question may be asked: "Under what condition?" with possible answers of:

- Withhold the information all the time *except* in emergency
- Withhold the information all the time *including* in emergency

Once the patient's intended action has been established, it should be matched against their health status. Since the prototype system will be standalone, establishing the patient's health status requires asking them a further question: "Do you have any of the following condition(s)?". As before, the patient will be presented with choices of answers. This time these answers are based on the patient's previous answers. For example, if the patient had chosen to withhold the venereal disease record from anyone apart from those originally involved in direct care, the list of answers here will include different types of venereal diseases such as HIV, Herpes, Chlamydia infections, etc.

Controller

The function of this unit is to pass on the patient's answers to the inference engine and return the results back to the patient portal. Furthermore, it also functions as a mediator or controller between all the back end units. As such, this unit contains meta-rules which will govern things such as precedence and dependencies of all the supporting units (for example the different rule categories being used at the inference engine). The existence of this unit will allow later development and addition of other supporting units as and when they are needed.

The output of the operation will be:

- Decision to allow or disallow the intended action
- Appropriate suggestions for alternative safer action(s)
- Explanations on the consequences of their intended actions according to rules that were fired during the operation. To make sense for the patient, this explanation will have to be presented in human rather than machine language.

Inference engine

This unit consists of sub-units of medical, legal and ethical rule engines. It takes the patient's answers and calculates the risk of the information in question against the patient's health status and the level of risk of unavailability of such information to the party from whom the information is withheld.

The medical reasoning sub-unit may have further sub-units, each may correspond to the categories of medical conditions (e.g. venereal disease, mental health, etc).

Another sub-unit, called explanation facility, is needed to interpret the rules that were fired during the operation into human friendly language. It will then be passed back on to the controller and subsequently to the patient portal.

System development and evaluation

The effectiveness of the system will be determined by how well the rules are constructed. It is therefore imperative to closely collaborate with one doctor or consultant in developing these rules. The rules will undergo a review process by different healthcare professionals from different fields of study to examine the logic and feasibility of its implementation before being coded into the system.

The end user's input will also be of paramount importance if the system will be usable at all. They will be involved during the evaluation and refinement of the system, which will be in several stages. Firstly, a small number of local HCPs who have been involved in the project will check the validity of the knowledge base. Then some representatives from the members of the public (e.g. colleagues) will be involved in the evaluation using hypothetical scenarios to see the functionality, effectiveness and user friendliness of the system. Their feedback will undoubtedly be valuable in refining the system further before it can be introduced to the wider audience. Time permitting, the system will also be demonstrated in front of many different healthcare professionals and patient associations nationally to gain wider acceptance.

FUTURE DEVELOPMENT

There is a considerable difficulty in terms of time and effort to obtain a fully informed express consent from the general public (Singleton 2002). The eventual integration of the advisory system into the internet accessible EHR will be an aid in reaching out to the members of the public.

When this system is integrated to the EHR, it runs as a background process at the patient's machine while he is accessing his health records. The process is automatically started once the patient has logged into the EHR. The advisory system will be first triggered when the first time the patient is accessing his EHR, offering general explanation regarding consent for information sharing. This should be done before any part of the record is made available electronically. The next time the advisory system is triggered again is only when the patient is changing the access control setting of his EHR, i.e. when he is trying to withhold one or more of his data.

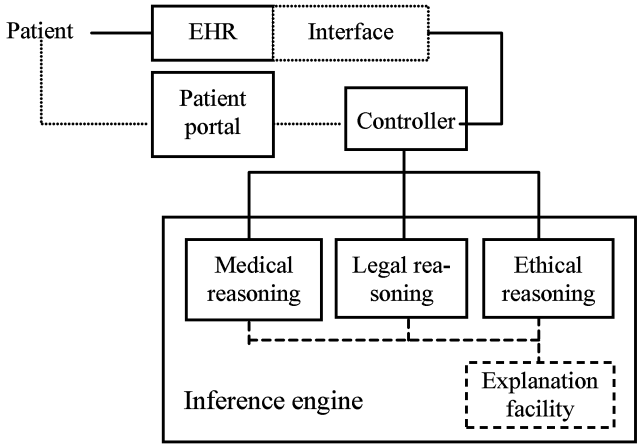


Figure 2: Future Development

The underlying principles of operation for the controller and the inference engine are similar to the prototype system. The difference is on the patient portal and the interfacing of the controller to the EHR (see Figure 2). The patient portal will no longer ask questions, but the controller will determine the patient's intended action and his health status directly from the EHR. The controller monitors the patient EHR and his activity in real time. The patient portal will come out as a pop-up window when the controller determines that the patient intended action will have serious consequences on his health or future care, or when there is better alternative action to be recommended.

A tight integration between the advisory system and the access control module of the EHR will make it possible to pose some degree of restrictions on the patient's action to prevent him from breaking the law but without violating his rights for privacy. For example when his intended action is clearly against the law, the advisory system will prompt the access control system to disable that action,

and give justifications to the patient. If the patient feels there is reason to dispute this restriction, he will be able to appeal this decision via a designated (human) data controller. On his favour, the patient's requirements to keep some of his information private, as expressed through the advisory system, will automatically be safely enacted and appropriately mapped into the access control mechanism.

CONCLUSION

The patient's autonomy on access control of their health information should be respected if the EHR is going to be attractive to the whole spectrum of individuals. On the other hand, patients need to know the consequences of giving and withholding access to their records, such that they will be able to make an educated choice in balancing between protecting their privacy and keeping up with the law and regulations, while at the same time making sure that their health information will be available to the right people at the right time, thus ensuring their own, and other's, safety. Achieving this goal would normally take a great deal of the patient's primary care doctor's time.

A patient advisory system will therefore be valuable in releasing most of these tasks from primary care doctors. This paper proposed a prototype model of such system, which can be used by the patient independently with only minimum help from his doctor. The system is very simple in its architectural design, but the real challenge will be in developing the rules to calculate the risks associated with the unavailability of health information to a particular patient with particular health status and history, whilst keeping up with the law and regulations as well as medical ethics. A close collaboration between the knowledge engineer, a healthcare professional and end user will be crucial in the success of the system development.

REFERENCES

Caldicott Committee. 1997. "Report on the review of patient-identifiable information." UK Dept. of Health http://www.doh.gov.uk/ipu/confiden/report/index.htm

Carr FP. 1997. "Web based EMR or Clinical Information System. Digital Med Inc." http://www.telemedical.com/webemr.htm

EU Parliament. 1995. "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data." *Legislative Document.* http://europa.eu.int/comm/internal_market/privacy/law_en.htm

Gaunt N. 2003. "Confidentiality and Consent. Use cases applicable to the shared electronic health record." http://www.nhsia.nhs.uk/erdip/pages/demonstrator/devo/devon_(34).pdf

Griew A. 2000. "A Compendium of Scenarios for Security." Commissioned by the NHS Information Authority. http://www.ihi.aber.ac.uk/web/publications/phss/download/security.pdf

HHS. 2003. "Medical Privacy – National Standards to Protect the Privacy of Personal Health Information." *US Dept. of Health & Human Services.* http://www.hhs.gov/ocr/hipaa/

HMSO[a]. 1998. "Data Protection Act 1998. Chapter 29." http://www.hmso.gov.uk/acts/acts1998/19980029.htm

HMSO[b]. 1998. "Human Rights Act 1998. Chapter 42." http://www.hmso.gov.uk/acts/acts1998/19980042.htm

NHSIA[a]. 2003. "Background to ERDIP." http://www.nhsia.nhs.uk/erdip/pages/backgroundtoerdip.asp

NHSIA[b]. 2003. "Data Quality." *Electronic Record Development and Implementation program. Lessons learned topic summary. No. 2.* http://www.nhsia.nhs.uk/erdip/pages/publications/ERDIPDataQualitySummary_3.pdf

Schoenberg R, and Safran C. 2000. "Internet based repository of medical records that retains patient confidentiality." *Information in practice.* BMJ 2000;321:1199-1203

Singleton P. 2002. "ERDIP Evaluation project. N5 - Patient Consent and Confidentiality." *Study Report.* http://www.nhsia.nhs.uk/erdip/pages/evaluation/docs/consentconfid/Consentstudyreport.pdf

AUTHOR LISTING

AUTHOR LISTING