15TH EUROPEAN CONCURRENT ENGINEERING CONFERENCE 2008

ECEC'2008

5th FUTURE BUSINESS TECHNOLOGY CONFERENCE

FUBUTEC'2008

EDITED BY

António E. S. Carvalho Brito

and

J. Manuel Feliz-Teixeira

APRIL 9-11, 2008

PORTO, PORTUGAL

A Publication of EUROSIS-ETI

Printed in Ghent, Belgium

15TH EUROPEAN CONCURRENT ENGINEERING CONFERENCE

5TH FUTURE BUSINESS TECHNOLOGY CONFERENCE

2008

PORTO, PORTUGAL

APRIL 9-11, 2008

Hosted by

UNIVERSITY OF PORTO

Organised by

ETI

Sponsored by

EUROSIS

Daimler AG

FEUP

Avnet 3D

Ghent University

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

Editors

General Conference Chair
Antonio Carvalho Brito
University of Porto
Porto, Portugal

General Program Chair
J. Manuel Feliz Teixeira
University of Porto
Porto, Portugal

Local Programme Committee

Manuel Augusto de Pina Marques, FEUP - Faculty of Engineering, University of Porto, Portugal
José António Barros Basto, , FEUP - Faculty of Engineering, University of Porto, Portugal
José Fernando Gonçalves, FEP - Faculty of Economics, University of Porto, Portugal
Guilherme Augusto Borges Pereira, School of Engineering, University of Minho

ECEC Honorary Chair

Uwe Baake
Daimler Buses, Evobus GmbH
,Neu-Ulm, Germany

ECEC International Programme Committee

1.E-Business in CE
Joachim Herbst, Daimler AG, Research and Technology, Ulm, Germany
Prof. Yo-Ping Huang, Tatung University, Taipei, Taiwan

2.Organization and Management
Abdelmalek Benzekri, IRIT, Universite Paul Sabatier, Toulouse, France
Iris Graessler, Robert Bosch GmbH, Automotive Electronics, Reutlingen, Germany

3.Supporting Technologies
Ricardo Goncalves, Univ. Nova de Lisboa, UNINOVA, Caparica, Portugal

4.Engineering Process Management and Modelling
Christian Andreas Schumann, University of Zwickau, Zwickau, Germany

5.Embedded Systems
Stefano Marrone, Second University of Naples, Caserta, Italy
Michael G. Salloker, FH JOANNEUM Gesellschaft mbH, Kapfenberg, Kapfenberg, Austria

6.Engineering Data Management and Information Modeling
Uwe Baake, Daimler Buses, Evobus GmbH, Neu-Ulm, Germany

7.Formal Methods
Mauro Iacono, Seconda Università di Napoli, Caserta, Italy
Rachel Moreau, Hasselt University, Diepenbeek, Belgium

8.Collaborative CE environments for virtual teams
Imed Kacem, University of Technology of Troyes, Troyes, France
Paul Stratil, Daimler AG, Stuttgart, Germany

9.CE Enhanced Lean Manufacturing
José António de Sousa Barros Basto, FEUP-University of Porto, Portugal
Jean-Claude Hochon, AIRBUS, Blagnac, France
Rik Van Landeghem, Ghent University, Zwijnaarde, Belgium

10.Practical Applications and Experiences
Irfan Manarvi, Center of Advanced Studies in Engineering, Islamabad, Pakistan
Jan Nesvadba, Philips Research Laboratories Eindhoven, Eindhoven, The Netherlands

FUBUTEC International Programme Committee

Statistical Analysis and Data Mining of Business Processes
Paulo Alexandre Ribeiro Cortez, University of Minho, Guimaraes, Portugal
Selwyn Piramithu, University of Florida, Gainesville, USA
Danny Van Welden, KBC Bank, Brussels, Belgium

Discrete Event Simulation and Queueing Systems in Business and Economics
Reza Azadegan, Urmia University, Urmia, Iran
J. Manuel Feliz Teixeiro, FEUP-University of Porto, Porto, Portugal
Henrikas Pranevicius, Kaunas University of Technology, Kaunas, Lithuania

Agents in Business Automation and Economics
Julie Dugdale, Institut IMAG, Grenoble cedex, France
Martin Fredriksson, BTH, Ronneby, Sweden
Isabel Praca, Polytechnic of Porto, Porto, Portugal

Simulation in Business and Economics
Henri Pierreval, IFMA, Aubiere Cedex, France
Enver Yucesan, INSEAD, Fontainebleau, France

Simulation in Business Games
Renate Sitte, Griffith University, Gold Coast, Australia

Simulation in OR and Knowledge Management
Chrissanthi Angeli, Technological Institute of Piraeus, Athens, Greece
Jacinto Davila, Universidad de Los Andes, Merida, Venezuela
Christophe Roche, Universite de Savoie, Le Bourget du Lac Cedex, France

Emergency Management and Risk Analysis Management
Konstantinos Kirytopoulos, University of the Aegean, Greece
Roger Smith, US Army PEO STRI
Lode Vermeersch, KBC Goldstate FMC Co, Shanghai, China P.R.

Simulation in E-Management, E-Government, E-Commerce and E-Trade
Vrassidas Leopoulos, Nat. technical University of Athens, Athens, Greece

Measuring the Costs and Effects of Environmental Policies for Business
Dietmar Boenke, Hochschule Reutlingen, Reutlingen, Germany
Elias A. Hadzilias, Université Catholique de Lille, Lille, France
Rinaldo Michelini, University of Genova, Genova, Italy

© 2008 EUROSIS-ETI

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

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

All articles published in these Proceedings have been peer reviewed

EUROSIS-ETI Publications are ISI-Thomson and INSPEC referenced

For permission to publish a complete paper write EUROSIS, c/o Philippe Geril, ETI Executive Director, Greenbridge NV, Wetenschapspark 1, Plassendale 1, B-8400 Ostend Belgium

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4, B-8400 Ostend, Belgium

Printed in Belgium by Reproduct NV, Ghent, Belgium
Cover Design by Grafisch Bedrijf Lammaing, Ostend, Belgium

EUROSIS-ETI Publication

ISBN: 978-9077381-39-7
EAN : 978-9077381-39-7

Preface

Concurrent Engineering and Business Technology are two of the most relevant fields of research in these days of profound change we are living in. The world and its economic and business forces show an incremental testing of the different paradigms, one against the other, leaving researchers worldwide no option but the intelligent adaptation of their solutions and concepts to the contours of the new reality. Concurrent Engineering, as an excellent tool for improving the efficiency of industrial and business processes, supported by an intelligent and creative utilization of technological advances, is, therefore, expected to contribute to surpassing such a challenge.

These conferences (ECEC-FUBUTEC) aim mainly at bringing together those, who around the world research these matters, in order to promote the spread of knowledge to the global community. We hope that this meeting will also contribute to a better visibility of these subjects.

Although this conference has been the result of the work of several personalities, we would like specially to mention the dedication of the committee members and the fundamental contribution of Mr Philippe Geril, for the excellent organization that allowed this conference to be held in Porto within the EUROSIS standards. We would also like to refer to the support given for this event by the Faculty of Engineering of the University of Porto.

We finally hope, you have an agreeable working experience during these couple of days at the Faculty of Engineering, and an exciting stay in Porto, a city as beautiful as ever.

The ECEC-FUBUTEC chairmen,



António E. S. Carvalho Brito



J. Manuel Feliz-Teixeira

CONTENTS

Preface ..... **VII**
Scientific Programme **1**
Author Listing **141**

DATA MANAGEMENT

Using Data Mining to Predict Secondary School Student Performance
Paulo Cortez and Alice Silva **5**

Electronic Health Record in the TTF Framework
Elena Ciancimino and Salvatore Cannella **13**

PERFORMANCE ESTIMATION AND FORECASTING

MPSOC Design Flow Based of Performance Estimation
Kamel Smiri, Abderrazek Jemai, Nihel Chibani and Mohamed Moalla **19**

Signal-Based predictive Condition Monitoring
Egon Müller and Steffen Leischnig **27**

SCHEDULING METHODOLOGY

A Robust What-if Bullwhip Analysis
Salvatore Cannella and Elena Ciancimino **33**

Holistic Metrics Applied to a Practical Supply Chain Study
António E. S. Carvalho Brito and J. Manuel Feliz-Teixeira **41**

Heuristic Approach as a Way to Improve Scheduling in ERP/APS Systems
František Koblasa, Luís S. Dias, José A. Oliveira and Guilherme Pereira **47**

EMERGENCY, RISK AND SAFETY CRITICAL SYSTEMS

The FLOREAN System: Emergent Flood Prediction
Ivo Vondrák, Jan Martinovič, Jan Kožusznik, Svatopluk Štolfa and
Jan Unucka **55**

Defining Project Risk Management Efficiency through Simulation
Konstantinos Kirytopoulos, Viktor Diamantas and Vrassidas Leopoulos **58**

CONTENTS

ICT Security Management and Risk Management
Jeanne Schreurs and Rachel Moreau.....**63**

STRUCTURES, ARCHITECTURES AND MODELS

Extension of Formal Concept Theory for Organization Structure Analysis
Jan Kozusznik and Ivo Vondrak.....**71**

Towards an Integrated Corporate Information System: Building up a Sales Product Data Model to cope with the emerging Complexity of E-Business
Christian-Andreas Schumann and Andreas Rutsch.....**74**

An Architecture for Collaborative Working Environments in Manufacturing Industry
Ana Correia, Cristina Grama, Dragan Stokic, Ana Campos,
Rui Neves Silva and Jon Aguirre **79**

WORKFLOW SIMULATION

Partial Recognizing Algorithm for Verification of Workflow Processes
Armen Kostanyan and Anna Varosyan **89**

Modelling Time-Constrained Collaborative Workflows based on timed Recursive Ecatnets
Kamel Barkaoui and Awatef Hicheur **95**

LOGISTICS APPLICATIONS

Assessment of Delivery Strategies by applying lexicographic Optimization and Design of Experiments: A Case Study
Herminia I. Calvete, Carmen Gale, Maria-Jose Oliveros,
and Belen Sanchez-Valverde **103**

A Linear Programming Model for the Storage Location Assignment Problem – A case Study
Carla Geraldes, Sameiro Carvalho and Guilherme Pereira **111**

Using 3-D Simulation in an Industry Internal Logistic Process
Vik Pavel, Luís S. Dias, José A. Oliveira and Guilherme Pereira.....**116**

CASE STUDIES

Simulation Building Blocks for Business Engineering of Mobile Field Services
Tennov Simanjuntak, Yan Wang, Marijn Janssen and Alexander Verbraeck **...123**

Quality Management in the Design of the Construction
José Carvalho, Jean-Claude Mangin and Gérard Sauce **.....128**

Definition of a Logistics Model in the Construction Industry – Case Study
Tiago Pinho, José Telhada and Maria Sameiro Carvalho **.....133**

SCIENTIFIC PROGRAMME

DATA MANAGEMENT

USING DATA MINING TO PREDICT SECONDARY SCHOOL STUDENT PERFORMANCE

Paulo Cortez and Alice Silva
Dep. Information Systems/Algoritmi R&D Centre
University of Minho
4800-058 Guimarães, PORTUGAL
Email: pcortez@dsi.uminho.pt, alicegsilva@gmail.com

KEYWORDS
Business Intelligence in Education, Classification and Regression, Decision Trees, Random Forest

ABSTRACT
Although the educational level of the Portuguese population has improved in the last decades, the statistics keep Portugal at Europe's tail end due to its high student failure rates. In particular, lack of success in the core classes of Mathematics and the Portuguese language is extremely serious. On the other hand, the fields of Business Intelligence (BI)/Data Mining (DM), which aim at extracting high-level knowledge from raw data, offer interesting automated tools that can aid the education domain. The present work intends to approach student achievement in secondary education using BI/DM techniques. Recent real-world data (e.g. student grades, demographic, social and school related features) was collected by using school reports and questionnaires. The two core classes (i.e. Mathematics and Portuguese) were modeled under binary/five-level classification and regression tasks. Also, four DM models (i.e. Decision Trees, Random Forest, Neural Networks and Support Vector Machines) and three input selections (e.g. with and without previous grades) were tested. The results show that a good predictive accuracy can be achieved, provided that the first and/or second school period grades are available. Although student achievement is highly influenced by past evaluations, an explanatory analysis has shown that there are also other relevant features (e.g. number of absences, parent's job and education, alcohol consumption). As a direct outcome of this research, more efficient student prediction tools can be be developed, improving the quality of education and enhancing school resource management.

INTRODUCTION

Education is a key factor for achieving a long-term economic progress. During the last decades, the Portuguese educational level has improved. However, the statistics keep the Portugal at Europe's tail end due to its high student failure and dropping out rates. For example, in 2006 the early school leaving rate in Portugal was 40% for 18 to 24 year olds, while the European Union average value was just 15% (Eurostat 2007). In particular, failure in the core classes of Mathematics and Portuguese (the native language) is extremely serious, since they provide fundamental knowledge for the success in the remaining school subjects (e.g. physics or history).

On the other hand, the interest in Business Intelligence (BI)/Data Mining (DM) (Turban et al. 2007), arose due to the advances of Information Technology, leading to an exponential growth of business and organizational databases. All this data holds valuable information, such as trends and patterns, which can be used to improve decision making and optimize success. Yet, human experts are limited and may overlook important details. Hence, the alternative is to use automated tools to analyze the raw data and extract interesting high-level information for the decision-maker.

The education arena offers a fertile ground for BI applications, since there are multiple sources of data (e.g. traditional databases, online web pages) and diverse interest groups (e.g. students, teachers, administrators or alumni) (Ma et al. 2000). For instance, there are several interesting questions for this domain that could be answered using BI/DM techniques (Luan 2002, Minaei-Bidgoli et al. 2003): Who are the students taking most credit hours? Who is likely to return for more classes? What type of courses can be offered to attract more students? What are the main reasons for student transfers? Is is possible to predict student performance? What are the factors that affect student achievement? This paper will focus in the last two questions. Modeling student performance is an important tool for both educators and students, since it can help a better understanding of this phenomenon and ultimately improve it. For instance, school professionals could perform corrective measures for weak students (e.g. remedial classes).

In effect, several studies have addressed similar topics. Ma et al. (2000) applied a DM approach based in Association Rules in order to select weak tertiary school students of Singapore for remedial classes. The input variables included demographic attributes (e.g. sex, region) and school performance over the past years and the proposed solution outperformed the traditional allocation procedure. In 2003 (Minaei-Bidgoli et al. 2003),

online student grades from the Michigan State University were modeled using three classification approaches (i.e. binary: pass/fail; 3-level: low, middle, high; and 9-level: from 1 - lowest grade to 9 - highest score). The database included 227 samples with online features (e.g. number of corrected answers or tries for homework) and the best results were obtained by a classifier ensemble (e.g. Decision Tree and Neural Network) with accuracy rates of 94% (binary), 72% (3-classes) and 62% (9-classes). Kotsiantis et al. (2004) applied several DM algorithms to predict the performance of computer science students from an university distance learning program. For each student, several demographic (e.g. sex, age, marital status) and performance attributes (e.g. mark in a given assignment) were used as inputs of a binary pass/fail classifier. The best solution was obtained by a Naive Bayes method with an accuracy of 74%. Also, it was found that past school grades have a much higher impact than demographic variables. More recently, Pardos et al. (2006) collected data from an online tutoring system regarding USA 8th grade Math tests. The authors adopted a regression approach, where the aim was to predict the math test score based on individual skills. The authors used Bayesian Networks and the best result was an predictive error of 15%.

In this work, we will analyze recent real-world data from two Portuguese secondary schools. Two different sources were used: mark reports and questionnaires. Since the former contained scarce information (i.e. only the grades and number of absences were available), it was complemented with the latter, which allowed the collection of several demographic, social and school related attributes (e.g. student's age, alcohol consumption, mother's education). The aim is to predict student achievement and if possible to identify the key variables that affect educational success/failure. The two core classes (i.e. Mathematics and Portuguese) will be modeled under three DM goals:

i) binary classification (pass/fail);
ii) classification with five levels (from I very good or excellent to V - insufficient); and
iii) regression, with a numeric output that ranges between zero (0%) and twenty (100%).

For each of these approaches, three input setups (e.g. with and without the school period grades) and four DM algorithms (e.g. Decision Trees, Random Forest) will be tested. Moreover, an explanatory analysis will be performed over the best models, in order to identify the most relevant features.

MATERIALS AND METHODS

Student Data

In Portugal, the secondary education consists of 3 years of schooling, preceding 9 years of basic education and followed by higher education. Most of the students join the public and free education system. There are several courses (e.g. Sciences and Technologies, Visual Arts) that share core subjects such as the Portuguese Language and Mathematics. Like several other countries (e.g. France or Venezuela), a 20-point grading scale is used, where 0 is the lowest grade and 20 is the perfect score. During the school year, students are evaluated in three periods and the last evaluation (G3 of Table 1) corresponds to the final grade.

This study will consider data collected during the 2005-2006 school year from two public schools, from the Alentejo region of Portugal. Although there has been a trend for an increase of Information Technology investment from the Government, the majority of the Portuguese public school information systems are very poor, relying mostly on paper sheets (which was the current case). Hence, the database was built from two sources: school reports, based on paper sheets and including few attributes (i.e. the three period grades and number of school absences); and questionnaires, used to complement the previous information. We designed the latter with closed questions (i.e. with predefined options) related to several demographic (e.g. mother's education, family income), social/emotional (e.g. alcohol consumption) (Pritchard and Wilson 2003) and school related (e.g. number of past class failures) variables that were expected to affect student performance. The questionnaire was reviewed by school professionals and tested on a small set of 15 students in order to get a feedback. The final version contained 37 questions in a single A4 sheet and it was answered in class by 788 students. Latter, 111 answers were discarded due to lack of identification details (necessary for merging with the school reports). Finally, the data was integrated into two datasets related to Mathematics (with 395 examples) and the Portuguese language (649 records) classes.

During the preprocessing stage, some features were discarded due to the lack of discriminative value. For instance, few respondents answered about their family income (probably due to privacy issues), while almost 100% of the students live with their parents and have a personal computer at home. The remaining attributes are shown in Table 1, where the last four rows denote the variables taken from the school reports.

Data Mining Models

Classification and regression are two important DM goals. Both require a supervised learning, where a model is adjusted to a dataset made up of $k \in \{1, ..., N\}$ examples, each mapping an input vector $(x_1^k, \ldots, x_I^k)$ to a given target y_k. The main difference is set in terms of the output representation, (i.e. discrete for classification and continuous for regression). In classification, models are often evaluated using the Percentage of Correct Classifications (PCC), while in regression the Root

Table 1: The preprocessed student related variables

Attribute	Description (Domain)
sex	student's sex (binary: female or male)
age	student's age (numeric: from 15 to 22)
school	student's school (binary: *Gabriel Pereira* or *Mousinho da Silveira*)
address	student's home address type (binary: urban or rural)
Pstatus	parent's cohabitation status (binary: living together or apart)
Medu	mother's education (numeric: from 0 to 4[a])
Mjob	mother's job (nominal[b])
Fedu	father's education (numeric: from 0 to 4[a])
Fjob	father's job (nominal[b])
guardian	student's guardian (nominal: mother, father or other)
famsize	family size (binary: ≤ 3 or > 3)
famrel	quality of family relationships (numeric: from 1 – very bad to 5 – excellent)
reason	reason to choose this school (nominal: close to home, school reputation, course preference or other)
traveltime	home to school travel time (numeric: 1 – $<$ 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – $>$ 1 hour).
studytime	weekly study time (numeric: 1 – $<$ 2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours or 4 – $>$ 10 hours)
failures	number of past class failures (numeric: n if $1 \leq n < 3$, else 4)
schoolsup	extra educational school support (binary: yes or no)
famsup	family educational support (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
paidclass	extra paid classes (binary: yes or no)
internet	Internet access at home (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
freetime	free time after school (numeric: from 1 – very low to 5 – very high)
goout	going out with friends (numeric: from 1 – very low to 5 – very high)
Walc	weekend alcohol consumption (numeric: from 1 – very low to 5 – very high)
Dalc	workday alcohol consumption (numeric: from 1 – very low to 5 – very high)
health	current health status (numeric: from 1 – very bad to 5 – very good)
absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20)

a 0 – none, 1 – primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education.
b teacher, health care related, civil services (e.g. administrative or police), at home or other.

Mean Squared (RMSE) is a popular metric (Witten and Frank 2005). A high PCC (i.e. near 100%) suggests a good classifier, while a regressor should present a low global error (i.e. RMSE close to zero). These metrics can be computed using the equations:

$$\begin{aligned} \Phi(i) &= \begin{cases} 1 & , \quad \text{if } y_i = \widehat{y}_i \\ 0 & , \quad \text{else} \end{cases} \\ PCC &= \sum_{i=1}^{N} \Phi(i)/N \times 100\,(\%) \\ RMSE &= \sqrt{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2/N} \end{aligned} \tag{1}$$

where $\widehat{y}_i$ denotes the predicted value for the i-th example.
In this work, the Mathematics and Portuguese grades (i.e. G3 of Table 1) will be modeled using three supervised approaches:

1. **Bin**ary classification – *pass* if G3$\geq$10, else *fail*;
2. **5-L**evel classification – based on the Erasmus[1] grade conversion system (Table 2);
3. **Reg**ression – the G3 value (numeric output between 0 and 20).

Figure 1 plots the respective histograms.
Several DM algorithms, each one with its own purposes and capabilities, have been proposed for classification and regression tasks. The Decision Tree (DT) is a

[1] European exchange programme that enables student exchange in 31 countries.

Table 2: The five-level classification system

Country	I (excellent/very good)	II (good)	III (satisfactory)	IV (sufficient)	V (fail)
Portugal/France	16-20	14-15	12-13	10-11	0-9
Ireland	A	B	C	D	F

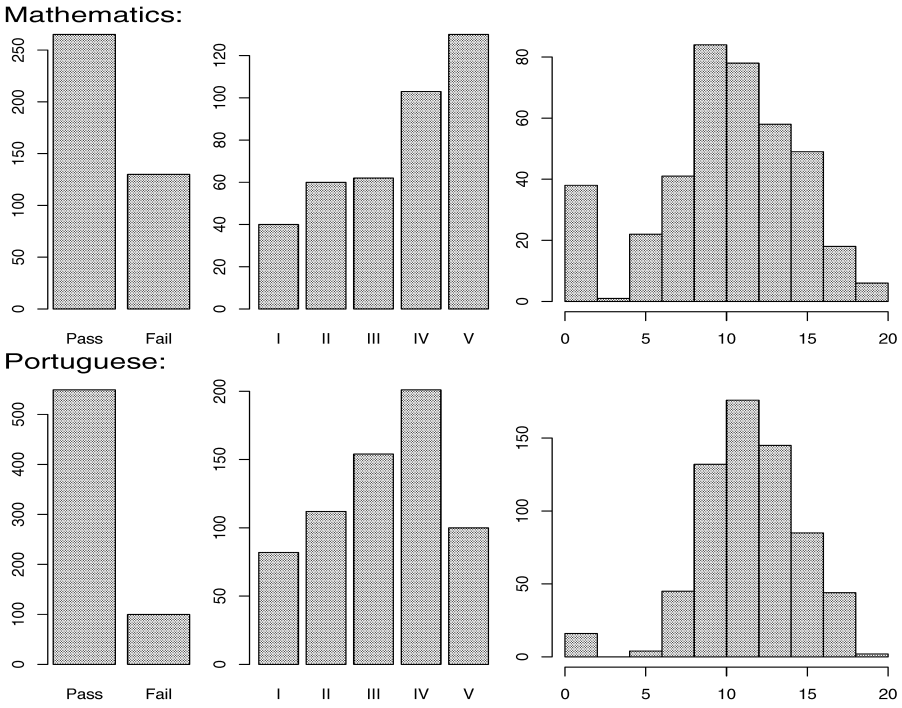


Figure 1: Histograms for the output variables (binary/5-level classification and regression)

branching structure that represents a set of rules, distinguishing values in a hierarchical form (Breiman et al. 1984). This representation can translated into a set of IF-THEN rules, which are easy to understand by humans. The Random Forest (RF) (Breiman 2001) is an ensemble of T unpruned DT. Each tree is based in a random feature selection from bootstrap training samples and the RF predictions are built by averaging the outputs of the T trees. The RF is more difficult to interpret when compared with the single DT, although it is still possible to provide explanatory knowledge in terms of its input variable relevance. Nonlinear functions, such as Neural Networks (NN) and Support Vector Machines (SVM), have also been proposed for DM tasks (Hastie et al. 2001), obtaining better results when a high nonlinearity is present. In this work, the NN model is based in the popular multilayer perceptron, with one hidden layer with H hidden nodes, while the SVM will use a gaussian kernel with one hyperparameter (γ). It should be noted that NN and SVM use model representations that are difficult to understand by humans. Also, NN and SVM are more affected by irrelevant inputs than the DT/RF algorithms, since the latter explicitly perform an internal feature selection.

Computational Environment

All experiments reported in this study were conducted using the **RMiner** (Cortez In press), an open source library for the **R** environment that facilitates the use of DM techniques (Figure 2). **R** is a free and high-level matrix programming language with a powerful suite of tools for statistical and data analysis (R Development Core Team 2006). It can run in multiple platforms (e.g. *Windows*, *MacOS* or *Linux*) and new features can be added by creating packages.

The **RMiner** library[2] presents a set of coherent functions (e.g. `mining`, `saveMining`) for classification and regression tasks (Cortez In press). In particular, the library uses the **rpart** (DT), **randomForest** (RF), **nnet** (NN) and **kernlab** (SVM) packages. As an example, the following **RMiner/R** code was used during the DT predictive experiments of the portuguese binary classi-

[2]http://www.dsi.uminho.pt/~pcortez/R/rminer.zip

fication:

```
source("rminer.R") # read the RMiner library
# load the data:
data=read.table("a-por-bin.dat",sep=";",header=T)
K=c("kfold",10) # 10-fold cross-validation
# execute 10 runs of a DT classification:
DT=mining(y~.,data,model="dt",Runs=20,method=K)
# show mean classification error on test set:
print(mean(DT$error))
# save the results (predictions, ...):
saveMining(DT,"DT-results")
```

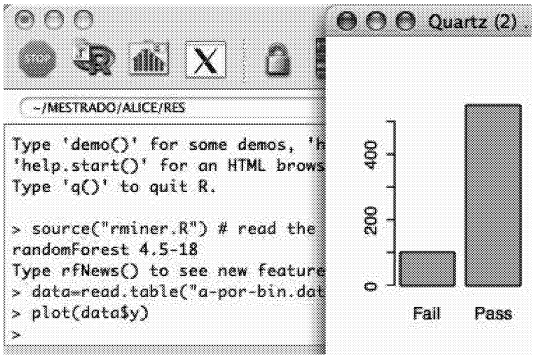


Figure 2: MacOS example of the R/RMiner tool.

RESULTS

Predictive Performance

Before fitting the models, some preprocessing was required by the NN and SVM models. The nominal[3] variables (e.g **Mjob**) were transformed into a *1-of-C* encoding and all attributes were standardized to a zero mean and one standard deviation (Hastie et al. 2001). Next, the DM models were fitted. The DT node split was adjusted for the reduction of the sum of squares. Regarding the remaining methods, the default parameters were adopted for the RF (e.g. $T = 500$), NN (e.g. $E = 100$ epochs of the BFGS algorithm) and SVM (e.g. Sequential Minimal Optimization algorithm). Also, the NN and SVM hyperparameters were optimized using an internal grid search (i.e. using only training data) where $H \in \{0, 2, 4, 6, 8\}$ and $\gamma \in \{2^{-9}, 2^{-7}, 2^{-5}, 2^{-3}, 2^{-1}\}$.
Since we suspected that the G1 and G2 grades would have a high impact, three input configurations were tested for each DM model:

- **A** - with all variables from Table 1 except G3 (the output);
- **B** - similar to **A** but without G2 (the second period grade); and
- **C** - similar to **B** but without G1 (the first period grade).

[3]Discrete with more than two non-ordered values.

To access the predictive performances, 20 runs of a 10-fold cross-validation (Hastie et al. 2001) (in a total of 200 simulations) were applied to each configuration. Under such scheme, for a given run the data is randomly divided in 10 subsets of equal size. Sequentially, one different subset is tested (with 10% of the data) and the remaining data used to fit the DM technique. At the end of this process, the evaluated test set contains the whole dataset, although 10 variations of the same DM model are used to create the predictions.
As a baseline comparison, a naive predictor (NV) will also be tested. For the **A** setup, this model is equal to the second period grade (G2 or its binary/5-level versions). When the second grade is not available (**B** setup), the first period grade is used (or its binary/5-level variants). If no evaluation is present (**C** setup) then the most common class (for classification tasks) or the average output value (regression) is returned.
The test set results are shown in Tables 3 to 5 in terms of the mean and respective t-student 95% confidence intervals (Flexer 1996). As expected, the **A** setup achieves the best results. The predictive performance decreases when the second period grade is not known (**B**) and the wost results are obtained when no student scores are used (**C**). Using only the last available evaluation (NV method for the first two input setups) is the best option for the Mathematics classification goals (both binary and 5-level) and Portuguese regression under the **A** input selection. This suggests that in these cases the non evaluation inputs are of no use. However, the scenario changes for the remaining experiments. The RF is the best choice in 8 cases, followed by the DT, which obtains 4 best results. In general, the nonlinear function methods (NN and SVM) are outperformed by the tree based ones. This behavior may be explained by a high number of irrelevant inputs, as shown in the next section.
As an example of the quality of the predictions, Figure 3 shows the confusion matrices for the DT algorithm, Portuguese class and **A** setup. In both binary and 5-level classification, the majority of the values are in/near the matrix diagonal, revealing a good fit. The figure also presents the RF scatter plot (predicted versus the observed values) for the Mathematics regression approach. The RF predictions are close to the diagonal line (which denotes the perfect forecast) within most of the output range (i.e. $[5, 20]$).

Descriptive Knowledge

Here the goal is not to infer about the predictive capabilities of each model, as measured in the previous section, but to give a simple description that summarizes the best DM models. Thus, the whole dataset will be used in the descriptive experiments. Furthermore, only the DT/RF algorithms will be considered, since it is more easy to extract knowledge from these models and they

Table 3: Binary classification results (PCC values, in %; underline – best model; **bold** – best within the input setup)

Input	Mathematics					Portuguese				
Setup	NV	NN	SVM	DT	RF	NV	NN	SVM	DT	RF
A	**91.9**$^{\dagger}{\pm 0.0}$	88.3± 0.7	86.3± 0.6	90.7± 0.3	91.2± 0.2	89.7± 0.0	90.7± 0.5	91.4± 0.2	**93.0**$^{\dagger}{\pm 0.3}$	92.6± 0.1
B	**83.8**$^{\dagger}{\pm 0.0}$	81.3± 0.5	80.5± 0.5	83.1± 0.5	83.0± 0.4	87.5± 0.0	87.6± 0.4	88.0± 0.3	88.4± 0.3	**90.1**$^{\dagger}{\pm 0.2}$
C	67.1± 0.0	66.3± 1.0	**70.6**$^{*}{\pm 0.4}$	65.3± 0.8	70.5± 0.5	84.6± 0.0	83.4± 0.5	84.8± 0.3	84.4± 0.4	**85.0**$^{*}{\pm 0.2}$

† – statistical significance under pairwise comparisons with other methods.
$*$ – statistical significance under a pairwise comparison with NV.

Table 4: Five-level classification results (PCC values, in %; underline – best model; **bold** – best within the input setup)

Input	Mathematics					Portuguese				
Setup	NV	NN	SVM	DT	RF	NV	NN	SVM	DT	RF
A	**78.5**$^{\dagger}{\pm 0.0}$	60.3± 1.6	59.6± 0.9	76.7± 0.4	72.4± 0.4	72.9± 0.0	65.1± 0.9	64.5± 0.6	**76.1**± 0.1	73.5± 0.2
B	**60.5**$^{\dagger}{\pm 0.0}$	49.8± 1.2	47.9± 0.7	57.5± 0.8	52.7± 0.6	58.7± 0.0	52.0± 0.6	51.7± 0.6	**62.9**± 0.2	55.3± 0.4
C	32.9± 0.0	30.4± 1.0	31.0± 0.7	31.5± 0.6	**33.5**± 0.6	31.0± 0.0	33.7± 0.6	34.9± 0.5	32.8± 0.6	**36.7**$^{\dagger}{\pm 0.6}$

† – statistical significance under pairwise comparisons with other methods.

Table 5: Regression results (RMSE values; underline – best model; **bold** – best within the input setup)

I.	Mathematics					Portuguese				
S.	NV	NN	SVM	DT	RF	NV	NN	SVM	DT	RF
A	2.01± 0.00	2.05± 0.02	2.09± 0.02	1.94± 0.04	**1.75**$^{\dagger}{\pm 0.01}$	**1.32**± 0.00	1.36± 0.04	1.35± 0.01	1.46± 0.03	**1.32**± 0.00
B	2.80± 0.00	2.82± 0.02	2.90± 0.02	2.67± 0.04	**2.46**$^{\dagger}{\pm 0.01}$	1.89± 0.00	1.88± 0.02	1.87± 0.01	**1.78**$^{*}{\pm 0.03}$	1.79± 0.01
C	4.59± 0.00	4.41± 0.03	4.37± 0.03	4.46± 0.04	**3.90**$^{\dagger}{\pm 0.01}$	3.23± 0.00	2.79± 0.02	2.76± 0.02	2.93± 0.02	**2.67**$^{\dagger}{\pm 0.01}$

† – statistical significance under pairwise comparisons with other methods.
$*$ – statistical significance under a pairwise comparison with NV.

achieved in general the best predictive performances.

Table 6 presents the relative importance (in percentage) of each input variable as measured by the RF algorithm (Breiman 2001). To clarify the analysis, only the five most relevant results are shown in the table. These five variables present an overall impact that ranges from 43% to 77%, which indicates that there is a high number of irrelevant inputs. As expected, the student evaluations have a high impact (from 23% to 46%) in the models. For instance, G2 is the most important feature for the **A** input selection example, while G1 is highly relevant for the **B** setup. Also, the number of past failures, which is related with previous student performance, is the most important factor when no student scores are available. Nevertheless, there are other relevant factors, such as school related (e.g. number of absences, extra school support or travel time), demographic (e.g. the mother's job) and social (e.g. going out with friends, alcohol consumption) variables.

Figure 4 plots the best four DT. Again, the student grades are the most relevant features, appearing at the root of the trees, and only a small number (2 to 5) of the inputs considered are used. There are some interesting rules that can be extracted from these trees, for instance:

1. if G1$<10 \wedge$ G2=9 $\wedge$ Mjob $\in$ {teacher, other} then *pass*;
2. if G1$<10 \wedge$ G2=9 $\wedge$ Mjob $\in$ {home, health, civil services} then *fail*;
3. if 12 $\leq$G2$<14 \wedge$ goout>1 then III;
4. if 12 $\leq$G2$<14 \wedge$ goout$=1$ then II;
5. if 11 $<$G1$\leq 13 \wedge$ absences<7 then 13;
6. if 11 $<$G1$\leq 13 \wedge$ absences≥ 7 then 11;

These rules show the influence of the mother's job (rules 1–2), going out with friends (rules 3–4) and the number of absences (rules 5–6).

CONCLUSIONS

Education is a crucial element in our society. Business Intelligence (BI)/Data Mining (DM) techniques, which

Table 6: Relative importance of the input variables for the best RF models

Setup	Relative Importance
C-Mat-Bin	failures: 21.8%, absenses: 9.4%, schoolsup: 7.0%, goout: 6.5%, higher: 6.4%
B-Por-Bin	G1: 22.8%, failures: 14.4%, higher: 11.9%, school: 8.1%, Mjob: 4.1%
C-Por-Bin	failures: 16.8%, school: 13.2%, higher: 13.1%, traveltime: 5.9, famrel: 5.7%
C-Mat-5L	failures: 18.3%, schoolsup: 9.5%, sex: 5.7%, absences: 5.6%, Medu: 4.5%
C-Por-5L	failures: 16.8%, higher: 9.9%, school: 9.3%, schoolsup: 6.9%, Walc: 6.6%
A-Mat-Reg	G2: 30.5%, absenses: 20.6%, G1: 15.4%, failures: 6.7%, age: 4.2%
B-Mat-Reg	G1: 42.2%, absenses: 18.6%, failures: 8.9%, age: 3.3%, schoolsup: 3.2%
C-Mat-Reg	failures: 19.7%, absenses: 18.9%, schoolsup: 8.3%, higher: 5.4%, Mjob:4.2%
C-Por-Reg	failures: 20.7%, higher: 11.4%, schoolsup: 6.9%, school: 6.7%, Medu:5.6%

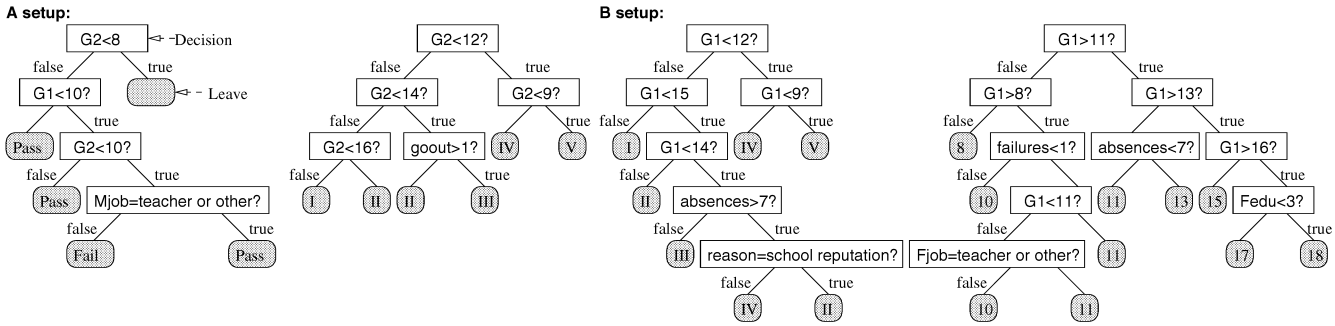


Figure 4: Examples of Decision Trees (A-Por-Bin, A-Por-5L, B-Por-5L and B-Por-Reg)

allow a high level extraction of knowledge from raw data, offer interesting possibilities for the education domain. In particular, several studies have used BI/DM methods to improve the quality of education and enhance school resource management.

In this paper, we have addressed the prediction of secondary student grades of two core classes (Mathematics and Portuguese) by using past school grades (first and second periods), demographic, social and other school related data. Three different DM goals (i.e. binary/5-level classification and regression) and four DM methods, i.e. Decision Trees (DT), Random Forests (RF), Neural Networks (NN) and Support Vector Machines (SVM), were tested. Also, distinct input selections (e.g. with or without past grades) were explored. The obtained results reveal that it is possible to achieve a high predictive accuracy, provided that the first and/or second school period grades are known. This confirms the conclusion found in (Kotsiantis et al. 2004): student achievement is highly affected by previous performances. Nevertheless, an analysis to knowledge provided by the best predictive models has shown that, in some cases, there are other relevant features, such as: school related (e.g. number of absences, reason to choose school, extra educational school support), demographic (e.g. student's age, parent's job and education) and social (e.g. going out with friends, alcohol consumption) variables.

This study was based on an off-line learning, since the DM techniques were applied after the data was collected. However, there is a potential for an automatic on-line learning environment, by using a student prediction engine as part of a school management support system. This will allow the collection of additional features (e.g. grades from previous school years) and also to obtain a valuable feedback from the school professionals. Furthermore, we intent to enlarge the experiments to more schools and school years, in order to enrich the student databases. Automatic feature selection methods (e.g. filtering or wrapper) (Witten and Frank 2005) will also be explored, since only a small portion of the input variables considered seem to be relevant. In particular, this is expected to benefit the nonlinear function methods (e.g. NN and SVM), which are more sensitive to irrelevant inputs. More research is also needed (e.g. sociological studies) in order to understand why and how some variables (e.g. reason to choose school, parent's job or alcohol consumption) affect student performance.

ACKNOWLEDGMENTS

We wish to thank the students and school professionals who voluntarily participated in this study.

	Pass	Fail
Pass	**480**	69
Fail	12	**88**

	I	II	III	IV	V
I	**44**	33	4	1	0
II	2	**54**	42	14	0
III	0	19	**84**	48	3
IV	0	0	24	**111**	66
V	0	0	0	1	**88**

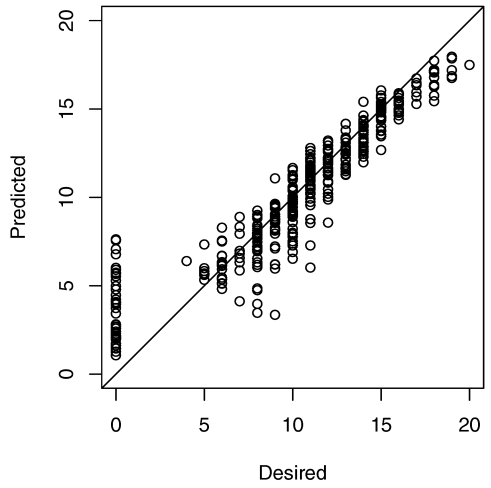


Figure 3: Example of the predictions on the test set: DT confusion matrices for A-Por-Bin and A-Por-5L setups (top); and RF scatter plot for A-Mat-Reg (bottom)

REFERENCES

Breiman L., 2001. *Random Forests. Machine Learning*, 45, no. 1, 5–32.

Breiman L.; Friedman J.; Ohlsen R.; and Stone C., 1984. *Classification and Regression Trees.* Wadsworth, Monterey, CA.

Cortez P., In press. *RMiner: Data Mining with Neural Networks and Support Vector Machines using R.* In R. Rajesh (Ed.), *Introduction to Advanced Scientific Softwares and Toolboxes.*

Eurostat, 2007. *Early school-leavers.* http://epp.eurostat.ec.europa.eu/.

Flexer A., 1996. *Statistical Evaluation of Neural Networks Experiments: Minimum Requirements and Current Practice.* In *Proceedings of the 13th European Meeting on Cybernetics and Systems Research.* Vienna, Austria, vol. 2, 1005–1008.

Hastie T.; Tibshirani R.; and Friedman J., 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag, NY, USA.

Kotsiantis S.; Pierrakeas C.; and Pintelas P., 2004. *Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. Applied Artificial Intelligence (AAI)*, 18, no. 5, 411–426.

Luan J., 2002. *Data Mining and Its Applications in Higher Education. New Directions for Institutional Research*, 113, 17–36.

Ma Y.; Liu B.; Wong C.; Yu P.; and Lee S., 2000. *Targeting the right students using data mining.* In *Proc. of 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Boston, USA, 457–464.

Minaei-Bidgoli B.; Kashy D.; Kortemeyer G.; and Punch W., 2003. *Predicting student performance: an application of data mining methods with an educational web-based system.* In *Proc. of IEEE Frontiers in Education.* Colorado, USA, 13–18.

Pardos Z.; Heffernan N.; Anderson B.; and Heffernan C., 2006. *Using Fine-Grained Skill Models to Fit Student Performance with Bayesian Networks.* In *Proc. of 8th Int. Conf. on Intelligent Tutoring Systems.* Taiwan.

Pritchard M. and Wilson S., 2003. *Using Emotional and Social Factors To Predict Student Success. Journal of College Student Development*, 44, no. 1, 18–28.

R Development Core Team, 2006. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-00-3, http://www.R-project.org.

Turban E.; Sharda R.; Aronson J.; and King D., 2007. *Business Intelligence, A Managerial Approach.* Prentice-Hall.

Witten I. and Frank E., 2005. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann, San Francisco, CA.

BIBLIOGRAPHIE

PAULO CORTEZ was born in Braga, Portugal and went to the University of Minho where he obtained an Eng. Degree (1995), a M.Sc. (1998) and a Ph.D (2002) in Computer Science. In 2001, he became Lecturer at the Dep. of Information Systems and Researcher at the Algoritmi R&D Centre. He participated in 7 R&D projects (principal investigator in 2) and he is co-author of more than fifty publications in int. conferences and journals. Research interests: Business Intelligence/Data Mining; Neural and Evolutionary Computation; Forecasting. Web-page: http://www.dsi.uminho.pt/˜pcortez

ALICE SILVA studied Informations Systems, obtaining a Degree (Bragança Polytechnic Inst., 2003) and a M.Sc. (U. Minho, 2007). Currently, she teaches Informatics at the secondary school of Arcozelo, Portugal.

ELECTRONIC HEALTH RECORD IN THE TTF FRAMEWORK

Elena Ciancimino
Salvatore Cannella
Faculty of Engineering, University of Palermo
Parco d'Orleans, 90128 Palermo, Italy
eciancimino@unipa.it, cannella@unipa.it

KEYWORDS
Health care, decision-support systems, model testing.

ABSTRACT
In the last decades the application of information and communications technologies in health sector has largely increased. Clinical information systems are recognised to boost the effectiveness and quality of the services offered, and their application is changing clinical routine and affecting patient outcome in several ways. An effective evaluation of clinical information systems is needed to assure continuous improvement policies.
In the field of information and communications technologies evaluation, one of the most accredited theoretical frameworks is Goodhue's Task-Technology Fit (TTF) model. TTF considers technology as a means for a goal-directed individual to carry out a task. According to the TTF view it is not the technology itself that affects performance, but the fit between information system characteristics, task requirements, and individual skills.
The aim of this paper is to present an adaptation of Goodhue's Task-Technology Fit evaluation questionnaire to the assessment of clinicians' use of an Electronic Health Record (EHR). The questionnaire aims at examining clinicians' TTF, perceived consequences of use of the EHR, and perceived impact on performance related to mandatory EHR use.

INTRODUCTION

The application of information and communications technologies in health sector is claimed to improve access to healthcare and boost the quality and effectiveness of the services offered (European Commission, 2004). Clinical information systems are increasingly used in healthcare routine and may affect patient outcome in several ways, so their evaluation is essential (Bürkle et. al., 2001). Furthermore, effective evaluations are necessary to ensure that the information system meets the requirements and users' needs within the workflow of a health care organisation (Kushniruk and Patel, 2004).
Very often the benefit from the adoption of a clinical information system, such as an Electronic Health Record (EHR), will not manifest in direct time or money savings, but rather in improved cooperation between departments, in increased quality of documentation, or in better patient care. Due to their intangible and interconnected nature, no clear indicators have been consolidated yet to assess such effects (Bürkle et. al., 2001).
Since it is extremely difficult to get objective measures of system performance, user evaluations are frequently used as measures of information systems success. Questionnaire surveys are useful and well established tool available to researchers who wish to investigate various aspects of actual contexts, provided basic rules about design and theoretical underlying models (Mandal et al., 2000). Goodhue et. al. (2000) literature review shows that, according to many researchers, user evaluations may be considered an appropriate surrogate for information systems success, especially when system use is mandatory. User satisfaction, or acceptance, seems to reflect whether a system adequately fits the characteristics of the user and the characteristic of the task which is to be performed; thus, it can be seen as an adequate indicator whether an information system really supports users in their clinical working processes (Ammenwerth et al., 2003).
This paper aims at presenting an assessing tool aimed at examining clinicians' EHR use. The study is based on the Goodhue's Task-Technology Fit (TTF) theoretical model, which considers technology as a means for a goal-directed individual to carry out a task.
This article is organized as follows. Next section introduces the theoretical framework at the basis of the analysis. Then the research hypotheses are stated, the construct measurement and the questionnaire are presented, followed by conclusions. The entire questionnaire is reported in the appendix.

THEORETICAL FRAMEWORK

Various models in literature have identified several dimensions involved in the evaluation of user interactions with clinical information systems from the end-user point of view (Despont-Gros et al., 2005), but many of them are criticised for lacking in strong theoretical underpinnings and empirical evidence of their efficacy is contradictory (Au et al., 2002). Furthermore, many studies are presented without giving access to the test or the questionnaire, not allowing performing further investigation and comparison in similar contexts.
The adopted theoretical model, Goodhue's Task-Technology Fit (TTF), was developed in 1995 and investigated by

several researchers in various settings. TTF theory views technology as a means for a goal directed individual to carry out a task, which is defined as the actions accomplished by an individual in turning input into output.
Combining research on user attitudes as predictors of utilisation and research on task-technology adequateness as a predictor of performance, the technology-to-performance chain model's innovative element consisted in suggesting that information systems affect performance depending upon the correspondence between the task requirements of the users and the functionality of the system (Goodhue and Thompson, 1995). The TTF model presumes it is not the technology itself that affects performance, but the fit between three constructs: information system characteristics, task requirements, and individual abilities. Thus users are expected to give high evaluation based not only on the characteristics of a system but also on the extent to which that system meets their task needs and their individual abilities (Au et al., 2002).
An important hypothesis of the TTF framework is that users are aware of the costs (e.g. mental effort of training and utilization) and the benefits of a clinical information systems (impact on accuracy, speed etc.), and that they make their evaluation on that basis. As Goodhue (1995) states, if users utilise a technology because of its instrumentality in their task, it is reasonable to believe they are capable of evaluating that technology's TTF from their personal experience.

RESEARCH HYPOTHESES

In the original work from Goodhue and Thompson, Task-Technology Fit was defined as the degree to which a technology assists an individual in performing his or her portfolio of task. Part of the technology-to-performance model was tested in a controlled experiment by Goodhue itself, finding support for the link between TTF and performance impacts. Staples and Seddon (2004) performed a broader verification of the model, both in voluntary and mandatory use. Their result showed that in a context where the utilisation of the technology is mandatory, TTF is positively associated with expected consequences of use and with performance impacts.
In this research framework the relationship between user evaluations of TTF, perceived performance and perceived consequence of use are examined. In the studied organisation the use of the EHR is mandatory, so we do not perform any analysis on utilization, as it was demonstrated to be significant only when the information system use is voluntary.
The research hypotheses are stated in the followings.
The better the correspondence between the capabilities of the information system, the task and the individual, the more positive the perceived consequence towards using the system. So the first hypothesis is:
H_1) TTF is positively associated with perceived consequence of use.
When the system matches the users' needs and abilities, it should be easier to achieve the required task and since it should have a positive impact on performance, both actual and perceived. Thus we hypothesise:
H_2) TTF is positively associated with perceived performance impact.
Figure 1 illustrates the relationships to be tested. The link expressed by the white arrow was proven to be true in Goodhue (1995), while black arrows stand for the relationships we wish to reconfirm to be significant.

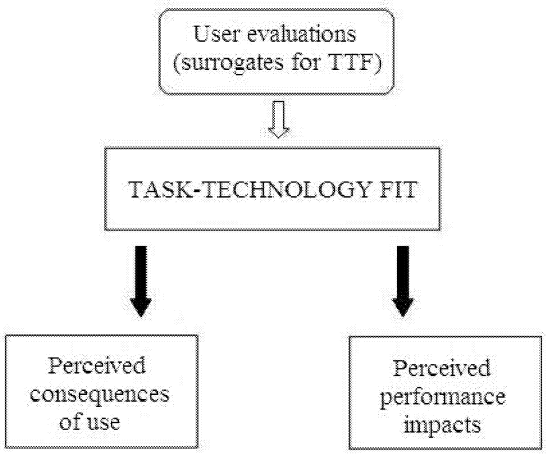


Figure 1: Research model

CONSTRUCTS MEASUREMENT

Task-Technology fit is a multi-faceted concept. User evaluation of TTF, which is proven to be a good surrogate for TFF itself when users receive feedback on it (Goodhue et al, 2000) is considered a second order construct, and it is evaluated through ten dimensions: consistency of the data in the Electronic Health Record, training for the task, help from the technical assistance personnel, system reliability, accessibility of the data in the system, meaning of data fields and elements, right and detailed data, ease of use, undestandability, familiarity with information systems and database. Each of these construct is measured with two items.
In this study the performance impact and consequences of use are based on the respondents' perception of how the system affects their work. Perceived consequences of use are measured with ten items, coming from Staples and Seddon 2004 as the subsequent set. The use of the Electronic Health Record is compared to the traditional health record in terms of quicker and easier accomplishment of the tasks, improvement of job performance, quality of care to the patient, error avoidance, control over the tasks, usefulness, and personal interest in using the system. Perceived impacts of the EHR on performance are also compared to the impact of a traditional paper record. It is measured with six items concerning efficiency, effectiveness, satisfaction with the system, cost-effectiveness, advantage vs. disadvantages.

THE QUESTIONNAIRE

Validated questionnaires to test the TTF hypothesis already exist, whose scales have demonstrated to have good psychometric properties in previous studies. The adopted item sets refer to Goodhue et al., 2000, for the TTF evaluation and to Staples and Seddon, 2004, for perceived

consequences of use and perceived impacts on performance. We choose to adopt already existing instruments and adapt them to the evaluation of the Electronic Health Record in order to enable comparison within the scientific community.
All EHR users in the organisation are planned to complete anonymously the questionnaire, consisting of 37 closed questions. A Likert scale is adopted, where 1=strong disagreement, 4=indifference, 7=strong agreement. Five open questions closed the survey, in order to get the widest feedback.
A representative sample of the different user groups is administered a pilot test, in order to assure that the questions were clearly stated and unambiguous. After the pilot sample, the questionnaire will be mailed to the whole clinician staff with a brief presentation of the project.

CONCLUSIONS

The aim of this work was to describe the adaptation of Goodhue's Task-Technology Fit IT evaluation questionnaire to the assessment of clinicians' perception towards the use of an Electronic Health Record.
The next step will consist in analysing the impact of the ten dimensions of TTF on the overall construct and the statistical relationships with perceived consequences of use and perceived performance impacts.

ACKNOLEDGEMENTS

The authors thank the Ph.D. program in "Sustainability, Quality, Safety and Logistic Management" of the University of Palermo

APPENDIX

Essential definitions

Electronic Health Record (EHR): a comprehensive medical record or similar documentation of the past and present physical and mental state of health of an individual in electronic form and providing for ready availability of these data for medical treatment and other closely related purposes. (European Commission, 2007). This technology, when fully developed, meets provider needs for real-time data access and evaluation in medical care. Together with clinical workstations and clinical data repository technologies, the EMR provides the mechanism for longitudinal data storage and access. A motivation for healthcare entities to implement this technology derives from the need for medical outcome studies, more efficient care, speedier communication among providers and management of health plans. This record may contain some, but not necessarily all, of the information that is in an individual's paper-based medical record. (Pam Pohly Associates, 2007).
Task-Technology Fit (TTF) theory: holds that IT is more likely to have a positive impact on individual performance and be used if the capabilities of the IT match the tasks that the user must perform (Goodhue and Thompson, 1995). Goodhue and Thompson (1995) developed a measure of task-technology fit that consisted initially of 8 factors: quality, locatability, authorization, compatibility, ease of use/training, production timeliness, systems reliability, and relationship with users. Goodhue and Thompson (1995) found the TTF measure, in conjunction with utilization, to be a significant predictor of user reports of improved job performance and effectiveness that was attributable to their use of the system under investigation. Since the initial work, TTF has been applied in the context of a diverse range of information systems including electronic commerce systems and combined with or used as an extension of other models related to IS outcomes such as the technology acceptance model (TAM), developed from Davis in 1993.

The questionnaire

Set 1: Task-Technology Fit (adapted from Goodhue et al., 2000)
Consistency
1. When it is necessary to compare data, there are inconsistencies that make the task more difficult than it should be.
2. Utilising the EHR is coherent with the workflow.
Training
3. I had sufficient training for tasks I usually accomplish with the EHR.
4. More training on this system would be helpful.
Help
5. The technical assistance personnel give me all the help I need.
6. It is always possible to get help when problems with EHR occur.
System reliability
7. Unplanned technical problems made tasks more difficult than they should be.
8. The EHR is subject to unexpected technical problems or crashes.
Accessibility
9. It is easy to get access to the data that I need.
10. I could get data I need quickly and easily.
Meaning
11. The exact definition of data fields is easy to find out.
12. The exact meaning of data elements is either obvious, or easy to find out.
Right data
13. The EHR contains the data I need to carry out my daily tasks.
14. Sufficiently detailed data is contained in the EHR.
Ease of use
15. It was easy to learn how to use the EHR to accomplish daily tasks.
16. The computer system and commands are convenient and easy to use.
Understandability
17. The arrangement or layout of the data is hard to understand.
18. It is easy to figure out what data is available
Familiarity
19. Before using the EHR, I already had some experience using clinical information systems
20. Before using the EHR, I was familiar with information systems.
Set 2: Perceived consequences of use (adapted from Staples and Seddon, 2004)
With respect to the paper-based record:

1. EHR enables me to accomplish my tasks more quickly
2. Using the EHR improves my job performance
3. Using the EHR enhances the quality of care provided to patients
4. Using the EHR increases the risk of medical error
5. Using the EHR makes it easier to complete my tasks
6. Using the EHR gives me greater control over my tasks
7. Overall, I do not find the EHR useful in the work I do
8. Using the EHR improves the quality of the work I do
9. The EHR is fun to use
10. The EHR is interesting to use

Set 3: Perceived performance impact (adapted from Staples and Seddon, 2004)

With respect to the paper-based record:

1. The EHR is a cost-effective solution for the organisation
2. The disadvantages of using the EHR outweigh the advantages
3. The EHR is efficient
4. The EHR is effective
5. Overall I am satisfied with the EHR
6. The EHR is worthwhile
7. I would have no difficulty telling others about the results of my use of this system.

Open questions

a. Have you ever had any problem with the EHR? If so, please describe briefly.
b. Is there any information you judge useful but that is not available in the system? If so, please describe.
c. Please express any concern regarding the use of the EHR.
d. Please give any advice about what you think would improve the EHR.

REFERENCES

Ammenwerth, E., Kaiser, F., Wilhelmy, E., Höfer, S. 2003. "Evaluation of user acceptance of information systems in health care - the value of questionnaires". *Studies in Health Technology and Informatics,* No.95, 643-648

Au, N., Ngai, E.W.T., Cheng, T.C.E. 2002. "A critical review of end-user information system satisfaction research and a new research framework". *Omega*, No.30, 451-478

Bürkle, T., Ammenwerth, E., Prokosch, H.U., Dudeck, J. 2001. "Evaluation of clinical information systems. What can be evaluated and what cannot?". *Journal of Evaluation in Clinical Practice*, No.7, 373-85

Davis, F.D. 1993. "User acceptance of information technology: System characteristics, user perceptions and behavioral impacts". *International Journal of Man-Machine Studies*, No.38, 475-487

Despont-Gros, C., Mueller, H., Lovis, C. 2005. "Evaluating user interactions with clinical information systems: A model based on human-computer interaction models". *Journal of Biomedical Informatics*, No.38, 244-255

European Commission, (2004). *"e-Health - making healthcare better for European citizens: An action plan for a European e-Health Area"*.COM(2004)356final. Brussels. Belgium

European Commission, Directorate-General Justice, Freedom and Security (2007). *"Working Document on the processing of personal data relating to health in electronic health records (EHR)"*. 00323/07/EN, WP 131. Brussels. Belgium

Goodhue, D.L. 1995. "Understanding user evaluations of information systems". *Management Science*, No.41, 1827-1844

Goodhue, D.L., Thompson, R.L. 1995. "Task-technology fit and individual performance". *MIS Quarterly*, No. 19, 213-236

Goodhue, D.L., Klein, B., March, S. 2000. "User evaluations of IS as surrogates for objective performance". *Information & Management*, No.38, 87-101

Kushniruk, A.W., Patel, V.L. 2004. "Cognitive and usability engineering methods for the evaluation of clinical information systems". *Journal of Biomedical Informatics*, No.37, 56-76

Pam Pohly Associates, 2007. *"The Glossary of Managed Care Terminology"*. Hays, Kansas, USA

Mandal, A., Eaden, J., Mayberry, M.K, Mayberry, J. 2000. "Questionnaire surveys in medical research". *Journal of Evaluation in Clinical Practice*, No.6, 395-403§

Staples, D.S., Seddon, P. 2004. "Testing the technology-to-performance chain model". *Journal of Organizational and End User Computing*, No.16, 17-36

Zigurs, I., Buckland, B.K., 1998 . "A theory of task/technology fit and group support systems effectiveness". MIS Quarterly, No. 22, 313-334.

BIOGRAPHY

ELENA CIANCIMINO was born in 1980. She graduated from the University of Palermo in 2005 with a First class Honours Degree with distinction in Management Engineering. She has been visiting student at Edena (Buenos Aires), a member of WorldWide Classroom Consortium for International Education & Multicultural Studies. Between 2005 and 2006, she was awarded a grant to perform an internship within the EU Leonardo Da Vinci mobility Programme. Since March 2006 she has been an engineering fellow at ISMETT (Mediterranean Institute for Transplantation and Advanced Specialized Therapies), Tissue & Stem Cells Laboratory project. Her research interests include supply chain modelling, robust design and assessment techniques. Current position: Ph.D. student at Department of Manufacturing, Industrial and Management Engineering of the University of Palermo, and visiting at Department of Industrial Management, High School of Engineering, University of Seville, Spain.

SALVATORE CANNELLA was born 1980 in Marianopoli, Italy. He holds a master's degree in Management Engineering (summa cum laude) from the University of Palermo, 2005. Since 2004 he has been committed to several international cooperation research projects, sponsored by the EU Socrates II programme, Spanish Ministry of Education and Science (MEC) and the Italian Ministry of University and Research (MIUR). He is member of Production Engineering Research workGroup (PERGROUP). His research is focusing on bullwhip effect reducing approaches, replenishment rules, IT-enabled supply chain issue, system dynamics methodology and what-if simulation in operations management. In 2006 he performed a training stage at Embassy of Italy in Panama under the provisions set forth by agreement between the Italian Ministry of Foreign Affairs and the Council of Italian Universities (CRUI/MAE internship). He is now Ph.D. student at Department of Manufacturing, Industrial and Management Engineering of the University of Palermo, and visiting at Department of Industrial Management, High School of Engineering, University of Seville, Spain.

PERFORMANCE ESTIMATION AND FORECASTING

MPSOC DESIGN FLOW BASED ON PERFORMANCE ESTIMATION

Kamel SMIRI[1, 3] Abderrazak JEMAI[1, 4] Nihel CHIBANI[1, 2] Mohamed MOALLA[1, 3]

[1]LIP2 Laboratory, Faculty of Science of Tunis, Tunisia
[2]School of Polytechnic of Tunisia, *B.P. 743 - 2078, Marsa, Tunisia*
[3]Faculty of Science of Tunis, Tunisia, *Campus University, 2092 Manar 2, Tunisia*
[4]INSAT, *B.P. 676, 1080 Tunis Cedex, Tunisia*
smiri_kamel@yahoo.fr abderrazek.jemai@insat.rnu.tn c_nihel@yahoo.com mohamed.moalla@fst.rnu.tn

KEYWORDS
MPSoC, Design method, Design Flow, multilevel methodology, Motion-JPEG Decoder, Performance Model.

ABSTRACT
In recent years, computational requirements for embedded applications are increasing exponentially. This complexity, coupled with constantly evolving specifications, has forced designers to consider intrinsically flexible implementations. For this reason, heterogeneous Multiprocessors Systems-on-Chip architecture (MPSoC) are gaining prevalence for embedded systems. More laboratories propose a platform based software design flow able to efficiently use the resources of the architecture and allowing easy experimentation of several mappings of the application onto the platform resources. Performance estimation plays a decisive role inside the design space exploration for ranking the design alternatives. The goal of this paper is to describe our experience on tools and environments of evaluation performance along the MPSoC design flow. We proposed performance model witch based on the annotation of the model Khan Process Networks of (KPN). The performance model evaluates the process execution time and the size of the exchanged data between processes. This performance model to aim the facilitate task of designers and guaranty the passage between their abstraction level. The effectiveness of this model was illustrated on the Motion-JPEG decoder application.

INTRODUCTION

According to Moore's law, more and more transistors will be integrated on a single chip. Such a huge transistor budget makes it increasingly difficult for engineers to design and verify the very complex chips that result, and in turn widens the gap between silicon capacity and design productivity. Multiprocessors Systems-on-Chip (MPSoC) architecture has thus become a solution for designing embedded systems dedicated to applications that require intensive computations (Pimentel, 2001) (Jerraya, 2006).

MPSoC are specialized to carry out specific tasks and are "embedded" in their environment. This is in contrast to personal computers or supercomputers which are general purpose and interact with users. Embedded systems typically have strict performance requirements relating to issues such as latency, throughput, jitter, memory usage, and energy consumption. Due to their widespread usage and performance-critical nature, the design of embedded systems is both relevant and challenging (Davare, 2007).

The choice of implementation architecture determines whether designers will implement a function as a hardware component or as software running on a programmable component. In recent years, computational requirements for embedded applications are increasing exponentially. This complexity, coupled with constantly evolving specifications, has forced designers to consider intrinsically flexible implementations. For this reason, and because hardware manufacturing cycles are more expensive, software based implementation has become more popular (Caspi, 2005) (Bouchima, 2006) (Popovici, 2007) (Guerin, 2007).

This paper is divided into four parts. The part one presents the software architecture and we presented the HW/SW architectures design flow for MPSoC proposed by TIMA-SLS laboratory. The part tow describes our experience on tools and environments of evaluation performance along the MPSoC design flow. The part tree presents our experience on the multimedia application Motion-JEG decoder. In the fourth part we proposed performance model witch based on the annotation of the model Khan Process Networks of (KPN).

SOFTWARE ARCHITECTURE

Given the increasing complexity of MPSoC applications, developing the SW monolithically is likely to be an unmanageable task. Indeed, this implies the use of low level language (Assembler) that supposes a meticulous knowledge of the underlying hardware architecture. First, this will cause SW development and debug to be harder. Furthermore, developing the SW as a whole does not allow SW reuse and portability. Indeed, this will cause the dependency of the application on targeted architecture. To tackle this problem, the use of an Operating System (OS) has been essential to decouple application specific computation from the manner in which this computation is performed on the underlying hardware architecture. In this way, a software layered approach is adopted to enable application software to run on different MPSoC architecture according to the software architecture. From the viewpoint of application software, the software architecture provides a virtual architecture on which the application software runs. The software virtual architecture includes: the Application layer; the Middleware; the Operating System (OS) and the Hardware Abstraction Layer (HAL). Each layer provides Application Programming Interfaces (APIs) for the layer above it and hence ensures an abstraction of the underlying hardware architecture to upper layers of software. Those APIs define a minimal standard set of common commands to make applications more portable.

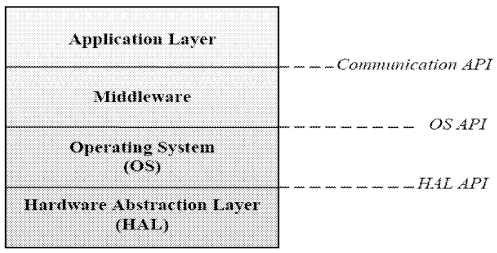


Figure 1: Software abstraction layers

As such, the software architecture enables several levels of software design reuse. First, the middleware API or OS API enables reuse of application software. Case when application software is written in an OS API, e.g., Portable Operating System Interface X or POSIX API, the same application software can be reused in other MPSoC designs in which the same OS API is provided (Bouchima, 2006). The same applies to the HAL API. If the same HAL API is used in OS and middleware code, OS and middleware can be reused over different designs (Jerraya, 2006).

MULTILEVEL METHODOLOGY FOR MPSOC DESIGN

Consisting of several types of processors from fully programmable microprocessors to configurable processing cores and customized hardware components, integrated on a single chip. The sheer architectural complexity of MPSoC based embedded systems, as well as their conflicting design requirements regarding good performance, high flexibility, low power consumption and cost greatly complicates the system design. It is now widely believed that traditional design methods come short for designing these systems due to following reasons (Pimentel, 2001):

- Classical design methods typically start from a single application specification, making them inflexible for broader exercise. In fact, MPSoC designs require very fast simulation, crossing the hardware/ software boundaries and far exceeding today's capabilities. Today software on an Instruction Set Simulator (ISS) for single processors can be simulated at multiple MIPS. Those capabilities are simply not scaling appropriately when using 10, 20 or even 50 processors and trying to meet the requirement of running a significant portion of the application (think 30 seconds of video and audio in a multimedia application). Today's embedded debug solutions are built with single processors in mind, not taking into account the requirements of debugging parallelism in an application distributed to 10, 20 or 50 processors.

- Compilers are targeted to single processors only. The problems of expressing parallelism in application programming, then mapping and distributing software on an MPSoC platform remain largely unaddressed.

- High level programming does not handle efficiently specific I/O and communication schemes, while low level programming explicitly managing specific I/O and communication is time consuming and error prone activity.

- Common evaluation practices still makes use of detailed cycle-accurate simulators for early design space exploration. Building these detailed simulation models requires significant effort, making them impractical in the early design stages. What is more, these low level simulators suffer from low simulation speeds which hinder fast exploration.

In order to handle those problems, the ever increasing complexity and time-to-market pressures in the design of multi-processors system-on-chips, the design abstraction has been raised to the system level to increase productivity. System level design enables making design decisions at higher levels of abstraction and reusing design components. Recently, there has been interest in transaction-level modeling TLM for system abstraction.

1. The HW/SW architectures for MPSoC design flow

TIMA-SLS of laboratory approach is derived from a Platform based Design method (Sangiovanni-V, 2004). Platform based Design focuses on the creation of abstraction layers in the design flow and investigates the semantic properties of mapping across these layers. It lies on a structured methodology for developing economically feasible software design flow. The proposed approach starts with a high level model of the application in allowing fast simulation of the application model. It generates automatically the software code and the corresponding software development platform. The abstract models of the architecture capture the special features of the final hardware platform allowing accurate estimation of performances. This approach provides incremental generation and validation of the different components of the software stack. This makes easier the debugging of both application and HdS in addition to opening new opportunities like communication mapping exploration (Popovici, 2007) (Guerin, 2007).

1.1. System Architecture level (SA)

It is the highest level. At this level, the software stacks have not been created yet. In this case, the software is made of a set of functions grouped into tasks (figure 2). The function is an abstract view of the behavior of an aspect of the application. Several tasks may be mapped on the same software subsystem. The validation is performed by the Simulink application. The initial application model is simulated using the Matlab's discrete simulation engine. The simulation at this level allows validation of the application's functionality. In fact, here, the global application's functionality can be verified and the existing bugs are tracked down and corrected.

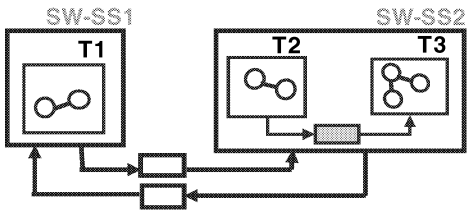


Figure 2: High level combined architecture /application model

1.2. Virtual Architecture level (VA)

At this level, the software stacks are only composed of the tasks and the communications are performed via SystemC channels. The validation is performed using a SystemC simulation mode. The communication primitives of the HdS API access explicit communication components. Each data transfer specifies an end-to-end communication path. The software is executed using an abstract model of the hardware architecture that provides an emulation of the HdS API. The

software development platform is composed of these abstract subsystems, explicit interconnection component and storage resources (figure 3). The simulation at this level allows validation of the final code of tasks and may give useful statistics about the communication requirements. The application's partitioning balance can be verified by analyzing the amount of data exchanged between the tasks.

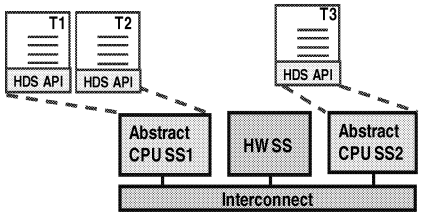


Figure 3: Virtual Architecture abstraction layer

1.3. Transaction accurate level (TA)

At this level, the software is linked with an explicit OS and specific I/O software to implement the communication units. The resulting software makes use of hardware abstraction layer primitives (HAL_API). The software is executed using a more detailed development platform to emulate the network component, the explicit peripherals used by the HAL API and an abstract computation model of the processor (figure 4). The validation is performed using a SystemC model as well. Here, the operating systems and the communication schemes can be validated by analyzing the number of bytes going through the hardware communication channels and the application tasks' response time. It may also provide precise information about the communication performances.

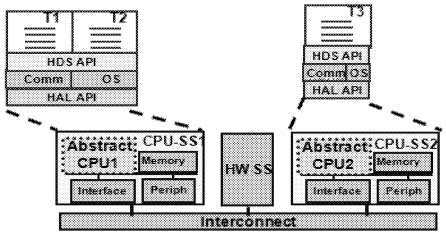


Figure 4: Transaction Accurate abstraction layer

1.4. Virtual prototype level (VP)

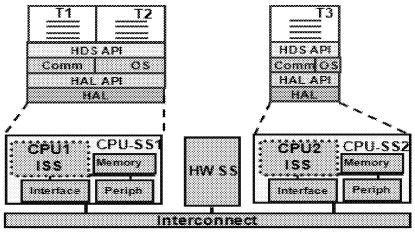


Figure 5: Virtual prototype abstraction layer

At this level, the software stacks are complete and the communications fully implemented. In fact, the HAL API and processor are implemented through the use of a HAL software layer and the corresponding processor part for each software subsystem. The communication consists of physical I/Os, e.g. load/store (figure 5). The validation is also performed using a SystemC model.

Here, the complete software stacks can be verified with the real device drivers and the final hardware platform's HAL. Even if this validation is the most accurate, it is also the slowest. It allows performance validation and it corresponds to classical hardware/software co simulation models with Instruction Set Simulators for the processors and RTL (Register Transfer Level) components for the hardware resources.

We notice that the MPSoC co-design flow proposed by the SLS group TIMA laboratory is very methodological and easy to exploit it. Because this flow is based on the refinement and the increment of the complexity details in the lowest abstractions level. And also use limited of the languages by the designers of the material part and the software part. The used languages can be limited to the SystemC, VHDL and C/C++. The problem of the MPSoC co-design flow is mentioned in System Architecture level (SA). We note that this level although it is very important for the validation functional. But the choice of Simulink\Matlab is advised against. Because one obliges the designer to master a new language, where complicates the task.

We propose in this context to respect the same steps of the flow. Just has replaced the Simulink\Matlab environment by the Disydent framework. Disydent is has framework dedicated to multiprocessor system one chip (MPSoC) platform-based design heart shared memory Multiple Instructions Multiple Dated (MIMD) architecture (ASIM-LIP6, 2007). Because Disydent Permit the validation of the Parallel model of the application that is described in C/C++ and according to Khan Process Network (KPN).

2. Mutek Architecture

Mutek is an open source that implements a POSIX compliant multiprocessor Kernel (ASIM-LIP6, 2007). Its architecture is depicted in Figure 6.

Scheduler organisations

The scheduler manages several lists of threads. It may be shared by all processors (SMP Symmetrical Multi-Processor). In the SMP scheduler organisation, there is a unique scheduler shared by all processors and protected by a lock. The different treads can run on any processor, with leads to task migration.

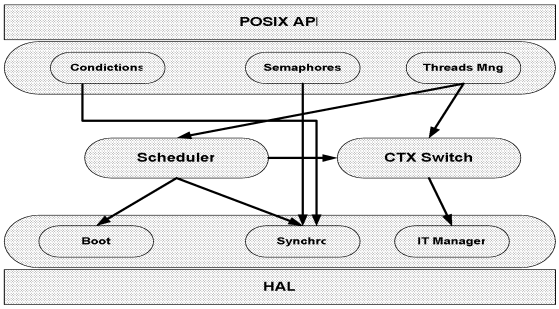


Figure 6: Mutek Architecture

Kernel protection and thread migration

The access to the scheduler must be performed through critical section, and under the protection of locks. Lock granularity is one major player in determining the balance between the overhead introduced by the locking mechanism and the opportunity to increase parallelism among different processors.

The SMP version of Mutek allows tread migration. The term migration is related to the ability of an operating system to resume a job on another processor after pre-emption.

Intuitively when a CPU finishes the threads currently allocated on it for scheduling, it can resume the execution of pre-empted thread that was previously executed on another processor. In that way, the system is dynamically balanced, reducing the mean response time of the system.

Thread synchronization

Synchronisation is required whenever shared data need to be accessed. In Mutek, this is done using different primitives: Spin-Lock, Mutual exclusion locks, condition variables and semaphores. Spin-Lock is the lowest level Mutual exclusion object. Thread waiting for Spin-Lock continually accesses element and do not return until they grad it. It is an active wait element (Senouci,2007).

A Mutual exclusion lock (Mutex) allows excusive access to shared resources such as global data. Threads attempting to access an object locked by a Mutex will be blocked until the thread lording the object releases it. Using condition variables, a thread can wait until (or indicate that) a predicate becomes true. A condition variable requires a Mutex to protect the data associated with the predicate.

Processor identification number

Processors generally provide a specialized register allowing their identification within the system. Each processor is assigned a number at boot time. This identification number is needed, for instance, at boot time, since some start-up actions should be done only once, such as clearing the blank static Storage (BSS) and creating the scheduler.

PERFORMANCE ANALYSIS

1. Role of performance evaluation in the Design Process

One of the major challenges in the design process of embedded systems is to estimate essential characteristics of the final implementation early in the design (Jerraya, 2006). This can help in making important design decisions before investing too much time in detailed implementations. Typically, the performance analysis or estimation is part of the design space exploration, where different implementation choices are investigated in order to determine the appropriate design trade-offs between the different conflicting objectives. The estimation of system properties in an early design phase is an essential part of the design space exploration (figure 6). Different choices of the underlying system architecture, the mapping of the applications onto this architecture, and the chosen scheduling and arbitration schemes will need to be evaluated in terms of the different quality criteria. In order to achieve acceptable design times though, there is a need for automatic or semi-automatic (interactive) exploration methods. As a result, there are additional requirements for performance analysis if used for design space exploration, namely (a) the simple reconfigurability with respect to architecture, mapping and resource sharing policies, (b) a short analysis time in order to be able to test many different choices in a reasonable time frame and (c) the possibility to cope with incomplete design information, as typically the lower layers are not designed or implemented.

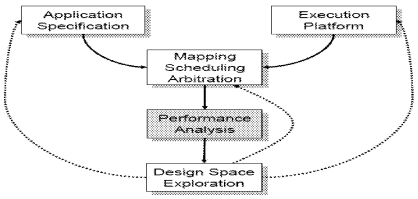


Figure 7: Relation between design space exploration and performance analysis.

2. Approaches to Performance Analysis

2.1. Simulation technique

These techniques are based on an evaluation of the system model by means of simulation. For this purpose, different levels of abstraction in the modeling of the components and their communication are used (Dey, 1997). The level of abstraction for modeling the communication allows for a trade-off between simulation time and accuracy. Examples of available approaches and software support provide the SystemC initiative that is supported by tools from companies like Cadence (nc-systemc) and Synopsys (System Studio). In simulation-based methods, many dynamic and complex interactions can be taken into account whereas analytic methods usually have to stick to a restrictive underlying model and suffer from limited scope. In order to determine timing properties of an embedded system, a simulation framework not only has to consider the functional behavior but also requires a concept of time and a way of taking into account properties of the execution platform, of the mapping between functional computation and communication processes and elements of the underlying hardware, and of resource sharing policies (as usually implemented in the operating system or directly in hardware). This additional complexity leads to higher computation times, and performance estimation quickly becomes a bottleneck in the design. Besides, there is a substantial set-up effort necessary if the mapping of the application to the underlying hardware platform changes, for example in order to perform a design space exploration. The fundamental problem of simulation-based approaches to performance estimation is the insufficient corner case coverage.

2.2. Trace-Based Performance Analysis Strategies

These are generally based on trace transformation techniques. A trace is a track of the system execution that contains information concerning computations and communications of the system components (Cagkan, 2006). These methods take as input a trace generated by an application process and generate, as output, a trace accepted by an architecture model and which contains the architecture level operations. Thus, a trace transformation provides the mapping of application level communication primitives onto architecture level communication primitives. This basic idea of collecting an execution trace and using it for performance estimation has been used in the field of high performance processor design, for instance for cache simulation. Traces are represented as a Communication Analysis Graph (CAG), whose analysis provides an estimation of the system performance as well as various statistics about the components and their communication. An important feature of this approach is its ability to model various dynamic effects of the communication architecture and to consider the inter-

dependencies between the computations, synchronizations, and data communications performed by the various components while estimating the system performance.

2.3. Static Performance Estimation approaches

These techniques try to avoid the computationally prohibitive alternative of exhaustive simulation, arguing that an efficient exploration of the system design space necessitates fast performance estimation. They include models of the communication time between components of the system and often assume systems where the computations and communications can be statically scheduled. The communication time estimations used in these systems are either overly optimistic, since they ignore dynamic effects such as wait time due to bus contention, or are overly pessimistic by assuming a worst case scenario for bus contention.

2.4. Analytical Performance Estimation approaches

The analytical approach formally investigates system capabilities (Jerraya, 2006). The subsystem under analysis is generally described at a high level of abstraction by means of algebraic equations. Mathematical theories applied to performance evaluation make possible a complete analysis of the full system performance at an early design stage. Moreover, such approaches provide fast evaluation because they replace time-consuming system compilation and execution. Building an analytical model could be Performance very complex. The dynamic behavior (e.g., program context and wait times due to contentions or collisions) and refinement steps (e.g., compiler optimizations) are hard to model. However, this approach may be useful for worst-case analysis or to find corner cases that are hard to cover with simulation.

We can quote two examples of these methods. The first method is Rate Monotonic Analysis (RMA). This method is a whole of technique mathematics which allows the developer's systems real time to answer real requirements time of application. The second method is Response Time Analysis (RTA). This last one is an analytical method which can be employed before the integration of the system to determine if real the time needs are satisfied.

RMA Method is implemented by two tools supported in the market with knowing: Timewiz and RapidRMA. Timewiz is a modelling tool of design operating under Windows OS, the simulation analysis of the performances which east, intends for the systems real time. It delivers results which can give information on the profitability of the necessary equipment and the software used. Timewiz makes it possible to parameterize the resources.

2.5. Profiling techniques

The profiling of a program make possible to identify the places where this one spends the most time, but also which are the functions which are carried out and how much time. All the tools for profiling function on the same diagram: they collect a certain number of information during the execution of a program. This information is dependent on the context of execution and a comparison could be carried out between two profiles only with equal context. Another consequence is that the profile could be used only after the end of the execution of the program.

Gprof: The Gprof commanded produces a report of execution for the program. It produces information on the graphs of calls and calculates the time spent for each routine. Times are then propagated along the graph of call. *Gprof* requires the recompiling and the binding of the application (Fabrice, 2007). To use the order *Gprof* it is necessary to compile the program with the profiling option *"- pg"*. At the end of the execution, we obtain a data file, *gmon.out*, which contains the information measured in internal format that *Gprof* can analyze.

Kprof is a utility witch ensures profiling. It is able to analyze ergonomically the C++ objects. Outputs and results of the profilers seem difficult to be read and to be analyzed whereas *Kprof* present them in forms of list or tree what makes comprehension easier in order of more effective optimization (Popovici, 2007).

Callgrind use the execution time technology calculation execution time employed by *Valgrind* to make necessary simulation and to generate the trees of calls. The files generated by *Callgrind* can be given the responsibility by *KCachegrind* to provide the performance results (Blume, 2006) (VALGRING, 2007).

MemProf It is a tool for shaping for the use of the memory. These two major characteristics are: (1) it makes it possible to give precise information concerning the allowance of the memory for each function of the program and (2) it can scanner the report and find the allocated blocks which are not well referred (MEMPROF, 2007).

We used different tools and method of evaluation performance in the context of the MPSoC Co-design of abstractions levels. We note that the profiling techniques are counselled for of abstractions levels. Concerning the simulation technique is always advice to the different levels of abstraction. On the other hand the use of the analytic methods is advised against to lowest abstraction levels.

3. Comparison of approaches for system performance estimation

The analytical approaches model a simplified system and can be evaluated analytically or by means of simulation. The analytical evaluation of analytical models aims at minimizing modeling time and solution time, while providing sufficient prediction accuracy to reduce the design space. They are suitable for the initial design stages where prediction cost is of a higher priority than optimum prediction accuracy, given the huge design space involved. The static performance estimation methods belong to this group. In the case that the system is too complex or there is a need for detailed modeling, the simulation is one of the most suitable alternatives to evaluate such systems.

The simulation-based approaches use either simulators for instruction level, cycle-accurate or RTL models. The simulators for instruction level models are not able to handle multi-processor systems due to the fact that timing details are only available at instruction boundaries. In contrast, the simulators for logic level models make it a very time consuming task to run software on it. Therefore, the most

commonly used approach is the cycle-accurate simulation, where many logic level details are omitted to speed-up the simulation. Nevertheless, the cycle-accurate simulation is also a time consuming approach.

Lastly, the trace-based techniques apply both analytical and simulation-based methods. They extract from the initial specification the information about the computations and communications of the system components. With this information and the parameters of the system resources, such techniques deliver an estimation of the final performance. Trace-based techniques are suitable for cases where the initial co-simulation is not easily reproducible. The extraction of traces from the initial co-simulation has direct influence on the inspection of diverse architecture combinations. Nevertheless, the traces to be extracted are fixed in advance as well as their sequence. A change during the evaluation phase is not possible.

EXPERIMENTATION: MOTION JPEG DECODER

The MJPEG decoder reads a stream of JPEG images from an input peripheral: a traffic generator named TG and writes pixels on an output peripheral: a digital-to-analogue converter named RAMDAC. The first functional bloc DEMUX dispatches the input stream to the other blocs. The decoding chain (fig.8) begins with the decompression bloc (VLD), then the zigzag rearrangement bloc (ZZ), followed by the inverse quantification bloc (IQ). Finally, the stream passes to the inverse discrete cosine transformation (IDCT).

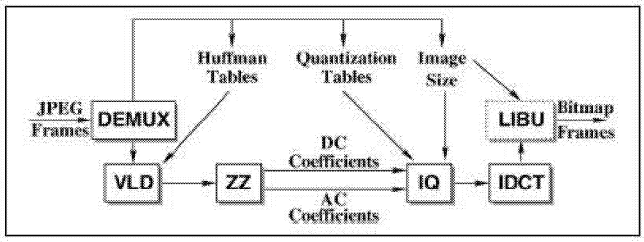


Figure 8: Motion-JPEG decoding principe.

We applied the three methods of performance evaluation on the Motion-JPG decoder: (1) Simulation techniques, (2) analytic methods (3) and Profiling techniques.

1. Simulation Techniques

We have using the Disydent design flow to validate the Motion-JPG decoder. Disydent is a framework dedicated to multiprocessor system on chip (MPSoC) platform-based design for shared memory Multiple Instructions Multiple Data (MIMD) architecture (ASIM-LIP6, 2007). Disydent is an open framework developed by the ASIM department of the LIP6 laboratory.

An extended MPSoC system includes both hardware and an operating system. This system is based on an “Asim0” platform. Asim0 contains n >=1 MIPS processors, a Pi-Bus interconnexion bus, a FIFO communication interface, an interruption controller, direct memory access controllers. The operating system «Mutek» is a multiprocessor multithread kernel supporting POSIX APIs. It proposes FIFO based communication layer allowing KPN communications. It includes three different schedulers. In the Symmetric Multiprocessor (SMP) kernel, there is only one scheduler allowing tasks to migrate between processors. The Asymmetric centralized Scheduler (Non SMP-CS) is only one scheduler affecting statically tasks to processors. The Asymmetric distributed scheduler (Non SMP-DS) instantiates one scheduler to every processor and tasks are statically affected to processors.

The sequential execution of the decoder on a multiprocessor platform realises 12.2 fps. Then, we need to duplicate the decoding performance by 2.6. The second step of the design flow is the parallel implementation. We have used three parallel models (fig 9).

Software implementation of the KPN model (fig. 4.a) on the asim0 platform with 1, 2, 3 and 4 MIPS processors, using the SMP kernel, had sped up the decoding process. But the execution time stagnates with the fourth processors at the value of 13.86 fps. We have been capable of evaluating the complexity of each task by using profiling techniques. We have found that the most important task is IDCT, with 72% of the execution time. Then we find the VLD task with 11.3% of the execution time.

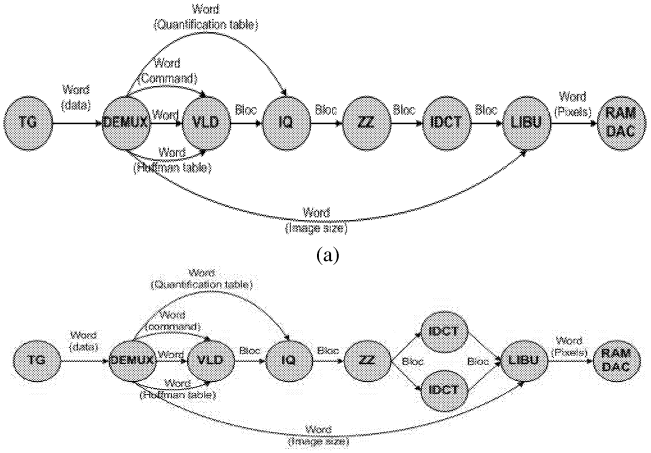
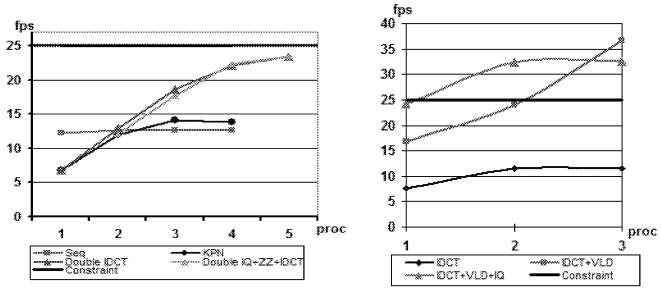


Fig. 9: (a) KPN model of Motion JPEG decode, (b) parallelization solution.

The model described in fig. 9.b is parallelized at bloc level, by duplicating IQ, ZZ and IDCT tasks. Using the asim0 platform with 5 processors and the SMP kernel of the «Mutek» (dynamic scheduler), this model reaches 23.35 fps. This implementation model presents an inconvenient: important communication time between processors. The double IDCT model (fig. 9.c) has accelerated the decompression time. The execution time reaches 23.3 fps with 5 processors on the platform.

We start by selecting processes candidates for a hardware implementation. Using tasks complexity profile and using the software implementation results, we have selected 3 groups: (IDCT), (IDCT+VLD) and (IDCT+VLD+IQ). Hardware implementation of group (IDCT) has realized 11.55 fps with 3 processors on the platform. So, we need a speedup of 2.5 in order to verify constraint. By mapping group (IDCT+VLD) into hardware, the constraint was almost verified using 2 processors on the platform (24.07 fps). In this case, a code optimization and a suitable configuration of the platform are necessary to reach 25 fps. The (IDCT+VLD+IQ) group had given as execution performance 24.12 fps with only one processor on the platform (fig. 10.b).



(a) (b)

Fig. 10. (a) Performance of the software implementation, (b) Performance of the hardware implementation of process groups.

2. Profiling techniques

The figure 11 presents us results of profiling technique while using the Callgrind tool for three different input files: F1.jpg, Plan.jpg and Unicycle.jpg. These results are obtained for a sequential specification for the Motion-JPEG decoder.

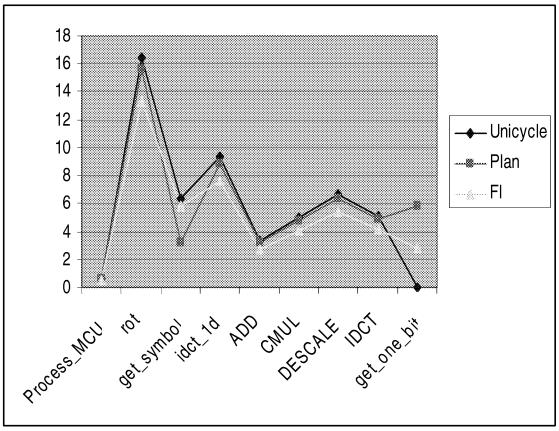


Fig 11. Results of profiling with Callgrind

We note that the execution time of the functions (e.g. IDCT, Rot, ADD, DESCALE,…) depends of picture type to decode.

3. Analytic methods of performance estimation

Sequential Implementation of MJPEG decoder

We Implement the MJPEG code on one node alone: CPU, a BUS, and a memory (Tab 1).

Tab 1: Assignment of the tasks with RapidRMA.

P	Name	Period	Work	...	Rel Dl	.....	Priority	Network	Active Res.	Dl Type
1	TG	Det: (50000)	Det: (1)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard
2	DMUX	Det: (50000)	Det: (596)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard
3	VLD	Det: (50000)	Det: (6868)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard
4	IQ	Det: (50000)	Det: (4708)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard
5	ZZ	Det: (50000)	Det: (4856)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard
6	IDCT	Det: (50000)	Det: (17144)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard
7	LIBU	Det: (50000)	Det: (4076)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard
8	RAMDAC	Det: (50000)	Det: (1)	0	50000	0 0	0	0	Node_0.CPU (R3)	Hard

RMA Approach proposes an assignment of the tasks for the priorities which will be at the base of the pipelines algorithms of scheduling. We specify following profiling time necessary for the execution of each tasks, the period of the CPU - here 500000- , as well as the node on which the task will be implemented. We note that the calculation of the task's work time is given as follow: Work (1frame) = 1/25 * %. Example: work of the task Demux= 1/25* 1, 5% = 0.0596.

Parallel implementation of MJPEG decoder

In the parallel implementation We choose to put into second the spot IDCT considering which it is greedy in resource, thus we will use two to see 3 CPU, and consecutively 2 to 3 memory and bus.

In the case of an only node: Sequential MJPEG. RapidRMA sequentially draws up the scheduling of the tasks according to their set of priorities. Optimal time of execution on only one node = 38.249. Less occupation of the CPU 1, better execution time Summary of the results 4= 17.144 marked in green.

We implement through worms these results, the improvement of the execution time in costs of the increase in the number of node. It result is perceptible of advantage with approach RMA if one puts in analysis the aspect of the priorities of the tasks as well as the type of this analysis, point-to-point or multi-task.

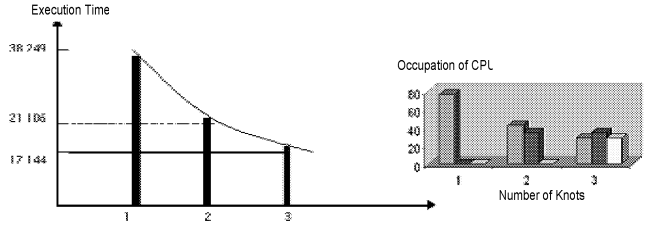


Figure 12: (a) Execution Time, (b) CPU Occupation

PROPOSAL OF A PERFORMANCE MODEL

We propose to add to the different abstraction levels considered co-design flow a temporal performance model. It consists in extending the Kahn process network with process execution time and the amount of exchanged data between processes. Figure 13 presents the proposed model for the MJPEG decoder. The annotations which describe the temporal performance are obtained by profiling techniques, analytical methode and simulation technique. The tools for profiling used are Kprof, Gprof, Callgrind, Memprof. The tool for simulation used is CASS of Disydent and SytemC language. The tool for analytical methode used is Timewiz and RapidRMA. The model of performance can be refined on the various levels of abstractions. The objective of this model is to help the designer to quickly reach the optimal solution and to guarantee it correctly function of application.

One envisages implementing and integrating our performance model in the MPSoC design flow. The performance model proposed can take a representation varieties and a structure variety along the MPSoC design flow. The limit of our model is that it will only be applied in the MPSoC design approaches platform based design.

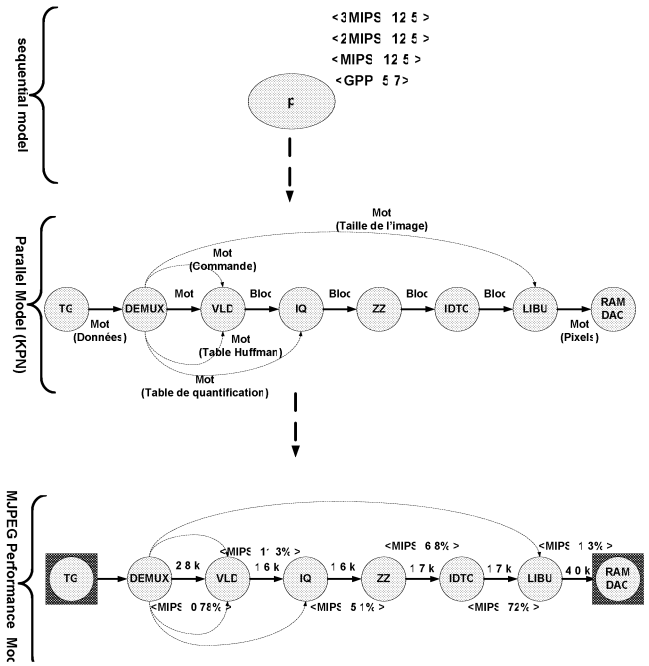


Figure 14: model of performance proposed.

CONCLUSION

We presented in this paper four parts. The first summarizes a bibliographic synthesis on MPSoC design flow existing. We presented a platform based software design flow allowing efficient software architecture and validation for material's architectures including heterogeneous MPSoC with specific I/O. In particular, we are interested to laboratory SLS-TIMA flow (INPG, Grenoble, France). The second part is reserved to the presentation of our experience on tools and environments of evaluation performance along the MPSoC design flow. We are interested to the three methods of evaluation performance: (1) the simulation technique, (2) the profiling techniques and (3) the analytic method. We proposed a gait to exploit these methods along SLS-TIMA MPSoC design flow. In the third part we present our experience on the multimedia Motion-JPEG decoder application. In the fourth part we proposed performance model witch based on the annotation of the model Khan Process Networks of (KPN). The performance model evaluates the process execution time and the size of the exchanged data between processes. This performance model to aim the facilitate task of designers and guaranty the passage between their abstraction level. The effectiveness of this model was illustrated on the Motion-JPEG decoder application.

REFERENCES

Bouchima, A. 2006. *"Modélisation du logiciel embarqué à différents niveaux d'abstraction en vue de la validation et la synthèse des systèmes monopuces".* Thesis, TIMA Laboratory.

Blume, H., Von-Sydow, T., All. 200|6. *"Application of deterministic and stochastic Petri-nets for performance modeling of NoC architecture".* J. Syst. Architect 2006.

Caspi P., Sangiovanni-Vincentelli A., All. 2005. *"Guidelines for a Graduate Curriculum on Embedded Software and Systems".* ACM Trans. On Embedded Computing Systems, Vol. 4, n. 3, pp. 587-611, Aug. 2005.

Cagkan, E., 2006. *"System-Level Modeling and Design Space Exploration for Multiprocessor Embedded System-on-Chip Architectures ".* Vossiuspers UvA. Amsterdam University.

Davare A., Densmore D., All. 2007. *"A Next-Generation Framework for Platform-based Design".* Design and Verification Conference (DVCON), San Jose, CA, February 21-23, 2007.

Dey, S., Bommu, S., 1997. *"Performance Analysis of a System of Communicating Processes".* Proc. IEEE/ACM International Conference on Computer Aided Design (ICCAD 97), San Jose, USA.

Fabrice, B., 2007. *"Profiling de programmes C et C++ avec GNU gprof".http://kcachegrind.sourceforge.net/cgi-in/show.cgi/ KcacheGrindIndex.*

Guerin, X., Popovici, K., All. 2007. *" Flexible Application Software Generation for Heterogeneous Multi-Processor System-on-Chip ".* DATE 2007.

Jerraya, A., 2006. *"Performance Evaluation Methods for Multiprocessor Systems-on-Chip Design".* Chapitre 6, TIMA Laboratory.

Knudsen, P.V., Madsen, J. 1999. *"Graph-Based Communication Analysis for Hard- ware/Software Codesign".* Proc. 7th International Workshop on HW/SW Codesign, Rome, Italy.

Pimentel, A. D., Lieverse, P., All. 2001. *" Exploring embedded-systems architectures with Artemis. Computer".* TIMA Laboratory.

Popovici, K., Rousseau, F. 2007. *" Multilevel Programming Environment for Multimedia Heterogeneous MPSoC Platform".* DATE 2006.

Popovici, K., Guerin, X., All. 2007. *"Efficient Software Development Platforms for Multimedia Applications at Different Abstraction Levels".* Rapid System Prototyping Workshop 2007, Brazil.

Sangiovanni-Vincetelli A., All, 2004. *" Benefits and Challenges for Platform-Based Design".* Proceeding of DAC 2004, USA.

Thompson, M., Pimentel, A. D., All. 2006. *"A Mixed-level Co-simulation Technique for System-level Design Space Exploration".* IEEE Workshop on Embedded Systems for Real-Time Multimedia, Seoul, Korea.

Paul, J. M., Thomas, D. E., All. 2005. *"High-level modeling and simulation of single-chip programmable heterogeneous multiprocessors".* ACM Transactions on Design Automation of Embedded Systems.

Smiri, K., Jemai, A. 2007. *"Performance Evaluation on Embedded Systems: Analytical methods and Profiling techniques".* Annual TMSS Conference, Hammamet, Tunisia.

Senouci, B., Kouadri A., Rousseau, F., Petrot, F. 2007. *"Hardware Spin-Lock Engine for POSIX Compilant SMP Operatin".* TIMA Laboratory.

ASIM-LIP6., 2007. www-asim.lip6.fr/recherché.disydent.

VALGRING.,2007.www.valgrind.org/docs/manual/cl-manual.html.

MEMPROF., 2007.www.gnome.org/projects/memprof/: memprof.

Kamel SMIRI received an Engineer degree from the The National Engineers School of Monastir, Tunisia in 2001 and the Master degrees from the University of MANAR II, Faculty of Science of Tunis, Tunisia, in 2005. From 2005 he prepared his thesis on Performance evaluation and MPSOC design flow based of performance estimation.

Abderrazak JEMAI received an Engineer degree from the University of Tunis, Tunisia in 1988 and the DEA and "Doctor" degrees from the University of Grenoble, France, in 1989 and 1992, respectively, all in computer sciences. From 1989 to 1992 he prepared his thesis on simulation of RISC processors and parallel architectures. Dr Jemai became an assistant professor at the ENSI university in Tunis in 1993 and a Maitre-Assistant Professor at the INSAT university in Tunis since 1994. He was the principal investigator for the "Synthesis and Simulation of VLSI circuits" project at the ENSI/Microelectronic group. He is the principal investigator for the "Performance evaluation of MPSoC" project in LIP2/FST Laboratory in Tunis.

SIGNAL-BASED PREDICTIVE CONDITION MONITORING

Egon Müller
Chemnitz University of Technology
Institute of Industrial Management
and Factory Systems
Erfenschlager Straße 73
D-09125 Chemnitz, Germany
egon.mueller@mb.tu-chemnitz.de

Steffen Leischnig
Robert Bosch GmbH
Manufacturing Sector 4
Am Börstig 2
D-96052 Bamberg
steffen.leischnig2@de.bosch.com

KEYWORDS
Diagnostics and Maintenance; Automated Inspection & Quality Control; Life-Cycle Cost and Quality

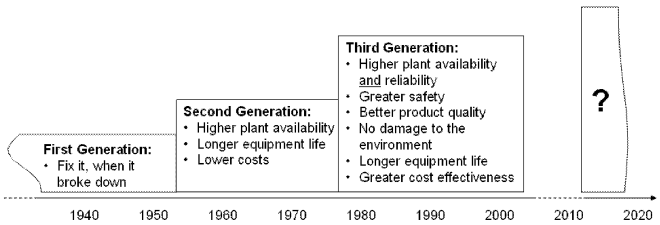
ABSTRACT

One of the most intense desires of mankind is to have the ability to forecast events. A pragmatic, as well as an economically desirable example is to know about the state and the imminent changes in the capability and effectiveness of production equipment. Forecasting of faults can result in reduction of costs, higher availability, performance and quality level of the asset. The appropriate tool to predict these changes is Condition Monitoring. This paper describes a successfully implemented solution by automatic and permanent interpretation of analogue signals in the production of Common Rail Injectors.

INTRODUCTION

History of Maintenance

Until the 1950´s, maintenance only meant: repair in case of fault. The tasks were carried out only by specially trained employees. By determining the recurring intervals, the specific behaviour of components was analysed, so that maintenance could be scheduled in the future. The biggest disadvantage of periodic replacements is that the components were replaced even before the end of their average service life. This resulted in the wastage of money and resources. In addition, due to the fact that only the maintenance personnel were permitted to repair the faults, the companies and institutions were compelled to develop concepts to integrate the operator's knowledge into the maintenance process. The greatest achievement of the 50´s and 60´s with regard to this was the development of Total Productive Maintenance (TPM) and Total Quality Management (TQM), which was chiefly driven by Japanese industry. (Birolini 1997)



Figures 1: History of requirements on maintenance (Moubray 1991)

During the decades which followed, many strategies were developed, most of them based on TPM or elements adapted from it. Reliability, availability, capability and productivity were focussed upon with great intensity. As a result, special strategies such as Reliability Centred Maintenance (RCM) emerged.

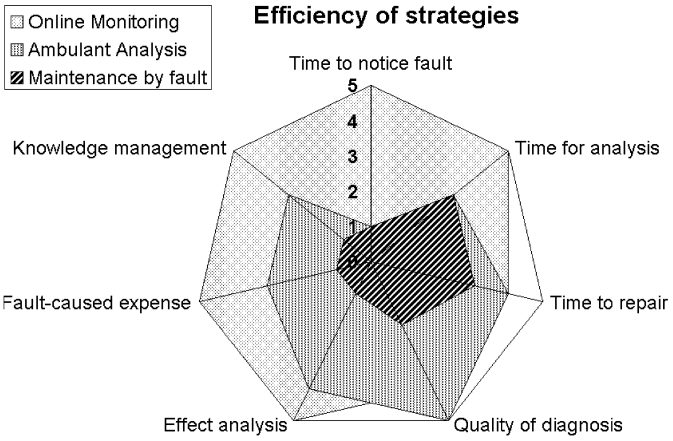
Recent requirements

Due to the changes in targets, tools are needed to be developed to ensure their realization. The above mentioned base concepts of TPM or TQM remain more a philosophy than an instrument. To fulfil the demands of currently used or upcoming strategies, considerable amount of data becomes necessary. Condition monitoring systems as part of technical analysis provide data on the state of components and assets. The main task and advantage of condition monitoring, in contrast to offline diagnosis, is to provide data continuously to avoid faults and unexpected states by detecting them immediately or even before they occur. It thereby particularly fulfils the need for improvements in reliability and availability.

MOTIVATION AND EXPECTATIONS

The project discussed below is under progress in discrete industrial production. The machines, which had to be monitored or analysed, are function test stations in the manufacturing of Common Rail Injectors. Several hundreds of identically constructed machines check in several plants worldwide, whether the fully assembled products meet their defined requirements.
Functional test of injectors is carried out by the recreation of a single cylinder engine without combustion. The injected volume for each "shot" of the magnet valve is measured. The physical conditions are extreme with respect to pressure, accuracy and time - an injection up to 2000 bar has to measure precisely within milliseconds the volume of a drop which is generated by a spray. The measuring instrument and a large number of peripheral elements are extremely sensitive to mechanical and electrical influences and their behaviour could change with minimal deviations in boundary conditions like temperature or in the properties of the fluid used.
For reasons of planned changes in maintenance strategy, a percentage reduction of fault based maintenance, as well as due to the urgent need for reduction in maintenance and

quality costs, Condition Monitoring is considered to be a suitable and a promising solution for these machines. The potential of Condition Monitoring was never in doubt. The decision was needed on the organisation of timed repair process. Whether, an ambulant analysis or a permanent online Condition Monitoring serves the purpose best, was to be decided. Fig. 2 shows the aspects which supported the decision for a permanent Condition Monitoring system. For realising a possibility for real predictability, only the continuous capturing of data is worth considering.



Figures 2: Comparison between the aspects of online and offline diagnosis

CONDITIONS FOR IMPLEMENTATION

The methods of forecasting events in the future are numerous, many of them are well known, although only rarely in practice. Simple examples comprise of extrapolation, trend analysis or the tools contained in Statistical Process Control (SPC). All of them use statistical or mathematical methods which function only with valid, complete and extensive data.
On the one hand, this data could be acquired with more or less technical effort. On the other hand, the timed behaviour of parameters, which need to be predicted, becomes decisive. The listed examples could easily forecast the failures such as sliding average values. Spontaneous failures could not be predicted in this way. The point is, to acquire at least that number of values per period which are enough to reconstruct timed changes in behaviour. Harry Nyquist described it in his theory for measurement technology in 1927 (see Equ. 1). The extent of data, which will be needed, depends on the speed of change. Some spontaneous failures could be predicted by combining direct or indirect parameters, which indicate the forthcoming events and the experience what is most likely to happen in the present constellation.

$$f_{record} >= 2 * f_{max} \qquad (1)$$

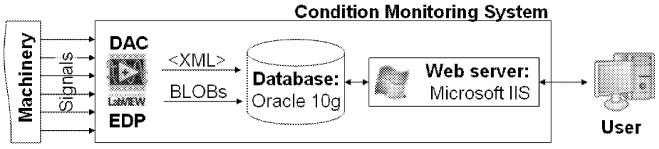
The presented example, Condition Monitoring of test stations in injector production, doesn't primarily take sliding into consideration, so that the simple methods for prediction could be used. The challenge was rather, in the rating of recorded signal sets to get the current state and condition of the elements in the machinery. To rate them, the focus on single characteristics was not sufficient. Complex curves, their form and profile have to be interpreted.

DATA-ACQUISITION, PROCESSING AND VISUALIZATION

The developed Condition Monitoring system consists of 3 basic elements: data acquisition, database and visualisation. As mentioned above, specific engineering was necessary because of the need to interpret complex signal characteristics.
In comparison, simple signals without specific characteristics could be easily monitored by using SPC directly on the acquired data. The example for that is temperature monitoring or FFT (Fast Fourier Transform) on antifriction bearings. To realise this, usually a small sampling rate is sufficient. To rate and estimate complex characteristics, much higher sampling rates are necessary with consequently increased requirements on processing and data-handling equipment.

Hardware architecture

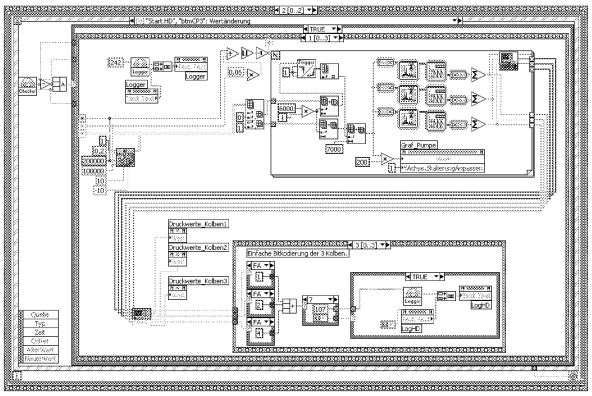
Data is acquired by a measurement PC with standard analogue-digital converter (DAC), signal amplification and data conditioning. For reducing costs, a digitally controlled multiplexer was specially designed to switch analogue signal channels of the connected machines periodically. The entire transportation of data is over Ethernet. Data storage comprises of a cluster of Oracle database servers. A Microsoft IIS web-server generates specific user reports (see Fig. 3).



Figures 3: Data flow and elements of data handling

Software architecture

Software is designed as a 3-Tier-application. Received data sets or signal sequences are interpreted by algorithms implemented in LabVIEW (National Instruments, USA), refer to Fig. 4. The user report pages are generated dynamically by using .NET. All data is read from the database. No direct access by clients to measurement PC is necessary.

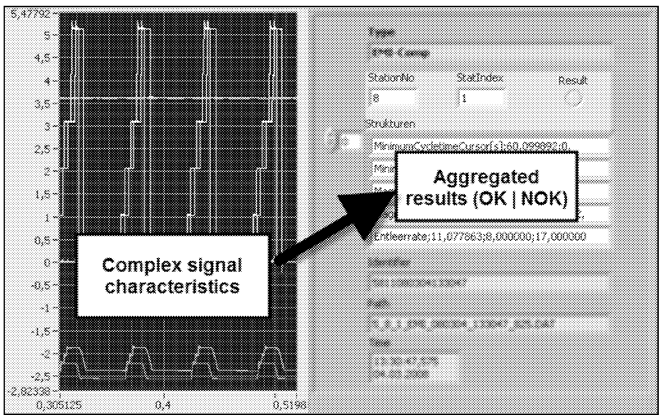


Figures 4: Code in LabVIEW

A possible better solution, particularly with regard to system design, is to integrate the algorithms into the machine. Therefore, the control equipment of a machine should be able to capture and rate signal sets. This is not normally possible in running machines without addition or replacements of components, but with regard to maintainability, an appeal is made to machine building companies to consider integration in new machines.

Processing of data

To reduce the amount of data, the periodically and permanently recorded signal sets are rated on the measurement PC. Only bad-rated signal sets are transmitted to database to provide the user a possibility for further analysis. Good-rated signal sets are not stored further or processed. The interface used for publishing results of rating by electronic data processing (EDP) is an ASCII-file containing only the compressed and highly aggregated results of capability rating.

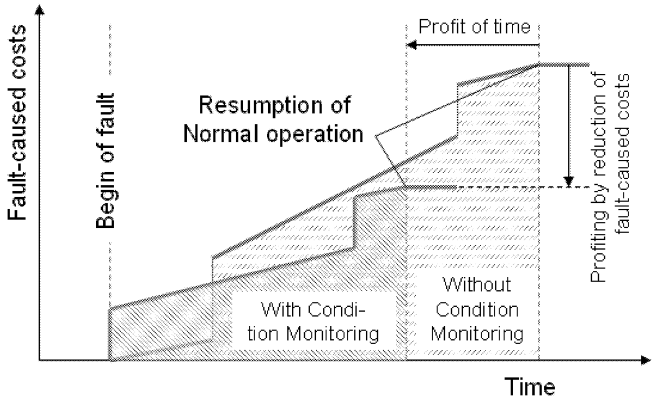


Figures 5: Rating of signal sets

Prediction of faults

The abovementioned aggregated results are, in contrast to the origin signal sets, suitable to be observed by SPC. The results and not the prognosis of forthcoming changes in behaviour or breaking limits, are visualized on dynamic websites generated by the web server.

Fig. 6 shows the profitability by using a Condition Monitoring System. The differences in cost evolution are caused by current information about the fault and, if possible, by the reason for it. Maintenance or repair could be done immediately after an occurrence, normally without further analysis. Profit is possible in two dimensions: reduction of Mean Time to Repair (MTTR) and reduction of total costs for repair. Total costs are reduced by reducing the costs of its elements e.g. material, staff, analysis and trouble-shooting, cost of idleness or follow-up run of the machine.

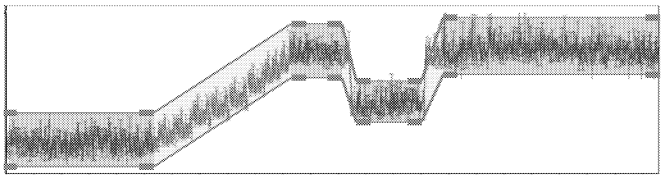


Figures 6: Comparison of cost evolution with and without Condition Monitoring (Proksch et al. 2001; Adam 1989)

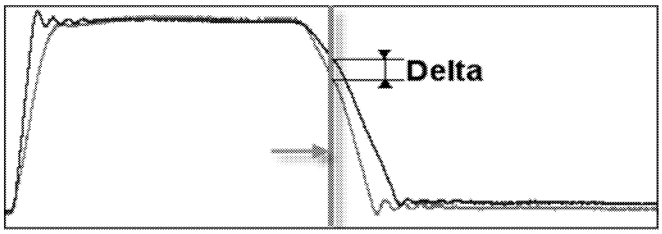
The advantages of Condition Monitoring are described by providing information about faults when they occur. Prediction of faults may save some more costs. The obvious examples are to avoid consequential damages to neighbouring parts of the machine or to arrange repairing at planned intervals to prevent idle times. The less obvious examples are the benefits of less effort in scheduling maintenance tasks and long-term effects like more harmonised production, quality improvement, higher availability and process dependability.

CONCLUSION AND FUTURE WORK

The presented example works in practice as a prototype, in one testing line of 16 machines. Roll-out on other lines has been started. After a few months of operation, imminent failures could be already predicted by using statistical methods on continuously collected ratings of specified criteria. So, the main condition for the introduction of planned, good structured and efficient condition-based maintenance is fulfilled.
In the analysis, proprietary algorithms provide data in the best manner with respect to data quality and performance. Pattern-based methods are more flexible and universally appropriate. The comparison between recorded signal set and ideal sample pattern could be done using envelopes (Fig. 7), interpolation points or edge analysis. To adapt Condition Monitoring easily to interpret and to rate changes or differences in signal curves from specified behaviour, pattern based comparison (Fig. 8), neural networks or analysis by fuzzy sets could be applied. (Williams et al. 2007)



Figures 7: Rating with envelope



Figures 8: Pattern based comparison

Further, studies are needed, especially to solve the integration of Condition Monitoring into new machines. A number of current approaches towards improvement of availability and reliability require information on the machine's state to work properly. In addition, the development of algorithms for rate signal curves using form factors and complex criteria is expected to have a high and rising potential.

REFERENCES

Adam, S. 1989. *Optimierung der Anlageninstandhaltung : Verfügbarkeitsanforderung, Ausfallfolgekosten und Ausfallverhalten als Bestimmungsgrößen wirtschaftliche sinnvoller Instandhaltungsstrategien*. Erich Schmidt, Berlin.

Birolini, A. 1997. *Quality and Reliability of Technical Systems*. Springer-Verlag, Berlin ; Heidelberg ; New York.

Moubray, J. 1991. *Reliability-centred Maintenance*. Butterworth-Heinemann, Oxford ; London.

Proksch, R.; E. Westkämper; H.J. Bullinger. 2001. *Modellbasierte Planung und Bewertung von Instandhaltungsstrategien am Beispiel einer Serienfertigung mit geringer Variantenzahl*. Jost Jetter Verlag, Stuttgart.

Williams, J.D. et al. 2007. "Statistical Monitoring of Nonlinear Product and Process Quality Profiles." *Quality and reliability engineering international*, Volume 23, Item 8, 925-941.

AUTHORS BIOGRAPHY

EGON MÜLLER was born in 1952 in Saxony, Germany. Until 1978 he studied on Zwickau Technical University where he graduated in 1984. He left his first professorship in Zwickau University of Applied Sciences in 2002 to get in the lead of the Institute of Industrial Management and Factory Systems of Chemnitz University of Technology. There he has the chair for Factory Planning and Factory Management.
egon.mueller@mb.tu-chemnitz.de
www.tu-chemnitz.de/mb/InstBF/

STEFFEN LEISCHNIG studied Systems Engineering on Chemnitz University of Technology. In 2006 he started his doctorate in Factory Management as cooperation between the Institute of Industrial Management and Factory Systems and the Robert Bosch GmbH, where he already applied and implemented several close-to-production IT-systems.
steffen.leischnig2@de.bosch.com
www.research.bosch.com

SCHEDULING METHODOLOGY

A ROBUST WHAT-IF BULLWHIP ANALYSIS

Salvatore Cannella
Elena Ciancimino
Faculty of Engineering, University of Palermo
Parco d'Orleans, 90128 Palermo, Italy
cannella@unipa.it, eciancimino@unipa.it

KEYWORDS
Operations research, differential equations, statistical analysis, continuous simulation.

ABSTRACT

The well-known deleterious effect of the amplification of variance of order rates in multi-echelon systems, commonly known as demand amplification phenomenon or bullwhip effect, still presents new challenges and continues to fascinate the operations management community. Since 2000, while the existence of bullwhip phenomenon has been widely confirmed, the discussion has shifted from strategies aimed at preventing the bullwhip effect to collaboration in the supply chain through collaborative planning, information sharing and VMI and other strategies. In the operational context this investigation period is called *the bullwhip avoidance phase*. Recently this research field is focusing on the study of robustness of bullwhip avoidance techniques under uncertainty, as environmental conditions often determine variations in processes, with regards to production and delivery lead time, and variations in the parameters of the decision policies.
This work aims at quantifying the efficacy of bullwhip dampening techniques and at verifying this efficacy against variations in the business context. The original value of the research consists in showing that the bullwhip dampening efficacy, provided by hi-tech applications and practices of supply chain collaborations, persists when operational parameters vary within a certain range.
The adopted research methodology consists of analytical modelling, continuous simulation, and design of experiment.

INTRODUCTION

As reported by Schmenner (2001), the deleterious effect of the amplification of variance of order rates in multi-echelon systems, commonly known as demand amplification phenomenon, was recognized as early as 1919 in the supply chain of Procter and Gamble. Seventy-eight years later Lee, Padmanabhan and Whang (1997a, b) published two of the most popular papers based on a study of the well-known company, coining the term "bullwhip effect" to indicate the phenomenon where the variance of the demand signal increases as the demand signal flows up the supply chain. Today the problem still presents new challenges and continues to fascinate the operations research community.
Holweg and Disney (2005), in their description of the evolving frontiers of bullwhip problem, report a chronological classification of investigation on the supply chain instability in the operational context. They identify six distinct developmental stages in the related literature:

- pre 1958 *the production and inventory control era*,
- 1958-1969 *the production smoothing era*,
- 1970-1989 *the development of control theory*,
- 1989-1997 *the 'beer game' phase*,
- 1997-2000 *the bullwhip rediscovery phase*,
- since 2000 *the bullwhip avoidance phase*.

In the last period, while the existence of bullwhip phenomenon has been widely confirmed, the discussion has shifted from strategies aimed at preventing the bullwhip effect to collaboration in the supply chain through collaborative planning, information sharing and VMI and other strategies. To contribute to the bullwhip avoidance literature, a robust what-if analysis focused on hi-tech applications and practices of supply chain collaboration is presented. This work aims at quantifying the efficacy of bullwhip dampening techniques and at verifying this efficacy against variations in the business context and at verifying this efficacy against variations in the business context. In the real world, environmental conditions often determine variations in processes, with regards to production and delivery lead time, and variations in the parameters of the decision policies. Shared real-time point-of-sales information, sales forecasts, inventory order policies and inventory reports to support multi-tiers integration decision rules represent the key element to achieve the bullwhip dampening ability regardless of the operational parameters.
The research methodology consists of analytical modelling, continuous simulation, and design of experiment. First, systems of nonlinear ordinary difference equations are used to model supply chain. Equations are solved by Vensim software that uses Euler integration as default simulation method. This tool evaluates the system state every constant time interval (Δt), then the new system state is recorded and statistics collected (Ventana Systems Inc., 2006). Finally nine experiments are presented, obtained by varying three parameters of the model with Latin square design. For each scenario two process performance assessments are calculated for every echelon, the Order Rate Variance Ratio and the

Inventory Variance Ratio, and a customer service metric, the Fill Rate.
The article is organized as follows. Section 2 reports a literature review on recent bullwhip effect solutions obtained by innovative supply chain operation practices. The simulation model, the adopted nomenclature and the related set of equations are explained in section 3. Section 4 presents the bullwhip metrics, the design of the experiments, validation and analysis of simulation data. Section 5 presents discussions and suggestions for future research.

'BULLWHIP AVOIDANCE PHASE' STATE OF ART

Supply chain instability, bullwhip effect and the influence of inventory level on demand have attracted the attention of researchers and practitioners in different fields of management science. In the last bullwhip investigation phase, the researchers in the field of demand amplification phenomena have deeply investigated the development of collaboration among partners in the supply chain and innovative information technologies solutions, both with empirical studies of real cases, analytical methods and what-if simulation. Among the contributions to *bullwhip avoidance* are those by Chen et. al (2000), Disney and Towill (2002), Dejonckheere et al. (2004), Chatfield et al. (2004), Gonçalves et al. (2005), Villegas and Smith (2006), Ouyang, (2007), Boute et al. (2007), Jakšič and Rusjan (2008) and Duc et al. (2008).
Chen et al. (2000) investigate the bullwhip problem in a multi-echelon supply chain with information sharing using stochastic approaches. They show the benefit of information sharing for the Order-up-to policies based on exponential smoothing forecast in an enriched scenario and propose an analytical metric to quantify the bullwhip effect.
Disney and Towill (2002) present an analytic solution to the bullwhip problem for the DE-APIOBPCS (Deziel and Eilon Automatic Pipeline Inventory and Order Based Production Control System) order policy. They show that the way to minimise the bullwhip problem with this policy is increasing the average age of forecasts, and reducing the rate at which inventory and WIP correction are accounted for in the production/distribution-ordering algorithm.
Dejonckheere et al. (2004) examine with control theory approach two supply chain information scenarios for Order-up-to and APIOBPCS replenishment policies. They quantify the benefit of sharing end customer demand forecasting by analytically demonstrating the relevant dampening of demand amplification along the tiers.
Chatfield et al. (2004) present a k-stage serial supply chain model to determine Order-up-to levels and order quantities for each stage. Using a simulation model, they perform a set of experiments that employ factors related to lead-time variation, information quality, and information sharing to conclude that information sharing reduces total variance amplification and stage (node to node) variance amplification. Furthermore, their results show that information sharing protects a supply chain against "cascading failures", the effect that stockouts at the upper echelons may have on multiple downstream nodes.
Gonçalves et al. (2005) via system dynamics approach (Sterman, 2000) focus on the causes of oscillatory behaviour in capacity utilization at a semiconductor manufacturer of Intel Corporation and explore the role of endogenous customer demand in influencing the company's production and service level. They suggest that typical policy prescriptions of lean inventory and responsive utilization policies hold only when demand is assumed to be exogenous.
Villegas and Smith (2006) show how an Automatic Planning System (APS) may be modified to better manage the trade-off between order and inventory oscillations. They state that in real supply chains APSs have introduced a new source of demand distortion, which is reflected as high variation in production order quantities as the desired inventory levels are maintained according to company policies. The problem is studied in a four echelon model that faces a seasonal predictable demand, using difference equation and mathematical optimisation models. The paper suggests ways in which the operating policies and the APS model can be modified to reduce the demand amplification phenomenon.
Ouyang (2007), via frequency domain approach, presents a system control framework for analyzing the bullwhip effect in multi-echelon supply chains with information sharing. He derives robust analytical conditions to predict whether or not the bullwhip effect will arise without knowing the customer demand, and also develops exact formulae for the variance of the order stream at any level of a multi-stage chain knowing only the spectrum of the customer demand process. He shows a special case in which sharing customer demand information across the chain significantly reduces, but does not completely eliminates, the bullwhip effect.
Boute et al. (2007) present a model with two-echelon supply chain: a single retailer holds a finished goods inventory to meet an i.i.d. customer demand, and a single manufacturer produces the retailer's replenishment orders on a make-to-order basis. They show that, by including the impact of the order decision on lead times, the order pattern can be smoothed to a considerable extent without increasing stock levels and avoiding bullwhip effect.
Jakšič and Rusjan (2008) demonstrate through transfer function method that demand forecasting with simple exponential smoothing effectively lowers the variability of orders over demand, and thus decreases the probability of the bullwhip effect occurring.
Duc et al. (2008) quantify the impact of the bullwhip effect as the phenomenon in which information on demand is distorted as moving up for a simple two-stage supply chain with one supplier and one retailer. Assuming that the retailer employs a base stock inventory policy, and that the demand forecast is performed via a mixed autoregressive-moving average model ARMA(1,1), they investigate the effects of the autoregressive coefficient, the moving average parameter, and the lead time on the bullwhip effect. They report two significant properties of the bullwhip effect: (1) the bullwhip effect does not always exist, but its existence depends on the values of autoregressive and moving average coefficients of the ARMA model; (2) the bullwhip effect does not always increase when the lead time increases.
As showed, the literature advocates that policies that stimulate information sharing and coordination are critical mitigants to supply chain instability, stockholding, and poor customer care. Supply chain connectivity, cooperation and coordination within the production/distribution network are indispensable to avoid bullwhip effect. In the next section

we present a multi-echelon model structure able to inhibit the bullwhip effect.

THE BULLWHIP AVOIDANCE MODEL

The multi-tier structure is composed by four serially-linked business entities i={1,2,3,4}, whose order policies are based on the integrated data on marketplace demand, tiers' inventory and work in progress. Echelon number four (retailer) accesses sale forecasts and point-of-sales data; its order policy is based on this information and on its inventory and work in progress reports. Echelon number one (manufacturer) gains data from the whole network: real-time sales forecast, and global item positions along the chain.
In the real business world the information structure of our model can be associated to a data-driven decision support system provided by a cyber communication platform that releases continuous feedback on the multi-echelon item flow data (Fig.1)

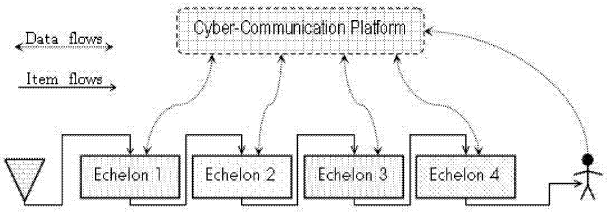


Fig. 1 Multi-tier Structure

Nomenclature

Wip_t^i	work in progress (includes incoming transit units) at echelon i at time t.
Inv_t^i	inventory of goods at echelon i at time t.
DO_t^i	units finally delivered at echelon i at time t.
Th_t^i	throughput at echelon i at time t.
d_t	market demand at time t.
$\hat{d}_t$	customer demand forecast at time t.
R_t^i	order quantity at echelon i at time t.
Bl_{t-1}^i	existing backlog of orders at echelon i at time t.
$VirtInv_t^i$	virtual inventory at echelon i at time t.
$VirtWip_t^i$	virtual work in progress at echelon i at time t.
$TVirtWip_t^i$	target virtual work in progress at echelon i at time t.
$TVirtInv_t^i$	target virtual inventory at echelon i at time t.
$\sigma_{R^i}^2$	variance of order quantity rate at echelon i
σ_d^2	variance of customer demand.
$\sigma_{Inv^i}^2$	variance of inventory at echelon i.
μ_{R^i}	mean value of order quantity at echelon i.
μ_d	mean value of market demand.
α	forecast smoothing factor.
T_p^i	physical production/distribution lead time (incoming transit time from supplier plus the production lead time).
T_c^i	cover time for the inventory control.
T_y^i	smoothing inventory parameter.
T_w^i	smoothing work in progress parameter

Assumptions

(a) A single item with no bill of material is analysed.
(b) There is no capacity constrain for any echelon.
(c) At every echelon the shipments sent by the supplier immediately become work in progress.
(d) For the first echelon's (manufacturer) no production lead time is considered. Note that in real business world this assumption implies that the manufacturer production plant is located next to the raw material warehouse.
(e) Backlogging is allowed as a consequence of stockholding; in each echelon the backlog will be fulfilled as soon as on-hand inventory becomes available.

Equations

$$Wip_t^i = Wip_{t-1}^i + DO_t^{i-1} - Th_t^i \cdot \quad (1)$$

$$Inv_t^i = Inv_{t-1}^i + Th_t^i - DO_t^i \cdot \quad (2)$$

$$Bl_t^i = Bl_{t-1}^i + R_t^{i+1} - DO_t^i \cdot \quad (3)$$

$$DO_t^i = MIN(R_t^{i+1} + Bl_{t-1}^i; Inv_{t-1}^i + Th_t^i) \cdot \quad (4)$$

$$Th_t^i = DO_{t-T_p}^{i-1} \cdot \quad (5)$$

$$\hat{d}_t = \alpha d_{t-1} + (1-\alpha)\hat{d}_{t-1}; \quad 0 < \alpha \le 1. \quad (6)$$

$$VirtWip_t^i = \sum_{j=i}^{4} Wip_t^j \cdot \quad (7)$$

$$VirtInv_t^i = \sum_{j=i}^{4} Inv_t^j \cdot \quad (8)$$

$$TVirtWip_t^i = \hat{d}_t \sum_{j=i}^{4} T_p^j \cdot \quad (9)$$

$$TVirtInv_t^i = \hat{d}_t \sum_{j=i}^{4} T_c^j \cdot \quad (10)$$

$$R_t^i = \hat{d}_t + \frac{1}{T_w^i}(TVirtWip_t^i - VirtWip_t^i) + \frac{1}{T_y^i}(TVirtInv_t^i - VirtInv_t^i); \quad R_t^i \ge 0, \forall i \cdot \quad (11)$$

Equations (1), (2) and (3) represent the state variables of the model. Work in progress and Inventory describe physical flows of items in downstream direction, while Backlog is representative of service level for each tier. Equation (4) expresses the dynamic of delivered orders, allowing backlog as stated in assumption (e). Eq. (5) represents the production/delivery lead time delay, represented by the parameter T_p. Eq. (6) models the exponential smoothing formula to forecast demand (Makridakis et. al, 1978), where d_{t-1} stands for the end customer demand and α is the smoothing parameter. The market demand forecast equation is shared by all the echelons.

Equation (7) is obtained by orders-in-the-pipeline at stage *i* plus the work in progress of downstream echelons. Equation (8) for an individual echelon is the sum of the Inventory at stage i and Inventory of subsequent echelons. Target Virtual Wip (9) and Target Virtual Inventory (10) are two further elements supporting data warehouse for the replenishment rule, generated by two operational parameters and the forecasted demand. Equation (11) models the order rule investigated in this work. It is a well-known variant of an Order-Up-To replenishment rule, namely APIOBPCS (John et al., 1994), largely utilized in industry and studied by academics (Wickner et al., 2007). This replenishment rule implies that at every time the ordered quantity is equal to the forecasted demand, plus a fraction of the discrepancy Target Virtual Wip and actual Virtual Wip ($1/T_w$), plus a fraction of the gap between virtual and actual Inventory ($1/T_y$). The peculiar feature of this order rule lies on taking into account pipeline feedback in the inventory policy and including smoothing parameters to gradually restore wip and inventory desired level.

OUTPUT ANALYSIS

Bullwhip metrics and design of experiments

We set up nine different experiments, obtained by varying three parameters of the model:

(i) *The forecast smoothing factor* α. This value represents the weighting factor of the exponential smoothing rule. The value of α is between 0 and 1. The higher the value of α, the greater is the weight placed on the more recent demand levels. The lower the α value the greater is the weight given to demand history in forecasting future demand.

(ii) *The physical production/distribution lead time* T_p. This parameter represents the incoming transit time at echelon *i* from supplier plus the production lead time.

(iii) *The smoothing inventory parameter* ($T_y=T_w$). T_y is representative of the discrepancy between actual inventory and target inventory levels, which is equal to the smoothing work in progress parameter T_w, representative of the discrepancy between actual WIP and target WIP levels. This peculiar design ($T_y=T_w$) was investigated by Deziel and Eilon (1967), and it is largely demonstrated in literature (Disney and Towill, 2002, 2006) that it lies well within the stable regime with extremely well behaved dynamic response.

We use a 3x3 Latin square design, in which each variable has three levels. Each factor varies between a maximum level and a minimum level, and it also assumes an average value. The chosen design is balanced and orthogonal. Design of experiments is used to check whether the estimated effects of changing the inputs of the simulation models agree with the experts' qualitative knowledge about the system (Kleijnen, 2005b)

The 3x3 Latin square used to design the experiment and the variables' levels are shown in Table 1. Cardinal numbers from 1 to 3 stand for the levels of α, Roman numbers for the levels of T_p, capital letters for the levels of the order policy smoothing parameters T_w and T_y.

	T_p		
A	I	II	III
1	A	B	C
2	B	C	A
3	C	A	B

α	**T_p**	**T_w=T_y**
0.2	1	2
0.5	2	3
0.8	3	4

Table 1. Experimental design and factors' levels

For each scenario we calculate two process performance assessments for every echelon: the Order Rate Variance Ratio (12) proposed by Chen et al. (2000) with the Disney and Towill variation (2003b), the Inventory Variance Ratio (13) proposed by Disney and Towill (2003a), and the Fill Rate (14) as customer service metric (Kleijnen and Smits, 2003).

$$ORVrRatio = \sigma^2_{R^i} \Big/ \sigma^2_d \tag{12}$$

$$InvVrRatio = \sigma^2_{Inv^i} \Big/ \sigma^2_d \tag{13}$$

$$FillRate_t = \left(DO_t^{retailer} \Big/ d_t \right) 100 . \tag{14}$$

The Order Rate Variance (OR Vr) Ratio metric is a smart and concise quantification of the order rate instability. The theoretical value of (12) is equal to 1 in case of absolute absence of demand amplification. When demand is amplified along the chain, the values of the Order Rate Variance Ratio show a geometrical or exponential trend in upstream direction.

The Inventory Variance (Inv Vr) Ratio is a supplementary measure of multi-echelon system instability obtained by comparing the inventory variance magnitude to the variance of the market demand. Equation (13) permits to explore the impact of the deleterious consequences caused by demand amplification and, at the same time, it offers additional value to investigate the effect of order policies on inventory levels. A geometrical or exponential increase of this metric is representative of the transmission of inventory instability along the network, caused by information distortion and unsynchronized decision planning between the trading partners.

The Fill Rate is defined as the percentage of orders delivered 'on time', that is, no later than the delivery day requested by the customer. Equation (14) is calculated as the fraction of demand immediately filled from the stock on hand.

Supply chain performance analysis

In continuous time domain we simulate the dynamics of the information-integrated structure and the response of the system for a step increase in final demand. The model runs are for a total of 100 time units, with a time step equal to 0.25; the demand signal from marketplace changes from 100 items per time unit to 200 items/time unit in the tenth simulation period.
Table 2 reports the values for the OR Vr Ratio and for the Inv Vr Ratio in the nine scenarios, measured in each echelon.

	Echelon 4			Echelon 3	
	OR Vr Ratio	Inv Vr Ratio		OR Vr Ratio	Inv Vr Ratio
1°Set	2,98	20,96		6,54	17,90
2°set	1,49	19,98		2,54	18,33
3°set	1,50	19,62		2,51	17,74
4°set	1,69	29,23		3,11	27,58
5°set	1,91	30,11		3,54	25,85
6°set	7,17	38,87		15,84	26,50
7°set	2,01	37,54		4,01	38,93
8°set	10,39	49,86		22,49	38,56
9°set	6,34	46,56		13,01	33,50

	Echelon 2			Echelon 1	
	OR Vr Ratio	Inv Vr Ratio		OR Vr ratio	Inv Vr Ratio
1°Set	7,59	14,67		5,34	14,23
2°set	3,32	16,55		3,55	15,78
3°set	3,38	15,73		3,76	14,92
4°set	3,68	25,66		3,17	23,57
5°set	4,29	23,16		3,79	21,87
6°set	16,58	21,82		7,09	20,92
7°set	4,68	33,49		3,39	28,03
8°set	22,07	33,03		6,51	27,39
9°set	615,73	30,89		7,47	26,89

Table 2. Bullwhip Magnitudes Outcome

The output indicates that when adopting a collaborative planning supply chain practice there is no geometric increase of the bullwhip with the level in the chain (Dejonckheere et al., 2004). The value of the two reported metrics clearly indicates that demand amplification magnitude does not get massive at higher nodes in the system. Furthermore, in our experiments the variance of inventory decreases along the chain in upstream direction, regardless of the parameter setting. Note that for the first echelon's (manufacturer) no production lead time is considered. Assumption (d) explains the values assumed by OR Vr Ratio at the manufacture stage.
The analysis of two process metrics shows that the bullwhip effect is dampened when the lead time is reduced, and when the smoothing inventory and work in progress parameters increase (Disney and Towill, 2003a). Furthermore, Inv Vr Ratio always decreases when analysing the structure from the retailer to the first echelon. Despite the similar behaviour along the nine sets, there is a considerable difference among the scenarios grouped by lead time: as the lead time increases the value of Inv Vr Ratio becomes higher. An additional significant outcome, in contrast with the result presented by Duc. et al. (2008), consists in the reduction of the lead time, which is closely related to the global performance of the supply chain as it induces demand amplification avoidance, inventory stability and better satisfaction of the actual marketplace demand.
Analysing the Fill Rate (FR) in each scenario enables to get an insight on the properties of the proposed model in terms of customer satisfaction. The FR is reflective of the network's ability to meet the market demand: when all orders are fulfilled in time, the FR at time t is equal to 100%. For every set in the first ten periods the market demand is satisfied no later than the delivery day requested by the customer. For the step increase in demand, a variation in the Fill Rate occurs.
We report in Figure 2 the values of the FR from period 5 until it gets stable to 100%. Diagrams are grouped by lead time.

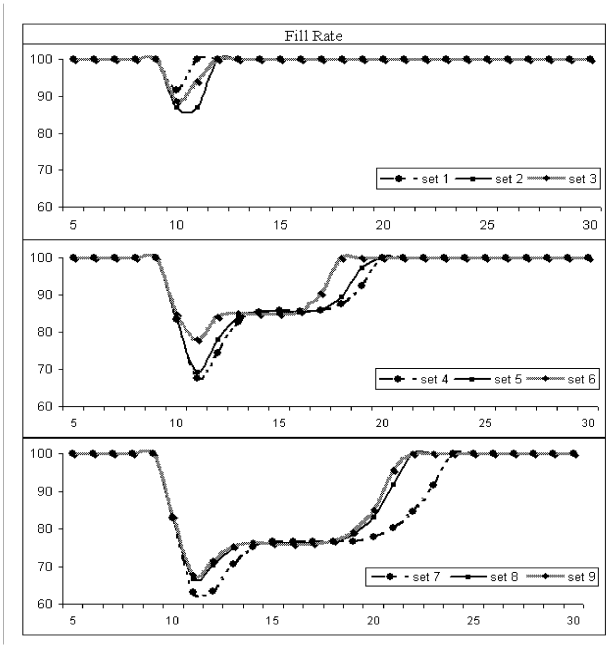


Figure 2. Fill Rates

The values of FR confirm the observation brought forth by the OR Vr Ratio and Inv Vr Ratio metrics. It is important to underline that a global metric high performance, measuring the effectiveness of the entire production/distribution network, implies also local improvement for each node. Among the nine simulated set, only the last group (T_p=3) presents a weak customer satisfaction metric. Nonetheless, even the worst scenario outperforms a non-collaborative traditional ordering structure (see e.g. Rubiano Ovalle and Crespo Márquez, 2003). Moreover, the FR measures confirm the importance of lead time reduction, not only in terms of demand amplification and inventory stability, but also with regards to customer satisfaction.

Model Validation

In order to validate the model, we perform a sensitivity analysis on the resulting data, guided by the statistical theory on design of experiments. Sensitivity analysis may serve

optimization and risk analysis for finding robust solutions (Kleijnen, 2005).
The factorial analysis is performed to evaluate the following null hypotheses: variations of a) the forecast smoothing factor, b) the physical production/distribution lead time, and c) of the smoothing inventory and work in progress parameters are not statistically significant on the considered performance measures. The null hypothesis for this experiment states that the results we obtain are due to chance alone and not to the different parameter values. In order to determine whether the model remains valid over different parameter settings, that is, the configuration is robust, an F test is performed to compare the mean values of two performance metrics, the OR Vr Ratio and the Inv Vr Ratio, along the nine scenarios. The values of the test are reported in Table 3.

	Echelon 4		Echelon 3	
	OR Vr Ratio	Inv Vr Ratio	OR Vr Ratio	Inv Vr Ratio
α	0.13	0.90	0.66	0.92
T_p	0.40	0.99	0.77	0.10
$T_y = T_w$	1.20	2.61	1.99	2.52

	Echelon 2		Echelon 1	
	OR Vr Ratio	Inv Vr Ratio	OR Vr Ratio	Inv Vr Ratio
α	0.07	0.93	0.93	0.94
T_p	0.41	0.99	0.92	0.10
$T_y = T_w$	1.25	2.55	2.69	2.63

Table 3. F test

The critical values for a 1-tailed *F*-test with $\upsilon_1=2$ and $\upsilon_2=2$ degrees of freedom are respectively 19.00 at 95% confidence level, 39.00 at 97.5% and 99.00 at 99%. As showed in Table 3, the obtained values for the F test are smaller than the critical values. The statistic F become small as the null hypothesis is true (Montgomery, 1991). According to the results of the F test, our data is consistent with the null hypothesis that all sample means are equal; this turns into stating that the performance metrics do not actually differ across the nine experimental settings. We derive from these results important insights into the model: we are able to affirm that this configuration behaves in a stable manner towards certain variations of the parameters α, T_p and T_w.

DISCUSSION

In 2006 Geary, Disney and Towill presented the most recent historical review on the demand amplification problem, in which they identify the ten major causes of bullwhip and discuss on the expected future impact of the famous phenomenon. They suggest that a way forward to systematically remove all avoidable causes of uncertainty is to re-engineer the supply chain. This requires the effective application of business systems engineering principles involving technical, cultural, organisational, and financial aspects of the project.
This work presented an operational and technical re-engineering of serially-linked supply chain to avoid the demand amplification phenomena. We have performed a robust what-if analysis focused on hi-tech applications and practices of supply chain collaboration. The supply chain business entities adopt a replenishment rule and order fulfilling decisions supported by a data-driven decision support system provided by a cyber communication platform that makes available constant feedback data on marketplace demand, tiers' inventory and work in progress.
To perform our what-if analysis we have set up a difference equations system implemented through a Vensim tool (Ventana Systems Inc., 2006). The simulation provides an assessment of the dynamics of the structure and the response of the information-integrated system to demand variations in terms of bullwhip effect.
Furthermore a Latin square design of experiment approach is adopted for both quantifying the benefits resulting from coordination, connectivity and cooperation through information technology-enabled business practices and demonstrating the robustness of such practices across a reasonably wide range of scenarios, characterised by different values of the forecast smoothing factor, the physical production/distribution lead time, the smoothing inventory and work-in-process parameters.
We have shown that the e-business technology infrastructure allows making real-time decisions on the basis of marketplace demand variations: shared real-time point-of-sales information, sales forecasts, inventory order policies and inventory provide a win-win business operational strategy to avoid bullwhip effect. We quantified the value of information technologies and supply chain collaboration practices to improve production/distribution network competitiveness, and furthermore we verified that competitiveness against variations in the structure setting.
The original value of this research consists in having shown that the bullwhip dampening efficacy, provided by hi-tech applications and practices of supply chain collaborations, persists when operational parameters vary within a certain range.
Finally, another relevant outcome coming from the present work is the close correlation between lead time reduction, inventory instability, customer satisfaction and bullwhip effect avoidance: the supply chain global performance improves when information or materials delay decrease.
Future research will involve testing the efficacy of hi-tech applications and practices of supply chain collaboration under several real supply chain conditions as production plant constraints and interruptions.

ACKNOWLEDGEMENTS

The authors thank the Ph.D. program in "Sustainability, Quality, Safety and Logistic Management" of the University of Palermo.

REFERENCES

Boute, R., N., Disney, S., M., Lambrecht, M., R., Van Houdt, B., 2007. An Integrated production and inventory model to dampen upstream demand variability in the supply chain. European Journal of Operational Research, 178 (1), 121-142.

Chatfield, D., C., Kim, J., G., Harrison, T., P, Hayya, J., C., 2004. The Bullwhip Effect – Impact of Stochastic Lead Time, Information Quality, and Information Sharing: A Simulation Study. Production and Operations Management, 13 (4), 340-353.

Chen, F., Drezner, Z, Ryan, JK., Simchi-Levi, D., 2000. Quantifying the bullwhip effect in a simple Supply Chain: the impact of forecasting, lead-times and information. Management Science, 46 (3), 436-43.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., Towill, D.R., 2003. The impact of information enrichment on the bullwhip effect in Supply Chains: A control engineering perspective. *European Journal of Operational Research*, 153 (3), 727-750.

Deziel, D., P., Eilon, S., 1967. A Linear Production – Inventory Control Rule. The Production Engineer, 43, 93-104.

Disney, S., M., Towill, D., R., 2002. A discrete transfer function model to determine the dynamic stability of a vendor managed inventory supply chain. International Journal of Production Research, 40 (1), 179-204.

Disney, S., M., Towill, D., R., 2003a. On the bullwhip and inventory variance produced by an ordering policy. Omega, the International Journal of Management Science, 31 (3), 157-167.

Disney, S., M., Towill, D., R., 2003b. The effect of vendor managed inventory (VMI) dynamics on the Bullwhip Effect in supply chains. International Journal of Production Economics, 85 (2), 199-215.

Disney, S., M., Towill, D.. R., 2006. A methodology for benchmarking replenishment-induced bullwhip. Supply Chain Management: An International Journal, 11 (2), 160-168.

Duc, T., T., H., Luong, H., T., Kim, Y.-D., 2008. A measure of bullwhip effect in supply chains with a mixed autoregressive-moving average demand process. European Journal of Operational Research 187 (1) 243-256

Geary, S., Disney, S., M., Towill, D., R., 2006. On bullwhip in supply chains - historical review, present practice and expected future impact. International Journal of Production Economics 101 (1) 2–18.

Gonçalves P., Hines, J. Sterman J. 2005. The impact of endogenous demand on push–pull production systems. System Dynamics. Review, 21, 187-216.

Holweg, M., Disney S., M., 2005. The Evolving Frontiers of the Bullwhip Problem, Proceedings of the Conference EurOMA: Operations and Global Competitiveness, Budapest Hungary June 19-22, 777-716.

Jakšič, M., Rusjan, B., 2008. The effect of replenishment policies on the bullwhip effect: A transfer function approach. European Journal of Operational Research, 184 (3), 946-961.

John, S., Naim, M., M., Towill, D., R., 1994. Dynamic analysis of a WIP compensated decision support system. International Journal of Management Systems Design. 1 (4), 283-297.

Kleijnen, J., P., C., 2005. An overview of the design and analysis of simulation experiments for sensitivity analysis. European Journal of Operations Research, 164 (2), 287-300.

Kleijnen, J., P., C., Smits, M., T., 2003. Performance metrics in supply chain management. The Journal of Operational Research Society, 54 (5), 507-514.

Lee, H., L., Padmanabhan, V., Whang, S., 1997a. Information distortion in a Supply Chain: The Bullwhip effect. Management Science, 43 (4), 546-558.

Lee, H., L., Padmanabhan, V., Whang, S., 1997a. The Bullwhip effect in supply chains. Sloan Management Review 38 (3), 93-102.

Makridakis, S., Wheelwright, S.C., McGee, V.E. 1998. Forecasting. Methods and applications. West Sussex, UK, John Wiley & Sons.

Montgomery, D., C., 1991. Design and analysis of experiments. New York, USA, John Wiley & Sons.

Ouyang, Y., 2007. The effect of information sharing on supply chain stability and the bullwhip effect. European Journal of Operational Research 182 (3), 1107-1121.

Rubiano Ovalle, O., Crespo Marquez, A. 2003. The effectiveness of using e-collaboration tools in the supply chain: an assessment study with system dynamics. Journal of Purchasing and Supply Management, 9 (4), 151-163.

Schmenner, R.W., 2001. Looking ahead by looking back: swift, even flow in the history of manufacturing. Production and Operations Management, 10 (1), 87–96.

Sterman J., D., 1989. Modelling managerial behaviour: misperceptions of feedback in a dynamic decision-making experiment. Management Science, Vol. 35 (3), pp. 321-339.

Ventana Systems Inc., 2006. The Ventana Simulation Environment: VENSIM 5.5D. Ventana Systems Inc., Cambridge, MA.

Villegas, F., A., Smith, N., R., 2006. Supply chain dynamics: analysis of inventory vs. order oscillations trade-off. International Journal of Production Research, 44 (6), 1037-1054.

Wikner, J., Naim, M., M., Rudberg, M., 2007. Exploiting the order book for mass customized manufacturing control system with capacity limitations. IEEE Transactions on Engineering Management, 54 (1), 145-155.

BIOGRAPHY

SALVATORE CANNELLA was born 1980 in Marianopoli, Italy. He holds a master's degree in Management Engineering (summa cum laude) from the University of Palermo, 2005. Since 2004 he has been committed to several international cooperation research projects, sponsored by the EU Socrates II programme, Spanish Ministry of Education and Science (MEC) and the Italian Ministry of University and Research (MIUR). He is member of Production Engineering Research workGroup (PERGROUP). His research is focusing on bullwhip effect reducing approaches, replenishment rules, IT-enabled supply chain issue, system dynamics methodology and what-if simulation in operations management. In 2006 he performed a training stage at Embassy of Italy in Panama under the provisions set forth by agreement between the Italian Ministry of Foreign Affairs and the Council of Italian Universities (CRUI/MAE internship). He is now Ph.D. student at Department of Manufacturing, Industrial and Management Engineering of the University of Palermo, and visiting at Department of Industrial Management, High School of Engineering, University of Seville, Spain.

ELENA CIANCIMINO was born in 1980. She graduated from the University of Palermo in 2005 with a First class Honours Degree with distinction in Management Engineering. She has been visiting student at Edena (Buenos Aires), a member of WorldWide Classroom Consortium for International Education & Multicultural Studies. Between 2005 and 2006, she was awarded a grant to perform an internship within the EU Leonardo Da Vinci mobility Programme. Since March 2006 she has been an engineering fellow at ISMETT (Mediterranean Institute for Transplantation and Advanced Specialized Therapies), Tissue & Stem Cells Laboratory project. Her research interests include supply chain modelling, robust design and assessment techniques. Current position: Ph.D. at Department of Manufacturing, Industrial and Management

Engineering of the University of Palermo, and visiting at Department of Industrial Management, High School of Engineering, University of Seville, Spain.

Holistic Metrics Applied to a Practical Supply Chain Study

António E. S. Carvalho Brito
J. Manuel Feliz-Teixeira

GEIN – Section of Management and Industrial Engineering
Faculty of Engineering of the University of Porto, PORTUGAL
Emails: acbrito@fe.up.pt, feliz@fe.up.pt

KEYWORDS: holistic metrics, simulation, decision making, complex behaviour, practical example, costs, additives, lubricants.

ABSTRACT

In this article we present a practical example where the concept of *Holistic Metrics* (Feliz-Teixeira & Brito, 2006) is used as a way for sensing the results obtained in a simulation study related to the Portuguese enterprise *GalpEnergia*. Previously described in Feliz-Teixeira & Brito (2005), that case is used again in this article in order to demonstrate the practical value of the *Holistic Metrics* concept. This is basically a simple method for estimating or characterizing the behaviour of *complex systems*, in particular when these are studied throughout simulation. Instead of treating the complex output data obtained from this kind of system along the *time domain*, this perspective encourages analysts to represent such data in the *frequency domain*, somewhat as is adopted in *Fourier Analysis* and in *Quantum Mechanics*.

1. Introduction

Currently, there are very few references of practical relevance about reliable metrics for helping in the interpretation of complex systems behaviour. These systems are often based on intricate structures where a high number of entities interact and properties emerge. Even if there are metrics for appropriately characterizing the nodes or individual parts of these structures, measures for the structure as a whole either fail or appear to be too simplistic. Hopp & Spearman (2001, pp.16), for instance, comment on this saying that "*too much emphasis on individual components can lead to a loss of perspective for the overall system*".

The *Holistic Metrics* technique presented here is expected to help surpass this vacuum by giving the analyst a more holistic perspective on the system's behaviour, as well as letting him/her choose almost freely the *complex states* in which such behaviour is to be *projected*. The final hope is that this method will lead to simpler processes in characterizing certain complex systems. For a detailed description of these concepts please refer to Feliz-Teixeira & Brito (2006).

2. A short explanation

In our *Holistic Metrics* proposal we represent the overall *system state* (ψ) in terms of certain *base functions* (ψ_i), and then we only have to find a way to measure the *probabilities* (α_i) associated with each of these functions. This can be simply done by means of simulation. An interesting aspect of this method is that each base state function (ψ_i) can even be arbitrarily chosen by the analyst, and the probabilities (α_i) can most of the time be easily estimated during the simulation through a process of counting. Final results will then be summarised in some expression of the form:

$$\psi = \alpha_1 \psi_1 + \alpha_2 \psi_2 + \ldots \alpha_j \psi_j + \ldots \alpha_n \psi_n \qquad (1)$$

which could be interpreted as follows: there is a probability of α_1 that the system will be found in the state ψ_1, a probability of α_2 that the system will be found in the state ψ_2, etc. This would be the final measure of the system, in a sort of characterization of expectations under certain conditions. This also

corresponds to *projecting* the system behaviour into the generalised vectors base of *state functions* (ψ_i). The amounts α_i simply correspond to the values of those projections.

Thus, similarly to what happens in *Fourier Analysis*, for example, where signals are to be decomposed into *sine* and *cosine* functions, the time dependent complex behaviour of our complex system will be decomposed into the functions ψ_i and that way moved to the *frequency* domain. A characterization of the system can then be made based on a type of generalised histogram, like the one presented in figure 1:

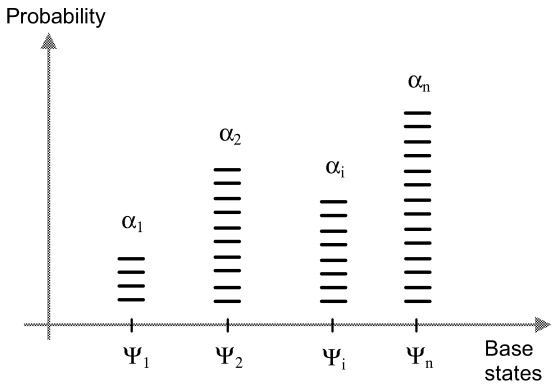


Fig. 1 Characterization of the system's behaviour

So, once the *base states* are well defined by the analyst, the characterization of the system is possible, no matter how complex the system is. We recall that in many practical cases the analyst is mainly focused on being sure that certain variables of the model do not cross some upper or lower limits, or, if they do, with which probability this happens.

3. The case under study

The practical case referred in this article concerns the procurement of additives for lubricants in the oil refinery of Porto, Portugal, and was part of a project of simulation in which we had the intention to help the company *GalpEnergia* decide which purchasing policy to adopt in its Lubricants-Base Plant. As noticed earlier, such a study was published in Feliz-Teixeira & Brito (2005). In it we have basically analysed the supply chain structure presented in figure 2, which was subjected to eight different scenarios of purchasing policies. These policies, in turn, where defined in terms of the minimal amount of materials to order, and analysed based on *purchasing costs*, *holding costs* and *stockout costs*.

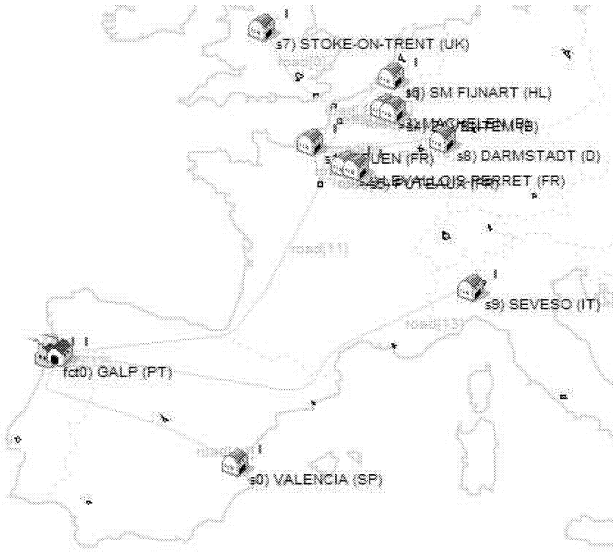


Fig. 2 The complete Supply Chain under study, with its 10 European suppliers

The results led us to conclude that "*the best approach to operate the plant's warehouse GALP(PT) would be to ensure a fortnightly inspection of its inventory and, once operating with the model (r_o, Q_o), to set r_o as the average demand (in each product) and Q_o as the respective expected accumulated demand between supplies, not forgetting to order always at least 8% of a truck. This would lead to a total expected savings of around 300 K€ in one year of operations*".

Such advice has resulted from comparing the following eight scenarios:

1) *hist(30)* – original historical data, model (r=ro, Q(t)).
2) *hist(30)_ss_qq* – historical data, model (r=avrgDem, Q=accDemSuppl)
3) *stat(30)* – statistical data, model (r=ro, Q(t)).
4) *stat(30)_ss_qq* – statistical data, model (r=avrgDem, Q=accDemSuppl)
5) *stat(15)* – statistical data, model (r=ro, Q(t)).
6) *stat(15)_ss_qq* – statistical data, model (r=avrgDem, Q=accDemSuppl)
7) *stat(7)* – statistical data, model (r=ro, Q(t)).
8) *stat(7)_ss_qq* – statistical data, model (r=avrgDem, Q=accDemSuppl)

Notice that the first two scenarios were directly configured with the historical data taken from the plant's database, while the others were driven by statistical versions of such data transformed into a *monthly*, *fortnightly* and *weekly* revision of inventory. All these scenarios have then been simulated and the one with best performances chosen (in the case, scenario 6 = *stat(15)_ss_qq*).

The next figure (Fig. 3) shows how the final

costs were changing along these eight scenarios, once focusing the analysis in the *purchasing*, *holding* and *stockout costs*. These values are already those taken from the best truck occupation level, which resulted from various other comparisons within each scenario and was found to be 8% of a truck. We will not consider this aspect in the *Holistic Metrics* analysis, since, for simplicity, we decided to reduce the problem to the choice of the best scenario. The results presented in figure 3 are already corrected by the multiplicative factors with which the demands of all the scenarios have been made equivalent (see Feliz-Teixeira & Brito, 2005).

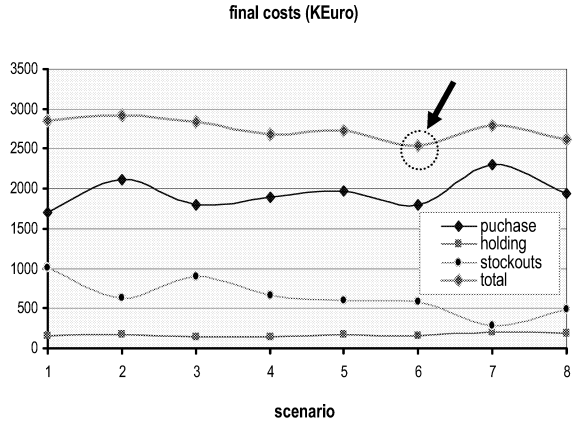


Fig. 3 Final results, of accumulated costs in a year

As a final remark of this section, one must remind that the previous results were obtained by considering only 108 of the 200 products used in the Lubricants-Base Plant; therefore a more detailed analysis would be necessary to give us a better confidence level on the process of decision.

4. Reasoning under *Holistic Metrics*

Management is the science of decisions, therefore in many cases it may almost connect with a certain artistic intuition. In effect, no complex decisions are solely made based on a single criterion, but instead on a weighted equation of criteria which may even contain several solutions. At the ultimate point, it is not infrequent that the manager, as people in their ordinary lives do, has to make a choice from several equal interesting things. This explains the usefulness of certain artistic skills in a manager, skills which are usually seen as intuition for business or simply as a certain ability to predict the future. Although this opinion can be seen as a promotion of *intuition*, it is for sure not in favour of *empiricism*, but instead in favour of good scientific knowledge with which the window of decision can be narrowed, thus transforming scientific knowledge into a tool for the final usage of an intuitive verdict.

Holistic Metrics, in a certain way, attempt to bring together such fields, an intuition which will depend on the analyst's sensibility, and the crude rationality of science. As referred to in previous articles, a marriage between Bergson[1] and Kant. So, the figure of the analyst will always be present in this method, and the results will depend on the skills and on the ability that he/she shows on "feeling" the system.

In an attempt to present a practical example, we will analyse again the case of the refinery described in the previous section. We first supposed that the manager was mainly interested in projecting the system's behaviour into 3 complex states related to the *purchasing*, *holding* and *stockout costs*, as in the previous analysis, but then we included one more state related to the *satisfaction* with its suppliers (Feliz-Teixeira, 2006, pp. 69). Later, however, we chose to double this *base* of *vectors* by considering conditions which express not only the *advantages* but also the *disadvantages*. So, in effect we will use 8 *base states*, instead of 4. The behaviour of the system will be projected onto these 8 *base states*.

5. Specifying the *base states* (ψ_i)

As mentioned in the original article of *Holistic Metrics* (Feliz-Teixeira & Brito, 2006), the *base state functions* (or *base vectors*) can freely be chosen by the analyst in terms of *certain conditions* that must be analysed in the system. In this case, we have considered that the analysis would be mainly focused on the same issues of the previous method, that is, on the *purchasing*, the *holding* and the *stockout costs*. After inspecting the data related to these factors, we could acquire some sensibility for comparing different scenarios, that helped us in choosing the vector's base. The next figure (Fig. 4) intends to help the reader to understand how this can be done. Notice that we have opted for two levels for data comparison: one for those situations

[1] Henry Bergson, a French philosopher from the beginning of the 20th century, who considered time as a history, instead of as an event or as a distance.

considered *advantages* (situations which we will represent by the (+) signal), and another for those considered *disadvantages* (represented by the (-) signal). For instance, in the case of figure 4, any situation in which the accumulated *costs of purchase* lie under the S_{down} level will be considered an *advantage*, while situations where these costs lie over the S_{up} level are considered *disadvantages*.

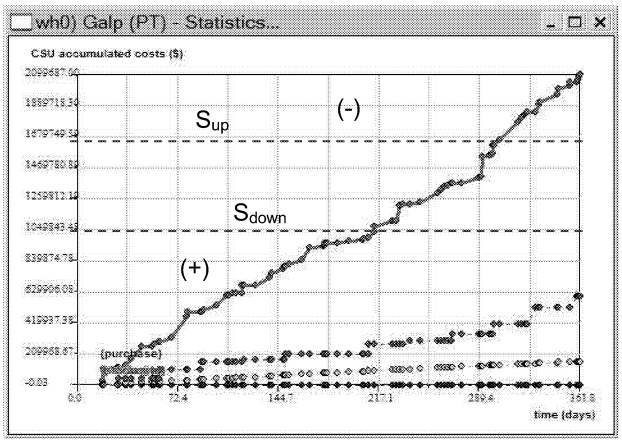


Fig. 4 Example on how to choose *base vectors*

Notice that the S_{down} and the S_{up} levels can be any, and the best way to choose them will depend on the sensibility and the intentions of the analyst. It is here where Kant starts to meet Bergson. A similar strategy was used for choosing the other *base functions*, which may simply be stated as (notice that these terms will be shown in a short form on the charts):

- Disadvantages:

ψ_1 = *(-) accPurchasingCosts > 1650 K€*
ψ_2 = *(-) accStokoutsCosts > 960 K€*
ψ_3 = *(-) accHoldingCosts > 150 K€*
ψ_4 = *(-) satisfSupplier < 90 %*

- Advantages:

ψ_5 = *(+) accPurchasingCosts < 1000 K€*
ψ_6 = *(+) accStokoutsCosts < 300 K€*
ψ_7 = *(+) accHoldingCosts < 50 K€*
ψ_8 = *(+) satisfSupplier > 95 %*

6. The scenarios finally simulated

As previously said, 8 scenarios have been used in the previous approach in order to search for the "best" solution. Here, we have appended to these, four other scenarios in which the *transportation costs* were considered *null*, in order to have some base for comparisons. So, in the end, the following 12 scenarios have been simulated:

1) *hist(30)* – original historical data, model (r=ro, Q(t)).
2) *hist(30)_ss_qq* – historical data, model (r=avrgDem, Q=accDemSuppl)
3) *stat(30)* – statistical data, model (r=ro, Q(t)).
4) *stat(30)_ss_qq* – statistical data, model (r=avrgDem, Q=accDemSuppl)
5) *stat(15)* – statistical data, model (r=ro, Q(t)).
6) *stat(15)_ss_qq* – statistical data, model (r=avrgDem, Q=accDemSuppl)
7) *stat(7)* – statistical data, model (r=ro, Q(t)).
8) *stat(7)_ss_qq* – statistical data, model (r=avrgDem, Q=accDemSuppl)

9) *hist(30)_zero* – original historical data, null transportation costs.
10) *stat(30)_zero* – historical data, null transportation costs.
11) *stat(15)_zero* – statistical data, null transportation costs.
12) *stat(7)_zero* – statistical data, null transportation costs.

To the simulator application was added the ability for counting the occurrences of the appropriate conditions, in order to compute an estimation of the probabilities α_i corresponding to the various ψ_i. Each scenario was simulated independently, and the data resulting from it recorded in two general tables.

7. Results and discussion

The simulator used in this study was developed by Feliz-Teixeira during his doctoral work, and it was specifically conceived for supply chain modelling and simulation[2]. With it, each scenario was simulated during one year of operation. Then, the output data was analysed using the *Holistic Metrics* tool included in it, and the final tables of results were created. Table 1 refers to the situations of *disadvantages*:

	(-) purchase > 1650 k€	(-) stockouts > 960 k€	(-) holding > 150 k€	(-) satisf < 90%
hist(30)_zero	0.00	6.31	8.74	3.53
hist(30)	6.47	6.31	8.74	4.71
hist(30)_ss_qq	18.16	0.00	9.17	2.70
stat(30)_zero	7.36	0.00	0.00	3.16
stat(30)	8.25	0.00	0.00	3.16
stat(30)_ss_qq	24.37	0.00	0.00	3.75
stat(15)_zero	7.06	0.00	2.92	3.91
stat(15)	7.81	0.00	2.92	3.91
stat(15)_ss_qq	0.00	0.00	0.00	2.70
stat(7)_zero	2.12	0.00	2.13	2.87
stat(7)	7.98	0.00	2.13	2.87
stat(7)_ss_qq	0.00	0.00	0.00	1.35
	%	%	%	%

Table 1 Results for the conditions of *disadvantages*

A similar table for the conditions of *advantages* has also been created:

[2] See http://www.fe.up.pt/~feliz/supplychainsimulator.htm

	(+) purchase < 1000 k€	(+) stockouts < 300 k€	(+) holding < 50 k€	(+) satisf > 95%
hist(30)_zero	64.24	17.80	24.76	88.24
hist(30)	62.78	17.80	24.76	85.88
hist(30)_ss_qq	45.52	55.69	22.75	87.84
stat(30)_zero	60.15	33.12	38.45	90.53
stat(30)	60.15	33.12	38.45	87.37
stat(30)_ss_qq	49.67	35.57	27.53	83.75
stat(15)_zero	62.47	38.29	33.87	82.03
stat(15)	62.46	38.29	33.87	81.25
stat(15)_ss_qq	67.59	64.58	31.02	83.78
stat(7)_zero	62.94	100.00	32.27	81.03
stat(7)	60.99	100.00	32.27	82.18
stat(7)_ss_qq	74.63	76.76	34.38	87.84
	%	%	%	%

Table 2 Results for the conditions of *advantages*

Finally, by organizing this data in the form of the appropriate graph, one gets to the final tool from which decisions can be made, that is, to what we may call the *Holistic Metrics Chart*, which in effect corresponds to the *spectrum* concept of *Fourier Analysis* or to the graph of *observable states* of *Quantum Mechanics*. As we will soon notice, from now on decisions will tend to request the help of a certain intuition, and that is where Bergson starts to surpass Kant.

The next chart represents the projections of those conditions considered *disadvantages*:

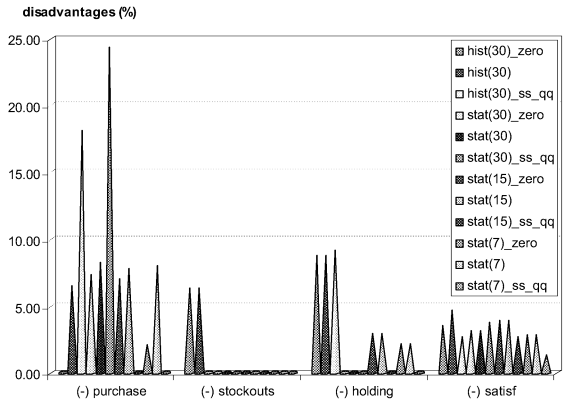


Fig. 5 Holistic Metrics Chart with the final projections for the *disadvantages*

As we can see by inspecting this chart, some scenarios could simply be rejected if the analyst's intention would be the simple minimising of disadvantages. In effect, scenario *stat(30)_ss_qq* shows the highest disadvantage projection in the *purchase* "direction", therefore we could simply reject it. Anyhow, wisdom would perhaps tell the analyst not to pursue this initial impulse, because the same scenario could probably also show a large projection in the *advantages* chart. Actually, that is not the case, as one can see by inspecting the *advantages* chart shown in figure 6, and so the rejection of this scenario could be considered a good decision.

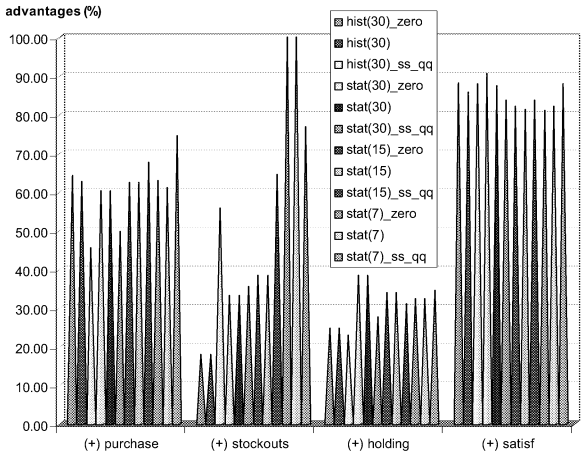


Fig. 6 Final holistic projections for the *advantages*

Even if it seems much easier to make decisions in the *disadvantages* chart, it is important to notice that this will depend on the choice of the various S_{up} and S_{down} levels for the definition of the *base vectors*. In respect to the vector *(-)stockouts* of the *disadvantages* chart (Fig. 5), for instance, we can see that it is perhaps not the best suitable for allowing us to compare the various scenarios, since many of the projections are in effect null. This base vector was chosen to be *accStokoutsCosts > 960 K€*, and so we would recommend that the analyst try a lower value in this condition, like, for instance, *700 K€*.

On the other hand, by inspecting the *advantages* holistic chart (Fig. 6), we can see that in certain "directions" the various scenarios behave with little disparity, meaning that decisions in fact have to be taken based on the analyst's instinct. In a first phase of our analysis, for example, one can even decide to remove certain of those vectors if we observe that their contribution to decision is low. In effect, *(+)purchase* and *(+)satisf* can be ignored on the *advantages* holistic chart, while the *(-)satif* can be ignored on the *disadvantages* chart. In doing so, we would stay with the following data in order to enter

the next level of decision:

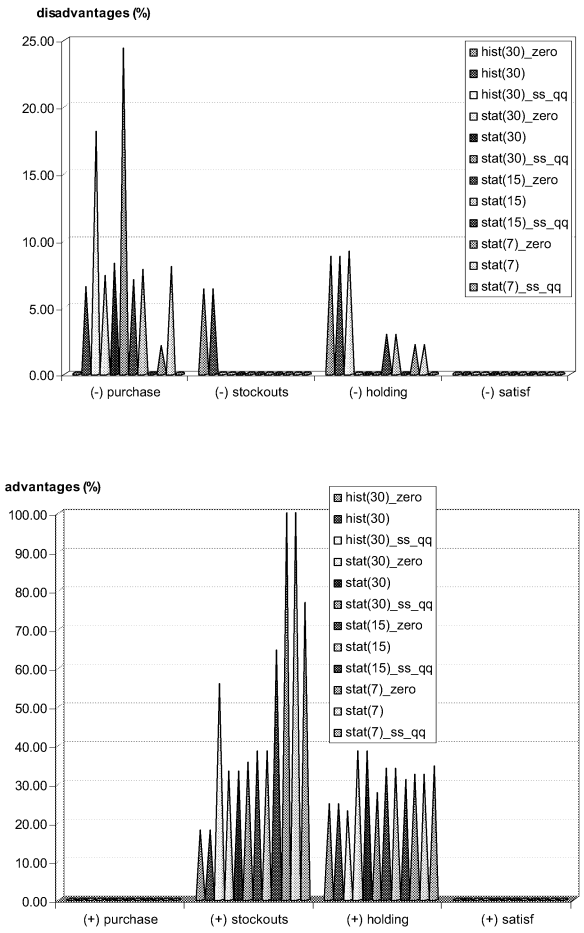


Fig. 7 Holistic data after rejecting certain vectors

From this we can see that, by working a little more in this process, it would not be difficult to reach the situation of making a choice based on the best performances concerning *stockouts*, naturally focusing the attention on scenarios *stat(15)_ss_qq*, *stat(7)* and *stat(7)_ss_qq*, as they are the most prominent. This means that slowly the analyst would probably be led to the same decision reached with the previous method, which was the scenario *stat(15)_ss_qq*. Notice that scenario *stat(7)_zero* is just hypothetical, so it may not be considered an option.

8. Conclusions

We have presented a practical and didactical example on using *Holistic Metrics* to analyse the behaviour of a real system. This is a method in which the analyst is encouraged to exploit his/her intuition during the strategy of approaching a final decision, even if all the measures in the process are in effect quantitative. The case taken as an example has also been analysed in a previous article by means of a standard method. By roughly comparing both processes, we were led to believe that, in the end, *Holistic Metrics* would probably lead to the same conclusions. However, a final choice of the best scenario was not obviously made, in this case, since it would depend on the analyst's feeling and because that was not the major objective of the article. We hope, however, that at least the reader could feel encouraged to finish such reasoning, and exercise intuition in order to finally get to a good choice.

Authors Biography:

J. Manuel Feliz-Teixeira graduated in Physics in the Faculty of Sciences of University of Porto, Portugal, and received an MSc and PhD from the Faculty of Engineering of the same university. His work has been related to various matters, from optical communications, solar energy and seismology to, more recently, the simulation of complex systems in management science, like warehouse and supply chain. His PhD thesis is on "Flexible Supply Chain Simulation".

António E. S. Carvalho Brito graduated and received MSc in engineering from the Faculty of Engineering of University of Porto and PhD in Simulation from the Management Department of University of Cranfield (UK). He teaches at the university since 1981 on the subjects of simulation, programming and systems information.

References:

Feliz-Teixeira, J. M. (2006). *Flexible Supply Chain Simulation (published thesis)*. Porto, Portugal: Publindustria Publisher.

Feliz-Teixeira, J. M., & Brito, A. E. S. C. (2005). *Using Simulation to Analyse the Procurement of Additives for Lubricants in the Oil Refinery of Porto, Portugal.* Paper presented at the Industrial Simulation Conference 2005, Berlin, Germany.

Feliz-Teixeira, J. M., & Brito, A. E. S. C. (2006, October 23-26). *Holistic Metrics, a Trial on Interpreting Complex Systems.* Paper presented at the ESM'2006 The European Simulation and Modelling Conference, LAAS, Toulouse, France.

Hopp, W. J., & Spearman, M. L. (2001). *Factory Physics* (second ed.): Irwin McGraw-Hill.

HEURISTIC APPROACH AS A WAY TO IMPROVE SCHEDULING IN ERP/APS SYSTEMS

František Koblasa

Dept. of Manufacturing systems, University of Liberec
Studentská 2
461 17 Liberec 1,Czech Republic
E-mail: frantisek.koblasa@tul.cz

Luís S. Dias
José A. Oliveira
Guilherme Pereira
Dept. Produção e Sistemas, Universidade do Minho
Campus de Gualtar
4710-057 Braga – P, Portugal
E-mail: {lsd|zan|gui@dps.uminho.pt}

KEYWORDS
Planning, Scheduling, ERP, Heuristics, Optimization, Local Search.

ABSTRACT

Modern companies usually use sophisticated Enterprise Resource Planning (ERP) systems for planning and scheduling their manufactory. ERP systems, thanks to the connection with Advanced Planning and Scheduling (APS), are solving several types of optimization problems such as basic scheduling problems, specific cutting problems or transport optimization.

This paper describes proceedings through solving Job Shop Scheduling Problem, which can improve scheduling in ERP/APS and also shows some possibilities of available improvements including the local search procedure example.

INTRODUCTION

Nowadays, the Enterprise Resource Planning (ERP) systems, together with Advanced Planning and Scheduling (APS) modules, solve a wide range of problems. They are prepared for many types of modern trends in the planning area such as JIT (Just In Time) and KANBAN using tools which TOC (Theory Of Constraints), TQM (Total Quality Management), etc. offer us.

Some ERP systems are able to deal with specific optimization problems, such as transport (travelling salesman problem), or cutting material (Cutting problem, Bin packing problem). Usually, ERP systems are also prepared, or specialized, to solve different types of schedule problems, as Flow Shop, for example in the chemistry manufacture, or Job shop for job-order manufacture.

In the matter of the Job shop scheduling, ERP systems usually use one or more dispatching rules (constructive algorithms), which bring them advantages and also disadvantages. But very few of them use advanced schedule optimization, such as Local Search or meta-heuristic methods (Oliveira 2006).

The general idea of proceeding through optimization with heuristic methods is to start with "constructive algorithm", then use its solution with one kind of search neighbourhood optimization that could be a Local Search procedure, or a meta-heuristic algorithm (Oliveira 2006). Each of these methods has general advantages and disadvantages, as shown in Figure 1.

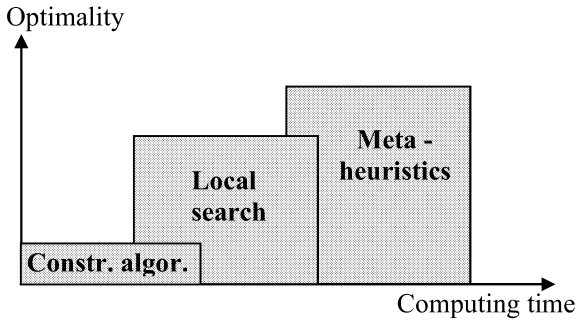


Figure 1: Optimality and computing time of heuristics methods

Constructive algorithms are very efficient, but their effectiveness is usually very poor. The Meta-heuristic, on the opposite side, performs very well, but it is much more time consuming. Generally ERP/APS systems are using dispatching rules that can be classified as Constructive Algorithm (CA). By using Local Search, which is a very simple algorithm, it is possible to improve the solution of CA without affecting the efficiency of ERP/APS (Oliveira 2006; Oliveira 2007).

CONSTRUCTIVE ALGORITHMS FOR SCHEDULING PROBLEMS

Constructive algorithms (CA) are used as a first approach to solve scheduling problems. CAs are easy to implement. They are perceptible and they are easy to explain at the shop floor, and for that their use is frequent, in spite of their acting. CAs start with an empty solution (plan) and using methods, as Greedy rules, they solve problems by constructing complete solutions (schedules), respecting constraints sets (technological order, material requirements, etc.), in a finite series of steps.

There are three main kinds of scheduling problems on the shop floor (French 1982; Pinedo 1995):

Open shop – Every operation is processed in one machine and the processing order of operations can be random (not constrained by technological order).

Flow shop – Every operation is processed in one machine in a defined processing order which is the same for every Job.

Job shop – Every operation is processed in one machine and the processing order is given by technological order. It is also possible to consider several completive constrains – setups, stocks, etc. This kind of problem can be solved by the Giffler and Thomson algorithm for active schedule generation, Non-delay schedules.

A general idea, in the use of CAs to solve scheduling problems, is to assign the values for all of decision variables, one by one, and at every step make the best available decision. The main problem is to find the way how to choose the best decision (set plan or rule) and it is obvious that it requires knowledge of the problem and its behaviour through computing.

Usually, in the real shop problems we use Non-Delay schedule generation that performs very well. It is also more acceptable when we justify a plan – machines are always filled with work whenever possible. We do not have additional work justifying "why is this machine idling", knowing that "overall plan will be finished earlier" is not a clear and understandable reason for many people. But we must know that this kind of plan, in some cases, could not find the optimal makespan (in the matter of minimizing overall finishing time).

By choosing priority rules we have many possibilities. For example, in the case of solving Job Shop by Active schedule generation, or Non-delay schedule generation, we have more than one hundred of different dispatching rules (Panwalkar and Iskander 1977). Each of these rules gives us, in the specific situations, optimal schedule but none of them can give us always the "optimal" solution.

Some known rules are:

SPT (shorter processing time) to finish small orders first;

MWKR (most work remaining) for the case that we whish not to focus on one job;

LWKR (least work remaining) when we like to finish job first to unblock a critical machine for other jobs;

FCFS (first come first served) to make a process of continuous flow;

Random – knowing that we have in the conflict set (using Active schedule generation) all "active operations" and we have no information that can support our decision to choose a rule. Random is the best in that instant.

Through computing can sometimes happen that operations can satisfy both plan generation (Active, Non delay) and dispatching rule. In that case we have several possibilities as selecting one randomly or the one with the highest user priority set up in ERP database and also by making hierarchy in the dispatching (priority) rules. For example, we can combine most usual rules like select first cases that satisfied MWKR and than SPT. Overall, we can use different rules and plans (active or non-delay) several times, thanks to constructive algorithms short computing times. However, we must know that we can get the same results with different schedules (Figure 2) as we get in the scheduling from model "Three mechanism" of "Virtual company" (Koblasa and Vavruška 2007).

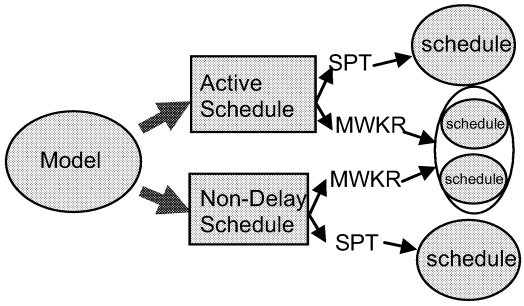


Figure 2: Same results with different "schemes"

Using both types of plan generation and several dispatching rules to get better plan can just seem like a time consuming matter (we generate several plans and we use only one). Anyway, if we would like to optimize, for example, with additional methods like Local Search, they can be very useful.

RANDOM WALK SEARCH

The Random Walk Search (RWS) method is a search that does not use knowledge about the problem. RWS is quite useful because it is very general and can be used in a large range of problems. We may wonder why we should use RWS, when we could use a search with some built in intelligence. The answer can be that there may not be any information we can use. We might just be looking for an answer and would not know we have found it until we see it (as optimal plan). However, it can be computationally quite expensive.

LOCAL SEARCH

Local Search is a method to improve the quality of the current solution of computationally hard optimization problems. It can be used on problems that can be formulated as finding a solution optimizing a criterion among a number of available solutions (schedules). Local Search algorithm moves from a solution to another solution in the search space; until that, there is no more improvements, or an acceptable solution is found, or a time bound is elapsed.

Results of this method depend mainly on selecting the searching space (neighbourhood) and on the initial solution

(schedule delivered from CA). If the neighbourhood is too small it gives us results very early, but it will probably be stuck in a poor local optima. On the other side, very large neighbourhood can give us very good results but only thanks to high costs due to computation time.

It is also possible to include knowledge about the problem in the Local Search methods. This strategy can improve both effectiveness (quality of optimum) and efficiency (computing time) of local search methods. Particularly, for the scheduling problem such kind of knowledge can be achieved from critical path in the Disjunctive graph.

The major improvement in Local Search comparing to random walk search is the particular knowledge about the problem. For some scheduling problems the local search procedure begins by getting critical path from current solution (schedule). Then any operations on the critical path are called critical operation. We can decompose the critical path into a number of blocks. Easily, the critical block is a set of activities on critical path scheduled on the same machine without idle time. For some objective functions only swapping operations that belong to the same block make sense.

Brucker (1994)

Brucker's branching scheme (Brucker et al. 1994) is based on swapping all around critical blocks searching for particular swap possibilities. In this swap neighbourhood all operations (except the first one) can be placed in first position and all operations (except the last one) can be placed in last position.

For example, with 4 operations in one block with [A B C D] order. Brucker's neighbourhood scheme gives us the following 6 neighbours:

First position – [B A C D], [C A B D], [D A B C];

Last position – [B C D A], [A C D B], [A B D C].

This neighbourhood is quite large and gives us acceptable results but it is more time consuming then e.g. Nowicki and Smutnicki, who only consider for this case [B A C D] and [A B D C].

Nowicki and Smutnicki (1996)

We also need, for the Nowicki and Smutnicki (1996) approach, to find critical path first. Then, if a job and a machine predecessor of a critical machine are both critical, we choose the predecessor which is first in the operation sequence. The neighbourhood of Nowicki and Smutnicki follows. There are given b blocks and l connections with blocks on another machine in the critical path. If $1<l< b$, then swap only the last two and first two block operations. Else, if $l=1$ swap only the (first) last two blocks of operations. If the swap improves schedule, it is accepted. Once a swap is accepted, the critical path may change and a new critical path must be identified.

This approach is more acceptable in the view of less computation time and also more justifiable knowing that TOC is very popular nowadays.

When we do not have possibility to compute critical path, we can extend the neighbourhood area just by including all blocks of operations to "critical path". It will extend searching space significantly in the case of bigger scheduling problem. At least we can make our neighbourhood smaller by "taking" only the first "acceptable" solution (first improvement strategy) instead of searching through the whole possible neighbourhood space (best improvement strategy). We must expect that it will definitely give us local optima or initial solution.

Single Swap Local Search

Single Swap Local Search is a simple example on how we can easily improve our solution. This Local Search example is not using any knowledge of the problem. We tested two schedule generations and three dispatching rules to generate initial solutions. The initial solution presented by job sequence on machine provides us constructive algorithm using:

i) Giffler and Thomson active schedule;

ii) Non delay schedule.

Together with dispatching rules SPT, MWKR and Random.

From this solution we search in the neighbourhood that is generated by single job sequence swapping through every machine (Figure 3).

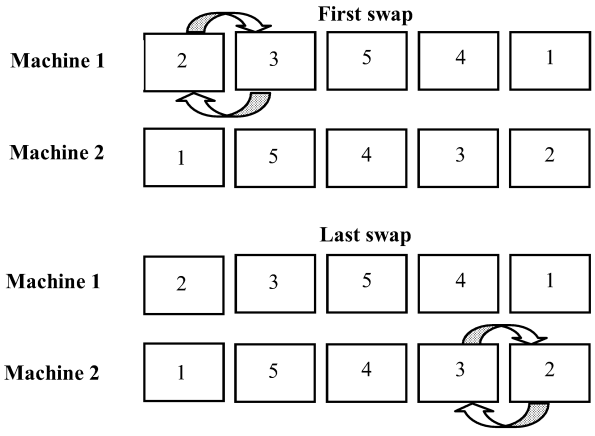


Figure 3: Single Swap Local Search procedure

This procedure is mostly generating unfeasible solutions, but also a few feasible ones. The solution that is better than the previous one is set as new initial, if it is found. This procedure continues until the best solution (makespan) in new neighbourhood is not better than the previous one.

This kind of Local Search procedure is not leading to optimal makespan, but it is improving schedules (Active schedule – Tables 1, 2 and Non delay schedule – Tables 3, 4), in most cases, to a better solution in short time. Single Swap Local Search procedure was tested on 10 well known instances of Job Shop Problem (the first number indicates number of jobs and the second indicates the number of machines):

i) H. Fisher, G.L. Thompson (1963): 6x6 (ft6); 10x10 (ft10); 20x5 (ft20).

ii) R.H. Storer, S.D. Wu, R. Vaccari (1992): 50x10 (sw11).

iii) S. Lawrence (1984): 10x5 (La2); 10x10 (La19); 15x10 (La21); 20x10 (La27); 20x10 (La30); 15x15 (La40).

Table 1: Active schedule generation – Constructive Algorithm

		Active schedule generation (CA)			
Problem	Optimum	SPT	MWKR	RAND (b)	RAND(10)
ft06	55	94	67	65	70
ft10	930	1429	1178	1287	1425
ft20	1165	1675	1498	1503	1623
sw11	2983	4682	4492	4163	4441
La2	655	1055	860	807	931
La19	842	1286	1084	1181	1280
La21	1046	1657	1267	1541	1629
La27	1235	2201	1567	1798	1928
La30	1355	2194	1565	1912	2062
La40	1222	1804	1549	1834	1954

Table 2: Active schedule generation – Local Search

		Active schedule (Local search)			
Problem	Optimum	SPT	MWKR	RAND (b)	RAND(10)
ft06	55	92	58	62	67
ft10	930	1293	1052	1219	1368
ft20	1165	1596	1475	1446	1549
sw11	2983	4585	4433	4098	4334
La2	655	914	829	747	806
La19	842	1182	1029	1160	1229
La21	1046	1581	1213	1466	1526
La27	1235	2146	1545	1657	1827
La30	1355	2140	1523	1873	1948
La40	1222	1681	1539	1691	1854

Table 3: Non-delay schedule generation – Constructive Algorithm

		Non-delay schedule generation (CA)			
Problem	Optimum	SPT	MWKR	RAND (b)	RAND(10)
ft06	55	88	62	66	69
ft10	930	1074	1108	1145	1250
ft20	1165	1265	1474	1446	1502
sw11	2983	3494	4342	4070	4247
La2	655	821	817	806	854
La19	842	913	1013	1005	1034
La21	1046	1324	1256	1283	1375
La27	1235	1784	1442	1603	1765
La30	1355	1792	1534	1716	1802
La40	1222	1476	1440	1499	1663

Table 4: Non-delay schedule generations – Local Search

		Non-delay schedule (Local search)			
Problem	Optimum	SPT	MWKR	RAND (b)	RAND(10)
ft06	55	72	60	64	66
ft10	930	1017	1108	1100	1213
ft20	1165	1265	1474	1434	1466
sw11	2983	3494	4298	4009	4178
La2	655	821	813	688	785
La19	842	909	1008	947	989
La21	1046	1257	1213	1220	1319
La27	1235	1705	1442	1520	1662
La30	1355	1666	1534	1567	1701
La40	1222	1377	1405	1489	1578

Ten schedules were created with random dispatching rule for both active and non-delay schedules, which were used as initial solution for Local Search. Mark (b) indicates the best result for the ten generated schedules.

The job shop optimal makespan is an active schedule but not necessary a non-delay schedule. Giffler and Thomson's algorithm, with the aid of local search or meta-heuristic procedure, can improve solution inside active schedules. Non delay schedule generation usually gives us better makespan then Active schedule generation in described instances. Our test also showed that Random dispatching rule is sometimes better then SPT or MWKR.

Single Swap Local Search improves the solutions in both Active and Non-delay schedule generation. The improvement percentage on this kind of procedure was, on average, for the case of Active schedule, approximately 5 % (Table 5) and 4% on Non-delay schedule (Table 6).

Table 5: Active schedule Local search improvement

	Active schedule solution improvement			
Problem	SPT	MWKR	RAND (b)	RAND(10)
ft06	2,13%	13,43%	4,62%	4,29%
ft10	9,52%	10,70%	5,28%	4,00%
ft20	4,72%	1,54%	3,79%	4,56%
sw11	2,07%	1,31%	1,56%	2,41%
La2	13,36%	3,60%	7,43%	13,43%
La19	8,09%	5,07%	1,78%	3,98%
La21	4,59%	4,26%	4,87%	6,32%
La27	2,50%	1,40%	7,84%	5,24%
La30	2,46%	2,68%	2,04%	5,51%
La40	6,82%	0,65%	7,80%	5,10%
Rule avg.	5,6%	4,5%	4,7%	5,5%
Average	5,1%			

Tables 5 and 6 show that the improvement of Active schedules was better than the improvement of ND. The Tables also show that in seven Non-delay schedules there was no improvement. It was caused by closeness of the initial solution to local optima.

Nevertheless, single swap Local Search, which is a very simple algorithm, shows that it is possible to improve solutions on average 4% in a very short time.

Table 6: Non-delay Local search improvement

	Non-delay schedule solution improvement			
Problem	SPT	MWKR	RAND (b)	RAND(10)
ft06	18,18%	3,23%	3,03%	4,35%
ft10	5,31%	0,00%	3,93%	2,96%
ft20	0,00%	0,00%	0,83%	2,40%
sw11	0,00%	1,01%	1,50%	1,62%
La2	0,00%	0,49%	14,64%	8,08%
La19	0,44%	0,49%	5,77%	4,35%
La21	5,06%	3,42%	4,91%	4,07%
La27	4,43%	0,00%	5,18%	5,84%
La30	7,03%	0,00%	8,68%	5,63%
La40	6,71%	2,43%	0,67%	5,10%
Rule avg.	4,7%	1,1%	4,9%	4,4%
Average	3,8%			

CONCLUSION

Today's research in the heuristic methods has very small impact on processes in the ERP (APS) systems. Common ERPs usually use one kind of plan based on MRP with one or few dispatching (priority) rules. Some of them have a simulation module where they advance their solution to reflect constraints (material, capacity, human resources).

Very few of them are using heuristics for planning and scheduling. The major problem is that advanced heuristics (including genetic algorithms) are more time consuming than basic constructive algorithms and cannot always guarantee the optimal solution. Time is not something today's companies like to spend. They usually need some feasible solution immediately. Knowing that today's heuristics are suitable for short term (operative) planning, where ranges of problems are much smaller than in medium or long term planning, we can use fast and simple heuristics.

Further research will focus on genetic algorithm based on Random Keys representation applied on Job-shop problem. The work will proceed with heuristic supported by knowledge of Job Shop Scheduling Problem (e.g. critical path), and we will study the connection of these heuristic methods with simulation software.

REFERENCES

Brucker, P.; B. Jurisch; and B. Sievers. 1994. "A branch and bound algorithm for the job-shop scheduling problem". *Discrete Appl. Math.* 49, 1-3, 107-127.

Fisher H. and G.L. Thompson. 1963. "Probabilistic learning combinations of local job-shop scheduling rules", J.F. Muth, G.L. Thompson (eds.), *Industrial Scheduling*, Prentice Hall, Englewood Cliffs, New Jersey, 225-251.

French, S. 1982. *Sequencing and Scheduling: An Introduction to the Mathematics of the Job-shop*. Ellis Horwood limited, West Sussex, England, ISBN 0-470-272

Koblasa, F. and J. Vavruška. 2007. "Educational Virtual Company at KVS". In *Konference WITNESS 2007*. Conference proceedings HUMUSOFT s.r.o. & VUT Brno FP.

Lawrence, S. 1984. "Resource constrained project scheduling: an experimental investigation of heuristic scheduling techniques" (Supplement), Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

Nowicki, E. and C. Smutnicki. 1996. "A fast tabu search algorithm for job shop problem". *Manag. Sci.*, vol. 42, pp. 797–813.

Storer, R.H.; S.D. Wu; and R. Vaccari. 1992. "New search spaces for sequencing instances with application to job shop scheduling". *Management Science* 38, 1495-1509.

Panwalkar, S.S. and W. Iskander. 1977. "A survey of scheduling rules". *Operations Research*, Vol. 25, No.1, 45-60.

Pinedo, M. 1995. *Scheduling - Theory, algorithms and systems*. Prentice Hall,Inc. ISBN 0-13-706757-7.

Oliveira, J.A. 2006. "A Genetic Algorithm with a Quasi-local Search for the Job Shop Problem with Recirculation". *Applied Soft Computing Technologies: The Challenge of Complexity*, Berlin / Heidelberg, Springer, 221-234.

Oliveira, J.A. 2007. "Scheduling the truckload operations in automatic warehouses". *European Journal of Operational Research*, 179, 3, 723-735.

BIOGRAPHY

FRANTISEK KOBLASA was born 1980 in Prague, Czech Republic and studied on Technical University in Liberec where he studied Manufacturing systems and obtained his master degree in 2005. He is studying PhD Manufacturing systems and processes program and he is enrolled in an ERASMUS-SOCRATES program at University of Minho - Department of Production and systems..

LUÍS S. DIAS was born in Vila Nova de Foz Coa, Portugal He graduated in Computer Science and Systems Engineering in the University of Minho, Portugal. He holds a PhD degree in Production and Systems Engineering from the University of Minho, Portugal. His main research interests are Operational Research and Simulation.

JOSÉ A. OLIVEIRA was born 1966 in Matosinhos, Portugal. He studied Mechanical Engineering at the University of Porto, Portugal. He graduated with a Ph.D. in Production and Systems Engineering at University of Minho, Portugal. His main research interests are Optimization with Heuristic Methods in Systems Engineering.

GUILHERME PEREIRA was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management in the University of Minho, Portugal. He holds an MSc degree in Operational Research and a PhD degree in Manufacturing and Mechanical Engineering from the University of Birmingham, UK. His main research interests are Operational Research and Simulation.

EMERGENCY RISK AND SAFETY CRITICAL SYSTEMS

The FLOREON System - Emergent Flood Prediction

Ivo VONDRÁK
Jan MARTINOVIČ
Jan KOŽUSZNIK
Svatopluk ŠTOLFA

Faculty of Electrical Engineering
and Computer Science
VŠB - Technick Univerzita Ostrava
17. listopadu 15, 708 33
Ostrava-Poruba

Jan UNUCKA

Faculty of Mining and Geology
VŠB - Technick Univerzita Ostrava
17. listopadu 15, 708 33
Ostrava-Poruba

KEYWORDS
Flood, Disasters Management, Modeling

ABSTRACT

The main goal of our system is to provide the end user with information about an approaching disaster. The concept is to ensure information access to adequate data for all potential users, including citizens, local mayors, governments, and specialists, within one system. It is obvious that there is a knowledge gap between the lay user and specialist. Therefore, the system must be able to provide this information in a simple format for the less informed user while providing more complete information with computation adjustment and parameterization options to more qualified users.

One system feature in high demand is the ability to display reliable and understandable graphical and textual information. Information for various types of users must be adapted to a desired format which is understandable to a particular group of people. For example, a specialist can ask for all available results from different simulation models in text format. This type of information may be useless, however, to the user who only wants to find out whether or not his house will be flooded.

INTRODUCTION

There are many types of natural disasters in the world. Many of which depend specifically upon geography. The flood is one of the worst and most recurrent types of natural disasters in our region [3, 4, 18]. Local governments require reliable models for flood simulations and predictions to save on ample funding that must be otherwise invested in post-flood repairs for impacted regions [4]. Therefore, the issue of flood prediction and simulation has been selected as a case choice for experimental development. This sophisticated system should combine all knowledge from existing models and systems while providing results to a wide range of people that are not experts in this particular field. The name of the system that is being developed is FLOREON (FLOods REcognition on the Net).

A wide range of professionals in various fields are already familiar with this system concept. Systems that are able to model floods and present them to qualified system users already exist. The concept of using a single, preferred and utilized software (or the combination of several types) plays a major role regarding the limited experience and viewpoints of some specialists [3, 18]. Furthermore, strict deadlines, resulting from the demand for promptly obtained modeling results to support timely decision making, present another significant issue. [3]. In addition, the individual, qualified user is not able to operate more models and assemble predictions at the same time. For these reasons, the goal of this system is to combine the results of all prediction and simulation systems and provide simulation results to people who are less informed concerning the complexity of this issue. These results should be accessible to a wide range of users including professionals from various fields. Inhabitants of flood plains should be able to view the results directly in a simple way while more complex information should be accessible to local representatives and rescue teams. Legal bodies responsible for entire regions (i.e. local governments) should be able to view as much information as possible. Thus, the front end of the entire system must be able to satisfy all types of users. A logical solution, and genuinely efficient method of achieving this goal, would be to combine several visualization tools and equipment in the system. All users would be able to access the system from one website. Local representatives and rescue teams can also use other types of thin equipment such as a local visualization applications. Governments will be able to use more sophisticated thin or fat equipment, that will enable access and visualization of all types of information, based on their needs.

The main concern of our project is to prove that a combination of the latest information technology, along with results from fields that are in some way connected to creating flood simulations and predictions, would be very useful. Fields of interest considered detrimental to supplying theses results include hydrology, meteorology, geoinformatics, remote sensing, integration of prediction models, development of new models, and others [3, 18]. The information system itself is focused on the development of a reliable architecture and a graphic user interface (2D, 3D) that should satisfy the needs of all types of users expected to eventually interact with the system.

The technological integration of the necessary system module is provided by means of internet technology which enables dynamic configuration and the construction of an open architecture [4]. We base the communication of modules within the system and communication with the other systems on Web Services [19].

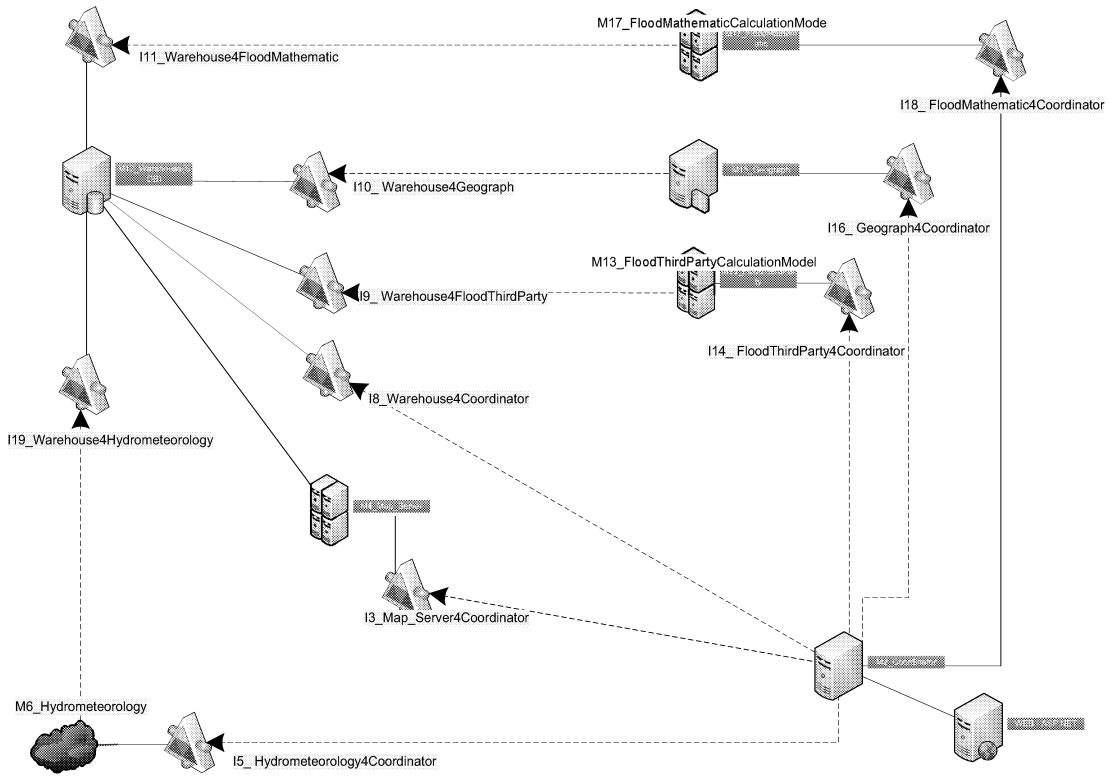


Fig. 1: System architecture

Although, the primary goal of this system is flood prediction, simulation and presentation, it is possible to expand this system to include modules that can represent a wide range of other disasters. Examples of some modules that may be added to the system include modules that simulate and predict air and water pollution, advective-dispersion processes, forest fire spreading, and etc…

The addition and removal of system modules (including alternative models for simulations and predictions) is based on the plug and play principle. All of these system futures should establish the base for a complex and unique system that enables users to tap into its potential for observing and controlling various types of disasters. Every type of model (rainfall-runoff, hydrodynamic etc…) can be replaced by other models. Reasons for considering such substations include the dropout of hydrometeorological data input for a particular model method (i.e. snowmelt module) and the added issue of method limitations during specific rainfall-runoff episodes – i.e SCS CN method performance differs from SMA method performance during rainfall-runoff events of varying duration and rainfall intensity.

ARCHITECTURE

The system architecture that has been developed for this project fulfills all desired features and enables any extension that might be added in the future. Moreover, our architecture includes an option for changing the technology that implements the architecture because the structure of the architecture itself is completely independent and does not rely on implementation (Fig 1.).

The main modules of the present system are the core, data warehouse, meteorological data module, hydrology module, geographic module, alternate numeric method module, visualization module, and web server module. The system is mainly based on a Microsoft .NET framework and written in C#. Other languages, such as C++ for special functions, are used when necessary.

All these modules must be connected and communicate in some way. Web Services seem to be the best technology currently available. Web Service technology ensures implementation independence, simple maintenance of connections, standardized communication, and simple module exchange. This technology enables the construction of a flexible and independent system.

Coordinator

A core module is responsible for the coordination of system scenarios and controls all modules that are plugged into the system.

Database Storage

A data warehouse stores all necessary data provided by a wide range of sources (geological data, meteorological data, etc…) or serves as a proxy to remote resources. The following technology is being used:

- PostgreSql database and its spatial extension PostGis
- Microsoft SQL Server
- Raster Provider - proprietary solution that will be soon published as an open source project [8].

Hydrometeorology

A meteorological module is a provider of meteorological data and weather predictions. We use a variety of sources:

- CHMU – Czech Hydrometeorological Institute provides measurement of hydrometeorological data (precipitation, temperature, water level stages, discharges, snow cover) obtained both from the station net and remote sensing methods (e.g. radar estimations of precipitation rates) [3]
- ALADIN – numerical weather forecasting system engaged by CHMU [1]
- Povodi Odry - company established by the government to provide measured discharge volumes in rivers and information about reservoir water levels and operations [15].

The hydrology module is mainly responsible for the computation of floods. More independent applications are used in our computation chain. These programs are connected to provide fully automated options. They involve the following applications:

- HEC-HMS (USACE) –Rainfall-runoff model
- MIKE 11 (DHI) – 1D hydrodynamic model
- MIKE 11 GIS (DHI) – GIS post-processing software, which inspects and provides measures of the flood occurrence and extent in the terrain
- ArcGIS (ESRI) – another platform for the post-processing of results

Additional information about the involved models is available on the internet or in [3].
An alternative, numeric method module is an experimental module that encapsulates new flood prediction methods developed by the group of mathematicians that we are cooperating on this project with. There is a 1D numerical approach as well as a more sophisticated 2D approach. The 2D numerical approach is based on the Finite volume method [2].

Geography and Visualization

The geographical module is responsible for the provision of all actual terrain data, maps, etc... It is a group of standalone tools used for data pre and post-processing:

- SharpMap for shapefile parsing [11]
- Quantum GIS [9] for data import and management

This module also combines all the information stored and created by the system and prepares this information for user presentation. This module uses 2D and 3D techniques to display results.

Web User Interface

A web server module communicates with users and transfers the visualization data to the user. The user can be equipped with any type of device. This module adapts the information to the user's device type.

CONCLUSION

We have already implemented the two scenarios described above. The system fully computes and predicts floods automatically and periodically. Results are accessible on the Internet with any modern browser. The user only needs to select a specific area and input the date of a given situation. The date can be set from the past or for the future. The system returns pre-computed and stored results.
In our opinion, and based on the response of external subjects specializing in the field of flood prediction or the coordination of emergency units, this project is thoroughly useful. This system can be easily extended or connected to other systems and establish a portal for disaster control, management, and presentation. This project seems to be innovative in its unique connection and coordination of many results from many different disciplines.
The architecture of this system is completely open and general. Therefore, all future features can be easily added to this system without negatively affecting its present functionality. Even the implementation of the architecture itself can be rebuilt without significant malfunctions in the entire system.

REFERENCES

[1] ALADIN - Aire Limitée, Adaptation Dynamique, Development International, http://www.chmi.cz/meteo/ov/aladin/index.php
[2] Barth T, Ohlberger M, Finite volume methods: Foundation and analysis. Volume 1, chapter 15 in Encyclopedia of computational Mechanics, editors: Stein, E.; Borst, R., Hughes, T. J. R. West sussex, England: John Wiley & Sons http://citeseer.ist.psu.edu/barth04finite.html
[3] BEDIENT, P.B., HUBER, W.C. et VIEUX, B.C. (2007): Hydrology and floodplain analysis. 4th edition. Prentice Hall, London, 795 s., ISBN: 978-0131745896
[4] BRÁZDIL, R. et al. (2005): Historické a současné povodně v České republice. Brno-Praha. MU Brno a ČHMÚ, 369 p. ISBN 80-210-3864-0
[5] Czech hydrometeorological institute, http://www.chmi.cz/indexe.html
[6] Günter Oliver, Data Management in Enviromental Information Systems, Handbook of massive data sets, Dortrecht, The Netherlands : Kluwer academic publishers, 2007, 981 – 1091, ISBN 1-4020-0489-3
[7] Institute of Computer Science, Academy of Sciences of the Czech Republic, http://www.cs.cas.cz/
[8] MEDARD, http://www.medard-online.cz/
[9] QGIS – Quantum GIS, http://www.qgis.org/
[10] Raster provider project, http://www.codeplex.com/rasterprovider
[11]SharpMap project – Open source mapping engine for .NET 2.0, http://www.codeplex.com/SharpMap
[12] URL: http://www.cs.vsb.cz/en/index.php (2007-06-30)
[13] URL: http://gis.vsb.cz/gisen2006/ (2007-06-30)
[14] URL: http://www.hgf.vsb.cz/hgf/eng/kat541_eng .html (2007-06-30)
[15] URL: http://www.am.vsb.cz/index.php?lang=en (2007-06-30)
[16] URL: http://www.zcu.cz/index-en.html (2007-06-30)
[17] URL: http://www.pod.cz/ (2007-06-30)
[18] WOHL, E. W. ED. (2000): Inland Flood Hazards. Cambridge,Cambridge University Press. 498 p. ISBN 0-521-62419-3,.
[19] Damien Foggon, Dan Maharry, Chris Ullman: Programming Microsoft .NET XML Web Services (Pro-Developer) (Hardcover), Microsoft Press

Defining project risk management efficiency through simulation

Konstantinos Kirytopoulos
University of the Aegean
Department of Financial and Management Engineering
Fostini 31
Chios, Greece

Viktor Diamantas
National Technical University of Athens
Department of Mechanical Engineering
Iroon Polytechniou 9
Athens, Greece

Vrassidas Leopoulos
National Technical University of Athens
Department of Mechanical Engineering
Iroon Polytechniou 9
Athens, Greece

E-mail: kkir@central.ntua.gr, vdiamant@mail.ntua.gr, vleo@central.ntua.gr

KEYWORDS
Project Management, Risk Management, Monte Carlo Simulation, Risk Management Evaluation.

ABSTRACT

Project risk management efficiency is an important factor for project success. However, the traditional project control techniques do not incorporate project risk management evaluation. Such a technique, addressing the project objective of time, is proposed in this paper, based on the use of stochastic S-curves (SS-curves) and Monte Carlo simulation. Stochastic S-curves provide probability distributions for the time values required to complete the project at every selected report point. Additionally, a new evaluation graph, based in quality control graphs, is presented in order to provide a clear view of the efficiency of the project risk management process to the project manager.

1. INTRODUCTION

Project management success is the appreciation of the project management results by the relevant interested parties (Caupin et al., 2006). Although successful implementation of projects is a vital issue for project-based enterprises, managers usually confuse the success of a project with the project management success. These two instances of success despite being higly related they are not the same as project success might be related to programme objectives that might change from external factors, while project management success is closely related to the objectives assigned to the project manager and have to be met.

Project management success is usually measured against the four project objectives namely cost, time, scope and quality (Kerzner, 2003). However, it is the risk factor the one which mostly determines the success or failure of a project (Johnson, 2006). Project risk has been defined *as an uncertain event or condition that, if it occurs, has a positive or a negative effect on at least one project objective, such as time, cost, scope, or quality. A risk may have one or more causes and, if it occurs, one or more impacts* (PMI 2004, p. 238). Project risk management consists of four basic phases, which are risk identification, risk assessment, risk mitigation and follow up (Kirytopoulos et al, 2001). This paper proposes a method that focuses on evaluating project risk management with respect to time management.

Project time management includes the processes required to accomplish timely completion of the project (PMI, 2004 p. 123). The core processes of project time management include activity definition, activity sequencing, activity resource estimating, activity duration estimating, schedule development and schedule control. The outcome of project time management is a project schedule that includes start and finish dates of each activity. Even if no risks appear, the project is unlikely to finish on schedule if either the duration of activities or the definition of dependencies is not realistic (Diamantas et al., 2006).

Since the scope of this paper is relevant to time-risk management evaluation, the time management process that is of most interest is schedule control. According to best practices (PMI, 2005) earned value management (EVM) is one of the most efficient and frequently used techniques for integrated monitoring of project performance.

Inspired by the popularity and efficiency of EVM, previous research in the area of implementing simulation within the EVM concept, as well as a proposed synergy between EVM and risk management, the proposed technique combines EVM (S-curves), simulation and risk management in order to monitor and evaluate the efficiency of the risk management process.

The remainder of this paper is structured as follows. Section 2, provides a brief description of the standard EVM technique, while the stochastic S-Curve concept is introduced in Section 3. The proposed technique for the evaluation of the risk management process is defined, in detail, in Section 4. Section 5 provides conclusions regarding the proposed technique.

2. Earned Value Management

Liberatore et al. (2001) conducted a survey that showed that 53% of the construction respondents (working with a firm or organization in the construction industry) used EVM for control.

The EVM concept was first introduced in the late 60s when the US Department of Defense released its first official list of "Cost/Schedule Control Systems Criteria" C/SCSC The C/SCSC constituted thirty-five standards of compliance (called criteria) that were required from management control systems within private industry, to assure consistency in the reporting of performance on major United States government acquisition projects (Fleming and Koppelman, 2005).
The Practice Standard for Earned Value Management (PMI, 2005) notes that the key practices of EVM include establishing a performance measurement baseline and measuring and analyzing performance against that baseline. The three basic EVM elements are Planned Value, Earned Value and Actual Cost:

- Planned Value (PV) is the authorized budget assigned to the scheduled work to be accomplished.
- Earned Value (EV) is the value of work performed expressed in terms of the budget assigned to that work.
- Actual Cost (AC) is the total cost actually incurred and recorded in accomplishing the work performed.

The graphical representation of Planned Value, Actual Cost, and Earned Value, as they accumulate over time, resembles the form of an S-curve (see **Fig. 1**). In the earlier stages as well as in the latter stages of construction project execution, less work is performed and less cost is accumulated than in the middle stages as a function of time.

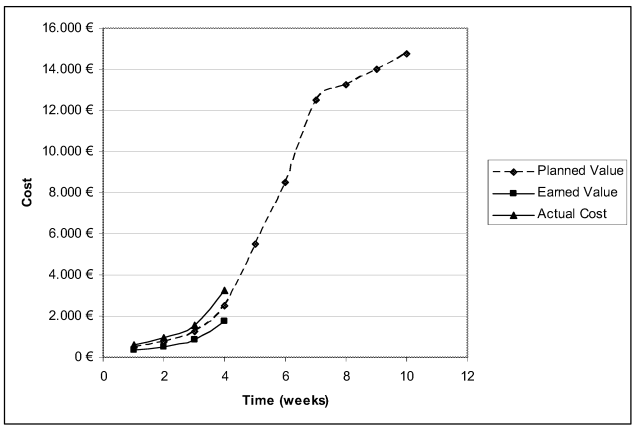


Fig. 1: Earned Value S-Curves

Based on the three basic elements, EVM calculates variances, indices and forecasts in order to evaluate the progress of a project. The variances include Schedule Variance (SV), Cost Variance (CV), and the indices include Schedule Performance Index (SPI) and the Cost Performance Index (CPI). The forecasts include, among others, Time Estimate at Completion – EAC_t and Estimate at Completion - EAC. The SV, SPI and EAC_t are used for schedule analysis while CV, CPI, EAC and ETC are used for cost analysis.

The SV is a measure of schedule performance in a project. It is calculated by subtracting the Planned Value (PV) from the Earned Value (EV). A positive value indicates a favourable condition, which means that the project is ahead of schedule while a negative value indicates an unfavourable condition, which means that the project is behind schedule. The Schedule Variance can be also expressed as a percentage – SV% by dividing the SV by the PV (PMI 2005).

The SPI is a ratio expressed measure of schedule efficiency in a project. It is the ratio of Earned Value (EV) to Planned Value (PV). An SPI equal to or greater than one indicates a favourable condition, and a value of less than one indicates an unfavourable condition.

Detailed description of CV, CPI and other cost related EVM indices is considered beyond the scope of this paper, since it focuses on the time aspect of a project.

EVM however, does not incorporate the stochastic nature of project activities. Several concepts of probabilistic project performance monitoring have been proposed (Barraza et al., 2000; 2004; Nassar et al., 2005) but still, none of the proposed concepts addresses the issue of project risk management evaluation.

3. Progress-Based S-curves and Stochastic S-curves

In EVM time is treated as an independent variable and cost and progress are treated as time-dependent variables, thus the EVM S-curves can be considered as time-based S-curves (TB-S-curves).

In order to include a probabilistic approach in the EVM concept, one has to use simulation (in this study Monte Carlo simulation) so as to come up with the possible outcomes of project performance. This possible outcomes are depicted at the Planned value which is now transformed to an area rather then a simple line (deterministic approach). An instance of this concept is illustrated in **Fig. 2.**

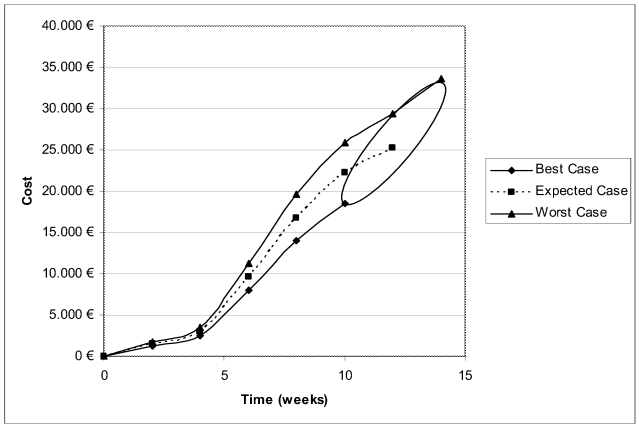


Fig. 2: Budget distribution in standard EVM S-curves (Hillson, 2005)

The results of the Monte Carlo simulation allow the best case project outcome to be determined and similarly a worst case profile can be produced. All other possible outcomes may also be calculated, allowing the "expected outcome" within this range to be identified. These can be shown as a set of three related S-curves, as in **Fig. 2**,

which take into account both the estimating embedded uncertainty (variability in planned events) and discrete risks (both positive opportunities and negative threats). The ellipse at the end of the curves represents all possible calculated project outcomes (90% confidence limit), with the top-right value showing worst-case (highest cost, later delivery time), the bottom-left giving best-case (cheapest and quickest), and the centre of gravity of the ellipse being at the expected outcome of project cost and duration. This ellipse is known by risk practitioners as the "eyeball plot" (Hillson, 2005).

Having constructed the budget distribution EVM S-curve (**Fig. 2**), Hillson (2005) recommends that the central S-curve in **Fig. 2** should be used as the baseline spend profile for EVM. He argues that "*this ensures that the EVM baseline fully reflects the risk associated with the project plan (including an appropriate amount for contingency which is automatically incorporated in the risk analysis results), rather than measuring performance against the raw 'all-goes-to-plan' plan*".

Barraza et al. (2000) presented an alternative chart where the earliest delivery time corresponds to the highest cost – case 1 – and the latest delivery time corresponds to the lowest cost – case 2 (**Fig. 3**).

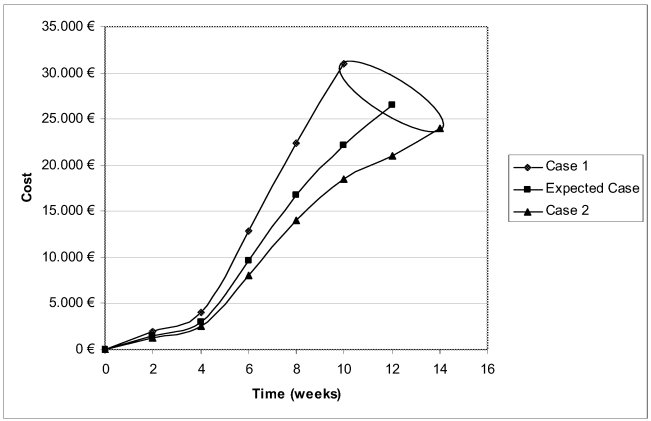


Fig. 3: Budget distribution in standard EVM S-Curves (Barraza et al., 2000)

Additionally, Barraza et al. (2000) identified a problem on this method. The problem is that *although* determination *of the planned duration distribution is possible for a specific project progress,* determination *of the possible correspondent budget distribution is quite difficult because of the overlapping budget envelopes for different project progresses*. For this reason, they proposed a new graphical representation (**Fig. 4)**.

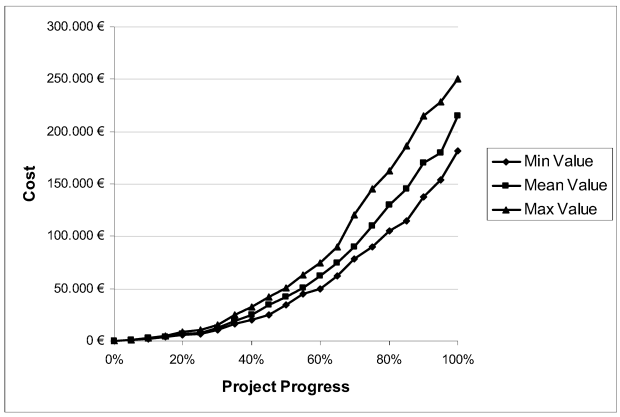


Fig. 4: Typical SS-curves

The new representation is based on the use of the planned percentage of accomplished work (progress) as the independent variable, because it depends only on the project scope, and project time and cost could be treated as progress-dependent variables. Progress-based S curves are, therefore, defined as plots of cumulative budget and planned duration against project progress.

4. Project Risk Management Evaluation Using Stochastic S-Curves

Standard EVM, as well as the approach proposed by Barraza et al. (2000; 2004), compare actual values to planned values in order to reach conclusions regarding the performance of the project.

In real projects the work performed is not always the same with the work planned. This fact would render problems in the evaluation of the risk management process. So, in order to evaluate the risk management process the actual values must be compared with the planned values for the actual work performed, that is the earned value in EVM terminology.

Inspired by Barraza et al. (2000; 2004), as well as Hillson (2004) a technique for the evaluation of risk management efficiency was developed. The proposed technique is based on the assumption that if the project risk management process is efficient the actual duration will fall in between a minimum and a maximum duration corresponding to a specific confidence interval of the duration distribution defined by the simulation. The confidence interval depends on users risk attitude and is defined on a project-wide basis. For instance, a company might consider project risk management efficient if the actual duration corresponds to a percentile between 20% and 80%, while another might narrow or extend this interval.

The proposed technique consists of three steps, which are repeated at each report point. The **first step** is the performance of the Monte Carlo simulation based on the actual worked performed. Input for the simulation is a project report spreadsheet that contains the progress and the initial stochastic estimations of duration for each project activity. The actual duration of the project is equal to the time elapsed between the initiation of the project

and the report point. An illustrative spreadsheet is presented in **Table 1**.

Table 1: Project progress report spreadsheet

Report Period	1	
Elapsed Duration	ED_1	
Activity ID	**Progress (%)**	**Time Distribution**
1	wp_1	td_1
2	wp_2	td_2
…	…	…
i	tp_i	td_i
…	…	…
n	wp_n	td_n

Keeping the priority constraints of projct activities, the outcome of this step is the calculation of the expected duration distribution for the actual work performed, which corresponds to the earned value SS-curves.

The **second step** is plotting the SS curves. It is noted that in this technique the SS-curves' points are being calculated in parallel to project reports and not from the start of the project since they depend on the actual work performed by each reporting period. Hence, the complete SS-curves are available at the completion of the final report of the project. An illustrative SS-curve graph for two reporting periods is shown in **Fig. 5**.

It is quite difficult to compare the duration to the corresponding distributions in the typical SS-curves graph. Therefore, a new graphical representation is suggested in the third step of the proposed technique.

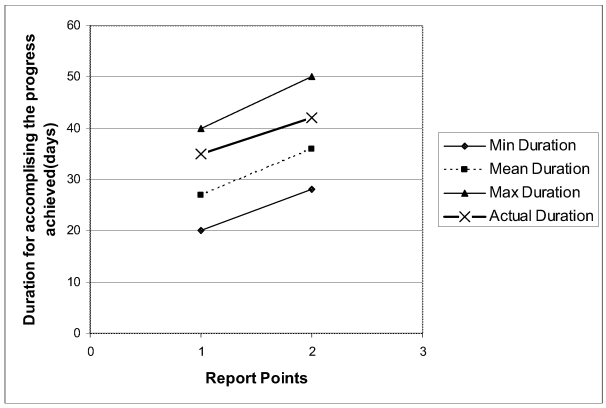


Fig. 5: Illustrative SS-Curves Graph

The **third step** is the plotting of a new graph that would allow a more clear view of the risk management efficiency. In the proposed graph the input data are the corresponding percentiles of the actual duration values in the calculated distribution at each report point, instead of the actual values themselves. The concept of the graph is based on quality control theory and the X bar R graphs.
The vertical axis of the proposed graph (**Fig. 6)** corresponds to the percentiles of the duration distributions instead of their numerical values, while the horizontal axis corresponds to the report points.

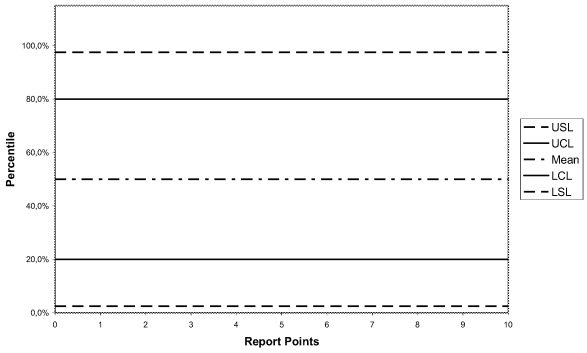


Fig. 6: Risk Management Evaluation

The proposed graph uses the concept of upper and lower specification and control limits. These limits are user defined depending on the risk management attitude of each user. For example, the lower and upper specification limits may be set to 2.5% and 97.5% respectively, and the lower and upper control limits may be set to 20% and 80% respectively. A user that desires a more strict control of the risk management process may narrow the area between the lower and upper control limit. An example of such a graph is shown in **Fig. 6**.

Although the assumptions of the original control graphs (X bar R graphs) might not be usable since the duration distributions are not normal (Kerzner, 2003), the trends of the input data are a clear indication of the efficiency of the risk management process.

When input data (corresponding percentage of actual duration) stay in one of the three areas of the chart for a number of reporting periods this could be seen as a clear indication of the efficiency of the risk management process, good or bad depending on the side of the diagram data resides.

If the data points are ascending or descending for a number of reporting periods then probably risk management process is getting better (descending points moving to mean value) or worsening (ascending points moving to mean value).

In the ideal situation the input data should fall within the envelope defined by the lower and upper control limits. Deviation from this would indicate potential problems within the risk management process.

It should also be noted that this graph provides a view regarding the actual performance of the project itself. If the mean value is used as a base for project evaluation, data above the centre line indicate that the progress is underperforming, while data below the centre line indicate over performance.

Additionally, this graph provides a clear view to the project manager by showing territories of 'caution' and 'danger' defined by firm bounds, instead of the ascending bounds of standard SS-curve graphs. An illustrative graph is presented in **Fig. 7**.

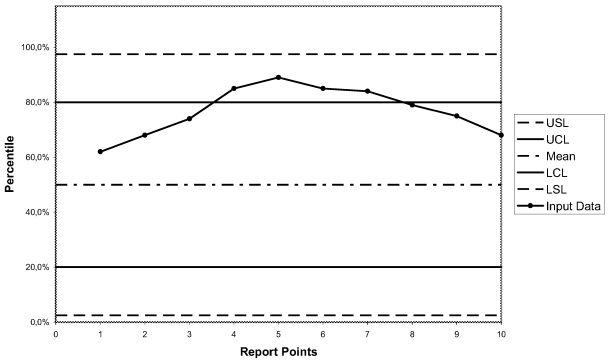


Fig. 7: Illustrative risk management evaluation graph

The deriving assumptions from **Fig. 7** are three. The risk management process was efficient in the beginning until the middle of the project. At this point the risk management process falters, and the actual duration corresponding percentile is outside the safe area. However, it is not beyond the upper specification limit, which would indicate a critical inefficiency. Due to actions taken by the project management the project falls back in the safe zone during the last reporting periods.

5. Summary and Discussion

Probabilistic planning and forecasting methods have been demonstrated to be more accurate than deterministic methods. However, project control is primarily conducted by deterministic approaches. Moreover, these approaches do not take into account the efficiency of the risk management process, which is vital for successful implementation of projects.

A new technique for evaluation of risk management, regarding the time objective of projects, has been suggested in this paper. The technique has been inspired by the SS-curves concept introduced by Barraza et al. (2000; 2004), Hillson's proposed synergy between EVM and risk management (2004), and the standard quality control charts.

The same technique can be used for evaluating risk management regarding cost, and thus an integrated evaluation of risk management can be achieved.

The main advancement to the original SS-curves techniques approach introduced by Barraza et al. (2000; 2004) is that in the technique suggested in this paper the whole distributions are used instead of using only the mean value.

Additionally, a major benefit of the use of S-curves, in both approaches, is the fact that both project management practitioners and researchers are familiar with S-curves, and thus, they are able to easily implement and further expand such techniques, respectively.

Implementation of the proposed technique in real case studies in order to verify and validate it is already under way. The results will be presented once a significant number of case studies have been monitored and safe conclusions have been derived. Another area for further research is the comparison of the performance of the proposed technique with other techniques dealing with the evaluation of risk management efficiency.

REFERENCES

Barraza G. A., Back E. W., and Mata F., 2000, "Probabilistic monitoring of project performance using SS-curves", *Journal of Construction Engineering and Management*, Vol. 126, No. 2, pp. 142-148.

Barraza G. A., Back E. W., and Mata F., 2004, "Probabilistic forecasting of project performance using stochastic S curves", *Journal of Construction Engineering and Management*, Vol. 130, No. 1, pp. 25-32.

Caupin, G., Knoepfel, H., Koch, G., Pannenbacker, K., Perez-Polo, F. and Seabury, C., 2006, *ICB – IPMA Competence Baseline version 3.0*, International Project Management Association, Nijkerk, The Netherlands.

Diamantas, V., Kirytopoulos, K. and Leopoulos V., 2006, "Simulation in Project Scheduling: Why PERT is not enough", *in Proceedings of the 3rd Future Business Technology Conference*, Athens, Greece, 17-19 April 2006, pp. 64-70.

Fleming Q.W and Koppelman J.M., 2005, *Earned Value Project Management* (3rd ed.), Project Management Institute, Newtown Square, PA, USA.

Hillson D., 2004, "Earned Value Management & Risk Management: A practical synergy", *in* Proceedings of the PMI Global Congress 2004 North America, Anaheim USA, 23-26 October 2004, CD-Rom proceedings.

Johnson J., 2006, My Life Is Failure: 100 Things You Should Know to be a Successful Project Leader, Standish Group International, West Yarmouth, MA, USA.

Kerzner H., 2003, *A Systems Approach to Planning, Scheduling, and Controlling (8th ed.)*. Wiley & Sons.

Kirytopoulos K., Diamantas V., and Leopoulos V., 2005, "Project Risk Identification Methods: Practices and Advices", *in Proceedings of the 17th Conference of the Hellenic Society of Operational Research*, Patra, Greece 16- 18 June 2005, Cd-rom proceedings.

Kirytopoulos K., Leopoulos V., and Malandrakis Ch., 2001, "Risk Management: A powerful tool for improving efficiency of project oriented SMEs", in *Proceedings of the fourth SMESME International Conference: Stimulating Manufacturing Excellence in Small & Medium Enterprises*, Aalborg Denmark 14-16 May 2001, pp.331-339.

Liberatore M. J., Pollack-Johnson B., and Smith C. A., 2001, "Project management in construction: software use and research directions." *Journal of Construction Engineering and Management*, Vol. 127, No. 2, pp. 101-107.

Nassar K. M., Gunnarsson H. G., and Hebab M Y., 2005, "Using Weibull Analysis for Evaluation of Cost and Schedule Performance", *Journal of Construction Engineering and Management*, Vol. 131, No. 12, pp. 1257-1262.

Project Management Institute, 2004, *A guide to the project management body of knowledge (PMBOK® Guide) (2004 ed.),* Project Management Institution, Newton Square, PA, USA.

Project Management Institute, 2005, *Practice Standard for Earned Value Management*, Project Management Institute, Newton Square, PA, USA.

ICT SECURITY MANAGEMENT AND RISK MANAGEMENT

Jeanne Schreurs, Rachel Moreau
Universiteit Hasselt
Agoralaan
3590 Diepenbeek.
Belgium.

jeanne.schreurs@uhasselt.be
rachel.moreau@uhasselt.be

KEYWORDS

Security, Risk

ABSTRACT

It's critical to know what are the critical resources and processes of the company and their weaknesses. Security has become a management problem. A security audit can offer a handy solution. We have developed a method and a system. The system is called BEVA. BEVA includes a security audit measuring the security situation of the organisation in 38 security factors. Based on this audit, it delivers an overall security score and the one for each security factor. It also supports management in detecting the critical security factors in the company. The ICT security budget is not unlimited. As a result not all needed security technologies can be implemented to improve all critical SF's. To create a priority list of the corresponding critical threats we need to refer to organisational risks. We calculate the company specific risk scores and by the way we identify the most critical threats and by the way the most effective security technologies that must be implemented

1. SECURITY AUDIT

1.1. ICT security is a management issue

As a consequence of more openness to the outside world, companies become more vulnerable for security threats. Studies show that some attacks bring along large costs. That is why companies become more and more aware that ICT security is a management problem. (9),(12) Actually a company needs a totally security approach. It is a must to know what are the critical resources and processes of the company and their weaknesses. A solution to this is a security audit.

1.2. Security audit

We have developed a security audit, called BEVA. It is covering the 10 domains as set forward in the ISO17799 model. (3) BEVA is a method to analyse critically the company and to uncover the weak spots of the security system. It positions the company on point of the security aspects in the different areas of business functions. We have developed a standard list that covers all aspects of security, structured in 10 domains.

Each of these areas consists of different security factors. In total 38 security factors are spread over the 10 domains. For each of the 38 factors, a number of subcriteria are formulated. (4), (13), (14) We developed a list of questions, covering the subcriteria we created. The questions are partly based on the "checklists in information management" SDU publishers.

10 domains
- Security policy - Organization of information security - Asset management - Human resources security - Physical and environmental security - Communications and operations management - Access control - Information systems acquisition, development and maintenance - Information security incident management - Business continuity management

2. THE AUDIT PROCESS AND SECURITY DECISION MAKING

2.1. Our security audit checklist

We based our security analysis partly on the Marion-AP method. But our list for the security factors is based on the standard ISO 17799. The 38 security factors are spread over the 10 domains , as set forward in the standard ISO17799 model. For example you have the domain "system access control" and in this domain you have the factors: Sf20 requirements for access, Sf21 management of user access, Sf22 user responsibility, Sf23 control of network access, Sf24 control access to OS, Sf25 control of access to applications and information and Sf26 use of mobile infrastructure. (Figure 1)

We start with the questioning of the key persons in the company using the audit checklist questionnaire.
An importance rate is given to the security factors from A (low importance) to E (high importance) (figure 1). For each subfactor a relevance score from 1 to 4 has to be given and each question has to be evaluated by a score between 1 and 4. For each subfactor a relevance score from 1 to 4 has to be given and each question has to be evaluated by a score between 1 and 4. When the questionnaire is completed, BEVA now calculates the security factor scores Sfis. If all the factor security scores are calculated also an overall security score Ss is given;

2.2. The calculation of security factor scores Sfis and the overall security score Ss (Figure 2).

Security Factor Sfi	Importance	Sub Factor	Relevance/weight 1 to 4	Code question	Question	evaluation 1 to 4
Domain: Access control						
Sf20. Business requirements for access controlPremise	B	access control policymanagement	3	20.1	Is the access control policymanagement based on the business security requirements?	3
				20.2	Are aspects of logical and physical access control included?	3
				20.3	Is it clear for users and service providers which rules are applicable?	2
Sf21. User access management	C	registration of users	2	21.1	Is there any formal user registration and de-registration procedure for granting access to multi-user IS and services?	1
		privilege management	1	21.2	are privileges and allocated on need-to-use basis?	3
				21.3	are privileges only allocated after formal authorisation process?	1
		user password management	4	21.4	should the allocation and the reallocation of passwords be controlled through a formal management process?	3
				21.5	are the users asked to sign a statement to keep the password confidential?	1
		review of user access rights	3	21.6	does there exist a process to review user access rights at regular intervals?	4
					Does there exist a procedure to block the	

Figure 1: Questions audit checklist

security factor Sfi	Sub Security Factor	Relevance/weight 1 to 4 w (i,j)	Code	Evaluation 1 to 4	Mean evaluation per Sf 1 to 4 eval (i,j)	Security factor Security score Sfis
Security domain Access control						
1. Requirements for access control	Access control policy	3	2.1.1	2		
			2.1.2	3		
			2.1.3	2	2,33	2,33
		3				**2,33**
2. Management of User access	User registration	2	2.2.1	3	3,00	
	Privilege-	1	2.2.2	2		
			2.2.3	1		
			2.2.4	2		
	management		2.2.5	3	2,00	
	User password	4	2.2.6	4		
			2.2.7	1		
			2.2.8	2		
			2.2.9	3		
			2.2.10	4		
	management		2.2.11	3	2,83	
	Review of user access	3	2.2.12	2		
	rights		2.2.13	2	2,00	2,46
		10				2,46
Security factor score: $Sf_{is} = \Sigma [(w(i,j) * eval(i,j) / \Sigma w(i,j)]$						
Security score: $Ss = \Sigma_i \Sigma_j [w(i,j) * eval(i,j)] / \Sigma i \Sigma j\, w(i,j)$						

Fig. 2. Calculation of security scores

In BEVA, we express the state of security into scores of the security factor (Sfi's). We do this for all the factors and in the end we give a general security score (Ss) over all security factors. We based our security analysis partly on the Marion-AP method.

The management team evaluates the company for all aspects on a one to four scale and at the same time measures the importance or relevance of all subfactors.
For all security factors Sfi, I=1,38 one or more subfactors do exist: SSfij, j=1,k.
For each subfactor one or more questions are included in the audit questionnaire.
Eval(i,j) is the evaluation of the subfactor Ssfij and is the mean value of the evaluations for the individual questions, corresponding to that subfactor.
W(I,j) are the weights and corresponds to the measure of relevance for the subfactors.
The security factor security score Sfis for factor Sfi is the evaluation of the security of the security factor Sfi and is equal to the weighted mean of the evaluations of the subfactors:
Sfi s = Σj [eval (i,j) * w(i,j)] / Σ w(i,j)
The overall security score Ss requires the weighing over all Sf evaluations and using the total weight:
Ss= Σi Σj [w (i,j) *eval (i,j)] / ΣiΣj w(i,j)

For example for the security factor management of user access see figure 2; it has a security factor score of 2.46;
The security score Ss is in this example = 2.39
Figures 3 and 4 show the scores of all the security factors. The red line in figure 5 states Ss the overall security score. The blue line connects the individual scores of the security factors. In the graph of Figure 4 the factors with score weaker than the company mean or overall security score can be identified
Based on these reports management can decide to concentrate on those security factors. Indeed if the situation on those points can be improved, as a consequence the overall score will increase too, and so the overall security of the company.

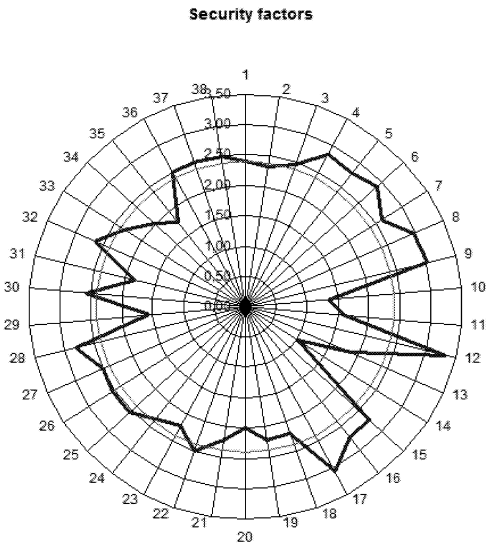


Fig. 3. Graph of the security scores

Security factor	weight	SfiS	Security factor	weight	Sfi
1	A	2,38	20	B	1,98
2	D	2,33	21	B	2,21
3	B	2,46	22	A	2,50
4	A	2,83	23	D	2,21
5	C	2,76	24	E	2,33
6	D	2,90	25	C	2,56
7	A	2,60	26	A	2,56
8	B	2,96	27	D	2,50
9	E	3,00	28	B	2,80
10	D	1,33	29	E	1,55
11	C	1,60	30	B	2,57
12	A	3,30	31	C	1,83
13	C	1,86	32	B	2,62
14	B	1,00	33	C	2,31
15	D	2,72	34	A	2,00
16	A	2,72	35	E	1,75
17	E	3,06	36	C	2,50
18	C	2,20	37	B	2,50
19	B	2,20	38	C	2,50

Fig. 4. Table of security scores for the security factors

2.3. Critical security factors and security decision making

Management has not only evaluated the security but it has also decided about the importance to them of the security factors or the security domains included in the audit.
As a consequence we can use those measures to differentiate between the weak factors as resulting from the previous analysis. We are also creating a report that combines the scores of the security factor with its importance. The differences are well stressed in this graphic and immediate points of action can be stated. (Figure 5)
We identify 4 regions and the Sfi can be found in one of them:

Red	Critical, improvement is urgent
Green	Continuation
Orange	No action/ no concern
Yellow	Not important, but follow up

Figure 5 highlights which security factors are critical, and for which improvement is needed. Critical means being evaluated low and being important. This feedback is given to the company management. Based on it action points can be developed. The goal is to increase the overall security score Ss by improving some low valued individual Sf security score. First we calculate a loss score for all threats. After adjusting the prior probability of occurrence of the threats, we calculate a risk score.

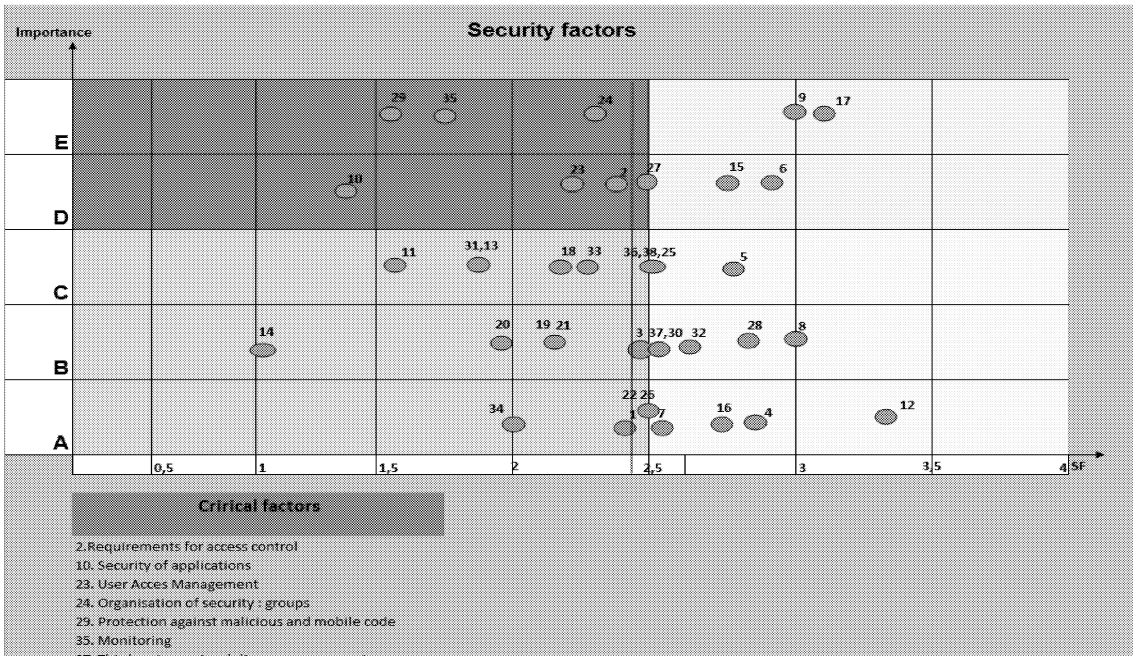


Fig. 5. Graph of security factors and their importance

3. ICT SECURITY DECISION MAKING AND ORGANISATIONAL RISKS

The ICT security budget is not unlimited. As a result not all needed security technologies can be implemented to improve all critical SF's.
We developed an ICT security management system. We developed a new methodology for security management decision making. To improve the security situation we will implement security technologies to decrease the probability of occurrence of the critical threats. (10), (11) first we must identify the critical threats corresponding to the critical security factors. And because not everything can be solved immediately due to a limited budget, we need some procedure to create a list of threats in descending order of probability of occurrence.

3.1. Losses due to the occurrence of threats: loss scores

Based on the table of total losses and on the occurrence percentage (used as the probabilities) of threats (cfr. CSI Survey-Annual Computer crime and security survey),
we can calculate a loss score by dividing the total loss by the occurrence percentage. We prefer to work with a loss score instead of the loss itself because first it is very difficult in practice to determine the losses resulting from the occurrence of threats and second we do not need the exact value of the loss, but rather a priority list. The yearly organised CSI/FBI-study delivers the following figures for the occurrence of the threats. Corresponding the exact losses we have calculated a loss score. Index value 100 is given to the highest loss due to "financial fraud";

3.2. Risks due to the occurrence of threats: risk scores

The occurrence level of the threats in the study can be interpreted as being the probability of the occurrence of the threats or the " mean" situation of a company.

Threat	Occurrence	Loss score
Financial fraud	0,12	100
Virus	0,52	9,17
System penetration	0,13	30,04
Theft of confidential data, from all causes but mobile device theft	0,25	23,25
Laptop/Mobile theft	0,5	4,41
Insider abuse of net access	0,59	2,78
Denial of Service	0,25	6,56
Phishing	0,26	6,01
Illigal software applications on the system (bots, trojan horses,…)	0,21	7,76
Unauthorised access to information	0,25	2,37
Instant messaging abuse	0,25	0,46

Fig. 6. Loss score for the threats

Threats	Security factors				
	Sf1 …	Sf20: requi-rements for access	Sf21: mgt of user access	…	Sf38
T1	x	x			
…					x
T4		x	x		
T5	x				
…					

Fig. 7. Relations between Sf's and threats

In the analysis of the security situation of our company, we can interpret those probabilities as the prior probabilities. If the overall security score of the company is high, the probability of occurrence of threats will be relative low, as will be figured out in the adjusted and smaller values for the company specific probabilities that will be calculated. We created a table that shows the relation between the treats and the security factors.
To calculate the adjusted probabilities we have to relate the threats with the relevant security factors. Adjustments will be made using the ratio of the mean value of the related Sf's to the Ss.

$$SsTi = \sum_{k=1}^{n} S f_k s / n \times Ss$$

n = number of related security factors to threat i

Pi= prior probability of threat i
1-Pi= prior probability of not-occurrence of threat i
Ss= overall security score
Pi'= adjusted prior probability
Adjustment factor: SsTi
1 - Pi'= SsTi (1- Pi)

We already calculated the loss scores. And because the risk is the expected value of the loss or can be calculated as the loss multiplied by the probability, we can calculate the risk scores of the company by multiplying the loss scores by the adjusted company specific probabilities.
Now we can range the threats in descending order of risk score. Priority will be given to the first categories. Dependent on the budget the most effective security technologies will be implemented, needed to improve the security situation by preventing the company against the occurrence of the threats
.

Threat	Pi	1-Pi	SsTi	1-Pi'	Pi'
Financial fraud	0,12	0,88	1,01	0,89	0,11
Virus	0,52	0,48	0,65	0,31	0,69
System penetration	0,13	0,87	0,93	0,81	0,19
Theft of data	0,25	0,75	1,02	0,76	0,24
Laptop/Mobile theft	0,5	0,5	1,07	0,54	0,46
Insider abuse of net access	0,59	0,41	0,90	0,37	0,63
Denial of Service	0,25	0,75	0,87	0,65	0,35
Phishing	0,26	0,74	0,85	0,63	0,37
Illigal software applications on the system (bots, trojan horses,…)	0,21	0,79	0,83	0,66	0,34
Unauthorised access to information	0,25	0,75	0,93	0,70	0,30
Instant messaging abuse	0,25	0,75	0,85	0,64	0,36

Figure 8. Adjusted probability of the threats

Threat	loss score	Pi'	Risk score
Financial fraud	1,00	0,11	10,76165
Virus	9,17	0,69	6,309383
System penetration	30,04	0,19	5,768547
Theft of data	23,25	0,24	5,532846
Laptop/Mobile theft	4,41	0,46	2,043806
Insider abuse of net access	2,78	0,63	1,752233
Denial of Service	6,56	0,35	2,296816
Phishing	6,01	0,37	2,2183
Illigal software applications on the system (bots, …)	7,76	0,34	2,649647
Unauthorised access to information	2,37	0,30	0,70927
Instant messaging abuse	0,46	0,36	0,16436

Fig. 9. Risk score

4. DECISION MAKING ABOUT THE IMPLEMENTATION OF SECURITY TECHNOLOGIES

The next stage is to resolve on to security technologies. First a table wit the most used technologies along the CSI-study.(5) (9)

Most used technolkogies (CSI-study)
AntiVirus Software
Firewall
VPN
AtiSpyware Software
Intrusion detection system
Ecryption for data
Vulnerability/patchmanagement
Server Based Acces control list
Static account login/password
Encryption for data in storage
Smart card/one time password token
Public key infrastructure
Specialized wireless security

Fig. 10 Most used technologies

Security technologies	Threats Ti			
	T1	T2	T3	…
M1= antivirus				X
M2= firewall		x	x	
…				
M14	x			
M15	x			

Fig. 11 Technologies to protect against the threats

The critical threats should be restrained by taking the necessary safety requirements. For each of the threats a security technology can be taken but it is also possible that one technology can resolve or effect several threats. It's important to evaluate the costs of the technology and the possible loss of a threat. It is obvious that the cost can't be higher than the expected loss.
After a period of approximately 3 months after implementing the security technologies, a new security audit should be taken. The new security score Ss is calculated and compared to the stated aimed Security score using the security technologies.
If there are security factors that score too low, these should be investigated and adjusted.

5. CONCLUSION

Following the standard ISO17799 model, an audit checklist has been developed. We developed a new methodology for the security management decision making. In the decision model we are using the output figures of the CSI study. A company can now very easily identify its critical security factors and the most serious security threats.

REFERENCES

1. Jean-Marc Lamère: la sécurité informatique; Dunod : La méthode MARION (Méthodologie d'Analyse de Risques Informatiques Orientée par Niveaux)
2. www.eisti.fr/~bg/COURSITACT/TXT/m_marion.txt
3. Information Security Management ; BS ISO/ IEC 17799:2005 ; SANS Audit Check List: author: Val Thiagarajan B.E., M.Comp, CCSE, MCSE, SFS, ITS 2319, IT Security Specialist.
4. (www.riskworld.net/7799-2.htm)
5. CSI2007 study about cybercrime: COMPUTER CRIME AND SECURITY SURVEY: by Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn and Robert Richardson
6. https://event.on24.com/eventRegistration/EventLobbyServlet?target=registration.jsp&eventid=27372&sessionid=1&key=42F39B89EE0B30BA951711A5E7A98EDD&sourcepage=register
7. http://mediaproducts.gartner.com/gc/webletter/computerassociates/vol3issue3_risk/index.html
8. Security Management: A New Model to Align Security With Business Needs; Sumner Blount, CA Security Solutions ;August 2006
9. K.C.Laudon & J.P.Laudon: Management Information Systems. Prentice Hall 2006.
10. Cavusoglu, Huseyin, B.Mishra, S.Raphunathan: "A model for evaluating IT security Investments". Communications of the ACM 47 no7- 2004.
11. D.W.Straub, R.J.Welke: "Coping with systems risk: security planning model for management decision making". MIS quarterly 22 no 4-1998.
12. L.Volonino & S.R.Robinson: "Principles and Practices of Information Security". Prentice Hall 2004.
13. http://www.knowledgeleader.com: internal audit tools and resources.
14. http://www.itcinstitute.com : IT audit checklist information security

STRUCTURES ARCHITECTURES AND MODELS

EXTENSION OF FORMAL CONCEPT THEORY FOR ORGANIZATION STRUCTURE ANALYSIS

Jan Kozusznik and Ivo Vondrak
Department of Computer Science, VSB-Technical University of Ostrava
tr. 17. listopadu 15
708 33 Ostrava-Poruba, Czech Republic
E-mail: {jan.kozusznik, ivo.vondrak}@vsb.cz

KEYWORDS
Business process, UML, Organization structure, Formal Concept Analysis, Singular value decomposition, Hass diagram.

ABSTRACT

Business processes represent the core of the company behavior. Business modeling deals with the process definition, simulation a verification. Obligatory part of the discipline is also organization structure definition. This paper shows how the engage formal – algebraic – aparat for structure analysis and simplification.

INTRODUCTION

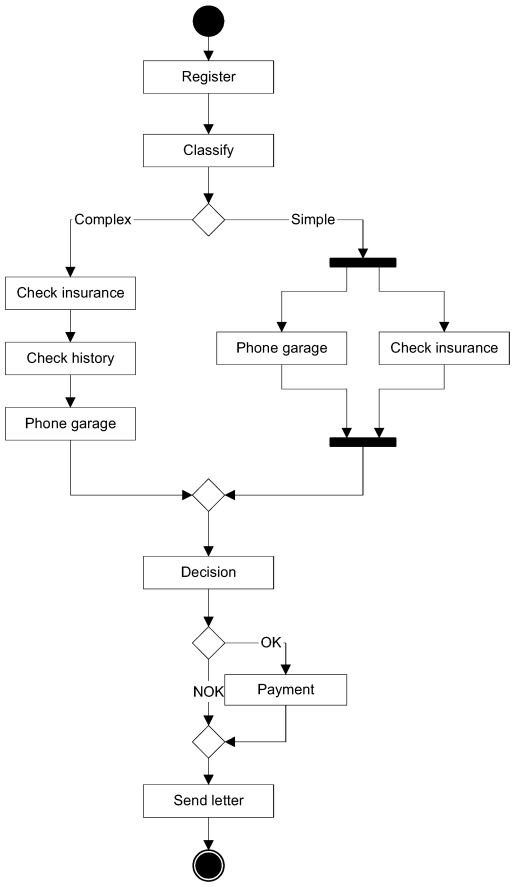
Business processes represent the core of the company behavior. Business process re-engineering is based on changes of the structure of business processes with the respect to obtain their higher efficiency. As a result of this business process re-engineering new organizational structure has to be defined to reflect changes in business processes.

There are many possibilities how these processes can be defined. Usually all modeling tools are focused on various kinds of business process aspects based on what abstraction is considered as the main (DeMarco 79, Peterson 77, Christie 95). Unfortunately, none of any views captures organization structure of roles implemented by human resources participating in processes being modeled. For example, BPM (Business Process Modeling) method (Vondrak 99) involves roles in a process specification but there is no option how the organizational structure implied by such models can be analyzed and evaluated. The organizational structure is usually defined using the best experience and there is a minimum of formal approach involved. This paper shows the possibilities of the theory of concept analysis and spectral analysis of matrices that can help to understand organizational structure based on solid mathematical foundations. In the paper, the theory of concept lattice is also extended by method of linear algebra called Singular Value Decomposition. This method of linear algebra is applied to concept lattice simplification.

The idea of formalization of terms of context and concept by means of theory of lattices is not quite new. Formerly isolated experiments at application so-called Galois´ lattices already existed, especially in the area of Information retrieval, but systematically built up theory did not arose until the work of R.Wille in about 1980 at the TH Darmstadt, Germany and his group. FCA is a mathematical approach to data analysis based on the lattice theory of Garret Birkhoff in 1993 (Birkhoff 93).

DEMONSTRATION EXAMPLE

For the formalization of term concept and context a theory of ordered sets and the theory of lattices are applied. The terms from these classic disciplines can be found in (Ganter 99). Conceptual analysis using for organize structure modeling can be found in (Vondrak 03). We map roles to objects and activities to attributes. For demonstration we use small example of claim handling process. Activity diagram of this process is in Figures 1.



Figures 1: Claim handling process

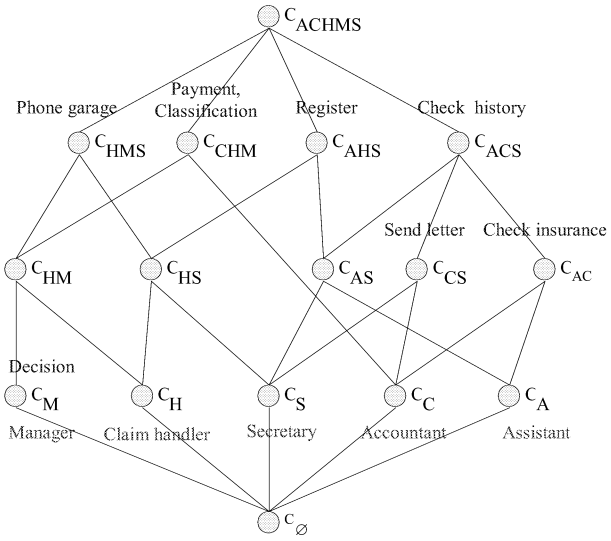
There are five roles in this process: *manager, claim handler, assistant of claim handler, accountant* and *secretary.* Ad-hoc definition of competency table follows:

	Register	Classify	Chck.i.	Chck.h.	Phone garage	Decision	Payment	Send letter
Manager		X			X	X	X	
Claim handler	X	X			X		X	
Accountant		X	X	X			X	X
Assistant	X		X	X				
Secretary	X			X	X			X

Table 1

Result Hass diagram representing concept lattice is in the Figures 2. This graph provides alternate views on the information contained in the above-described table. In other words, the concept lattice enables to visualize the structure "hidden" in the binary relation. Obviously, the more complex is the table of responsibilities the more difficult is to understand who is responsible for what. The nodes in our picture of concept lattice can be considered as a potential source of how the organizational units can be defined.

Based on such modification of concept lattices this theory provides exact and formally well-defined way how the organizational structure can be analyzed and re-designed.



Figures 2: Result Hass diagram

SINGULAR VALUE DECOMPOSITION

However, lattice constructed for more complex organizational structure is still messy. There is need of reducing size of concept lattice. The reduced concept lattice could be used as coarse view on organizational structure. Detailed view will provided only for specific part of structure. In this case, there arises some kind of hierarchical structure. In this paper, the approach for size reduction presented in (Gajdoš 04) is introduced. It is based on a Singular Value Decomposition.

Theorem 1. Let A be a matrix of $m \times n$ rank. Let $e_1 \geq e_2 \geq \cdots \geq e_r$ be eigenvalues of a matrix $\sqrt{AA^T}$. There exists matrices $U = (u_1, \ldots, u_r)$ and $V = (v_1, \ldots, v_r)$, and diagonal matrix $S = diag(e_1, e_2, \ldots, e_r)$ so that $A = USV^T$. The decomposition is called **singular value decomposition** (from this moment SVD) of matrix A and eigenvalues of a matrix $\sqrt{AA^T}$ are called **singular values** of matrix A.

Definition 1. Let us have k, $0 < k \leq r$ and singular value decomposition of matrix A

$$A = USV^T = (U_k U_0)\begin{pmatrix} S_k & 0 \\ 0 & S_0 \end{pmatrix}\begin{pmatrix} V_k{}^T \\ V_0{}^T \end{pmatrix}$$

$A_k = U_k S_k (V_k)^T$ is called **k-reduced singular value decomposition (rank-k SVD)**.

Theorem 2. (Eckart-Young). Among all $m \times n$ matrices C of rank at most k A_k is the one that minimizes $\|A_k - A\|_F^2 = \sum_{i,j}(A_{ij} - C_{ij})^2$.

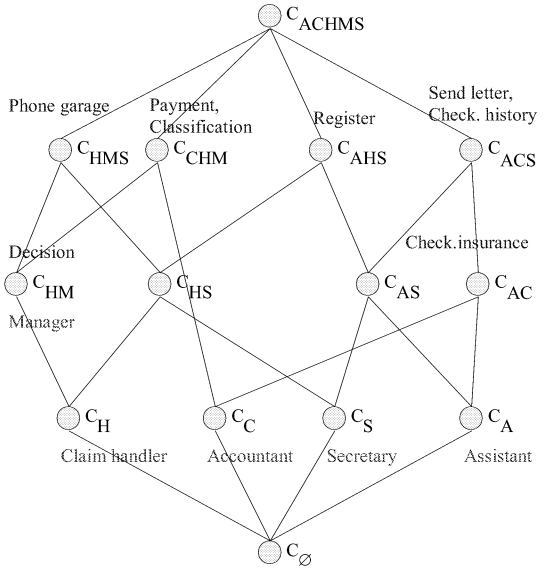
The SVD could be applied on relation matrix of objects and attributes that expresses context in FCA. A rank-k SVD application could be interpreted as a noise removing by ignoring small differences between row and column vectors of matrix A – it corresponds to small singular values, which are dropped by the choice of k.

Of course, SVD is mainly used for matrices with real numbers. Its application on a context, that contains only binary values 0 or 1, is also possible. However, resultant approximation matrix A_k contains real numbers, so there is used a *threshold t* discriminating between ones and zeros. In praxis, values smaller or equal then t are assigned to 0 and values higher then t are assigned to 1. Finally, this method contains two parameters k and t that have influence on resultant context.

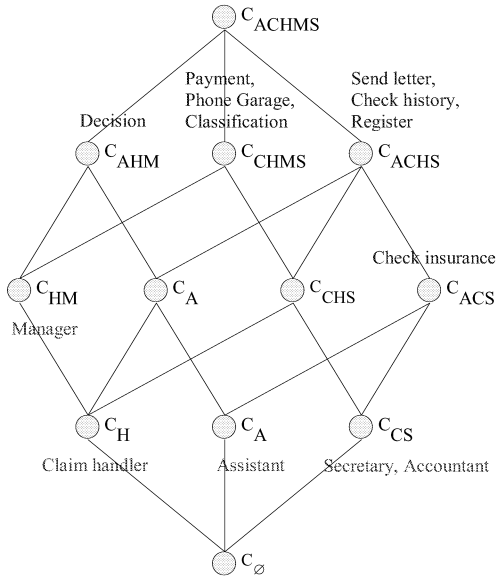
RESULTS

On the Figures 3, there is Hass diagram for concept after SVD application. It was set *k=3* and *t=0,3*. The result Hass diagram isn't as complex as original and it is also visible that activity "*Send letter*" and "*Check history*" are similar – they are contained in the same concept. It misses two concepts. E.g., *Manager* doesn't have any additional competency in comparison to *Claim handler.* In the original contexts, *Manager* has competency for performing of activity *Decision* while *Claim handler* doesn't have. It could show possibility for process simplification. On the other hand, we would like to have only one person with competency for *Decision* making in reality – it depends on particular case.

Different values for k, t provides different results how you can see on the Figures 4. This lattice is even simpler but it is probably misrepresented – *Secretary* and *Accountant* are in the same concept.



Figures 3: Concept lattice after SVD application



Figures 4: SVD with k=2, t=0,2

CONCLUSION

Presented rank-k SVD seems to be valuable. Lattice complexity reduction is the more important the more complicated is context. The presented example was quite simple and difference between original concept and reduced concept wasn't significant – more complex examples are better for demonstration of the SVD strength. However more complicated examples exceed extends of this paper.
SVD application is also valueable that it could lead us to simplification of organization structure. It shows how the persons are similar and fungible.

REFERENCES

Christie A. 1995. *Software Process Automation*. Springer-Verlag

Birkhoff G. 1993. *Lattice Theory*. 3rd edition, American Mathematical Society, Providence, RI.

DeMarco T. 1979. *Structured Analysis and System Specification*. Prentice-Hall, Englewood Cliffs, New Jersey

Ganter B., Wille R. 1999. *Formal Concept Analysis, Mathematical Foundation*. Sringer-Verlag.

Gajdoš P., Moravec P., Snášel V. 2004: "*Concept Lattice Generation by Singular Value Decomposition"*, in proceedings of Concept Lattices and their Applications (CLA 2004), Ostrava, Czech Republic.

Peterson J.L. 1977. "Petri Nets." *ACM Computing Surveys*, vol.9, no.3 (Sept): 223-251

Vondrak I., Szturc R., Kruzel M. 1999: "Company Driven by Process Models." *In Proceedings of European Concurrent Engineering Conference ECEC '99* (Erlangen-Nuremberg, Germany), SCS, Ghent, Belgium,. 188-193

Vondrak I., Kozusznik J. 2003: "Using Concept Lattices for Organizational Structure Modeling and Analysis" *European Concurrent Engineering Conference ECEC 2003* (Plymouth, United Kingdom), SCS, Ghent, Belgium, 5-9

TOWARDS AN INTEGRATED CORPORATE INFORMATION SYSTEM: BUILDING UP A SALES PRODUCT DATA MODEL TO COPE WITH THE EMERGING COMPLEXITY OF E-BUSINESS

Christian-A. Schumann and Andreas Rutsch
University of Applied Sciences Zwickau
Institute of Management and Information
08056 Zwickau, Germany
E-mail: christian.schumann@fh-zwickau.de

KEYWORDS
E-business, integrated enterprise information, multichannel e-commerce, product-oriented information

ABSTRACT

Today, supplying customers with products and related information and data is a unidirectional and mostly static problem for many final manufacturers. Especially in the field of e-business, information systems will need more access to data in the required context and volume enclosed in the enterprise internal layers and sources to respond to changing customer demands. The aim is giving customers the opportunity to change products and product ranges over their current limits. The establishment of a dynamic interaction loop between sales (customer interface) and the corporate information system is needed to better fulfil the fluctuating customer demand. Therefore a Sales Product Data Model (SPDM) is established upon the requirements of E-Commerce (EC). The steps of research and implementation are presented. The current laboratory architecture is shown. The architecture is based on a state-of-the-art integrated enterprise information system, including Computer Aided Design (CAD), Product Data Management (PDM), Part Management (PM), Enterprise Resource Planning (ERP), and E-Commerce subsystems.

E-COMMERCE AND ITS INFLUENCE ON NEAR-TERM PRODUCT DEVELOPMENT

The globalisation influences the technological change and the spatial organisation of the economic activities (Dunning 1998) of large-scale companies as well as increasingly of Small and Medium sized Enterprises (SME).

E-business concerns the development as well as the partly respective complete support, processing, and maintenance of performance exchange processes by electronic networks. Those processes are procedures including the transfer of material and immaterial goods and services into compensatory accounts. The main objective of the performance exchange processes from the viewpoint of an organisational participant is the sustainable yield bringing performance creation. (Wirtz 2001) E-commerce is, beside e-information, e-education, e-communication, and e-collaboration, a subset of e-business. It concerns performance exchange processes of development, brokering, and conclusion of commercial transactions among business units by electronic networks. (Wirtz 2001)

The main objectives of e-commerce are the increase of efficiency, the decrease of costs, and the user-friendliness of any commercial transaction. EC is an interesting alternative especially for SME with limited resources for the opening of new markets to improve the market access on different ways. (Wirtz 2001) But e-commerce is also a tough challenge due to high expectations especially with respect to lead times. E-commerce should finally lead to more customer impact on the up-to-date near-term product development.

E-commerce is often underestimated in its impact on old-fashioned commercial operations. So many companies are fatally wrong in facing the e-commerce challenge not until they participate in electronic markets. Especially the SME are characterised by a huge backlog in the recognition of the actual market situation. The conditions of the producer, the market, and the customer are essential for defining an e-commerce strategy. The star role approaches the key accounts. The e-commerce strategy should adhere to the specific conditions of the key accounts. Respecting arbitrage it is important to define different EC approaches depending especially on the interaction in pricing of neighbouring regions.

Therefore the development of a product-oriented information system must also include an integrated data and information management for multichannel e-commerce systems. It is essential in consideration of the international competition to close the gap in EC strategy by the SME.

CONTROLLING THE EMERGING COMPLEXITY VIA INTEGRATION

Each enterprise needs a general concept for the provision of the right data and information, supporting electronic and non-electronic commerce activities. (Schumann et al. 2004) The concept has to consider a wide range of commercial activities, methods, and tools. Typically, it includes catalogue data demands, customer intranet access specifics, and e-shop requirements in the form of data supply. Those are mainly presented as one-way, a passive kind of customer support via prepared, static data descriptions.

Another, more complex case is the use of product configurators. They include the customer-related design of the product (the configuration) and as a result the dynamic change of the data and its structures.

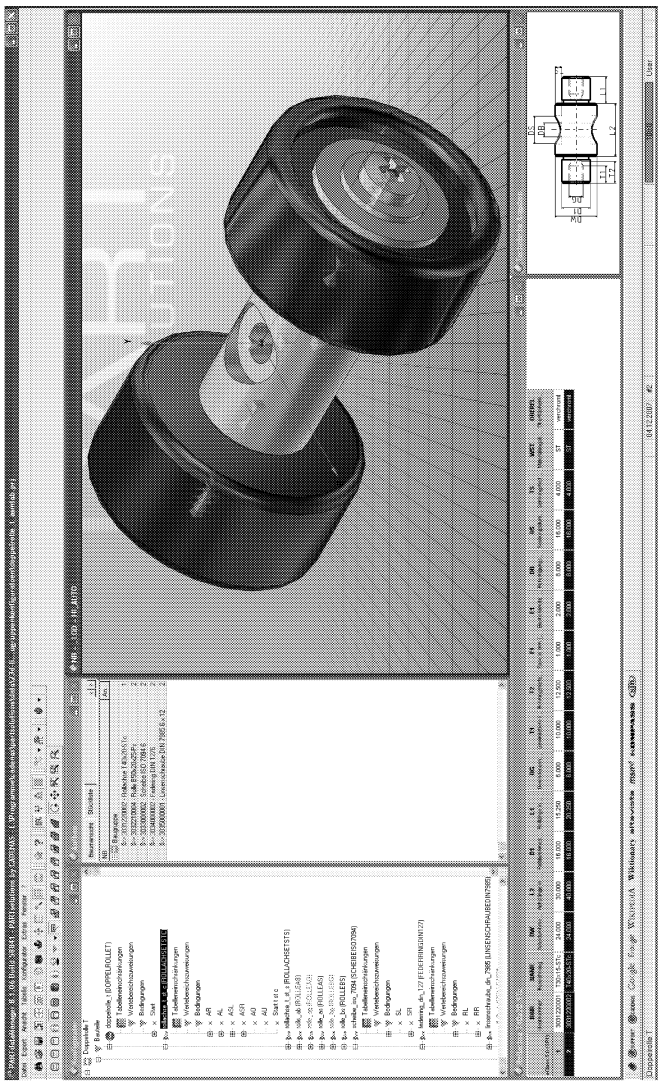


Figure 1: The twin wheel castor configurator

If the producer offers customer-depended manufacturing of products as unique specimen, he has to organise an active, bidirectional linking of the information and data flow between the enterprise data sources and systems, and the commercial support systems. The product configurators support capabilities such as guided selling, guidance of the user to recommended options, presentation of alternative features and options, constrained choices based on compatibility rules, display of price, lead time and availability, support of complex calculations and access to engineering algorithms, etc.. It is necessary to develop a network concept of data exchange between the data source systems and the distribution channels.

There are different classes of data sources and data source systems. The first and main class, the product data master class, implies the product description by product attributes and behaviours. Usually, the source systems are product describing systems like CAD systems and animation systems in interaction with sub domains of the product management systems such as ERP and PM systems. The second one is the product data relation class. It consists of the description of the assortments and the compositions of products configured for special application cases. The data sources are available in correspondent systems such as sub domains of PM systems, ERP systems, and product configuration systems.

The concept of a multichannel e-commerce solution is based on dual layer architecture. The two kernel layers for the multichannel e-commerce sources are the PDM layer and the product data description layer. The PDM layer contains the integrated data base as a single product information source and should be media-neutral. (Schumann and Rutsch 2006)

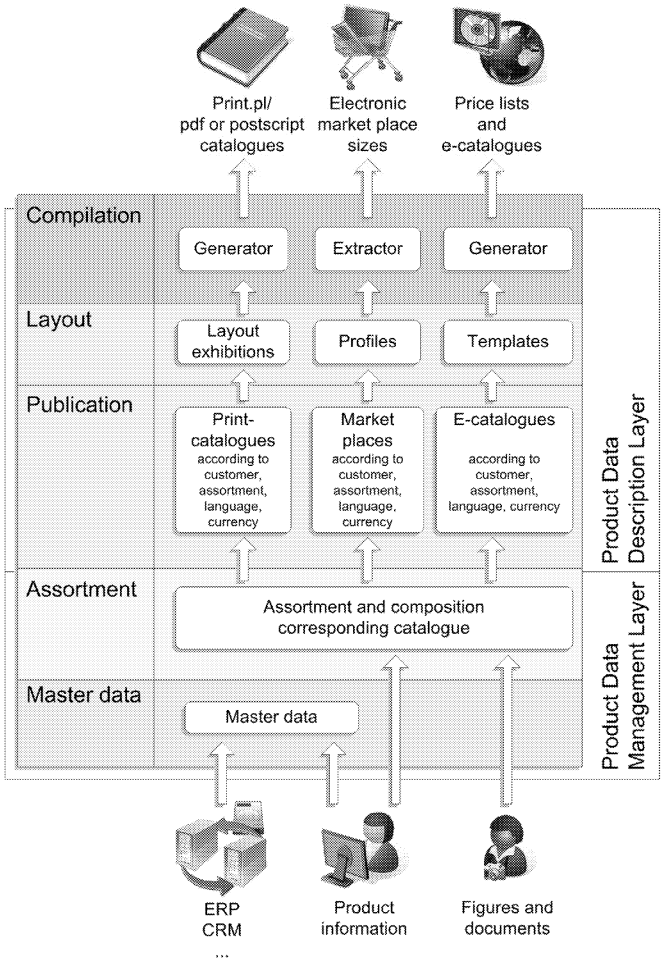


Figure 2: A multichannel e-commerce architecture

It can be realised as a centralised data base system or as an integrated, distributed data base system. In any case it should have a unified data base management system. The PDM layer is linked to the main enterprise information systems such as ERP, CRM, Computer-Aided (CAX), and controlling systems by special interfaces. The product data description layer involves the different opportunities for publishing product data as customer-related information.

In principle, the PM and the related PDM system are able to support the unidirectional as well as the bidirectional exchange of data from enterprise data source systems to the commercial support systems. Recently, the problem is the

availability of the required interfaces for bidirectional data exchange of the software packages. The issue is that the prevalent unidirectional data flow is only appropriate for the quasi static data support of multichannel e-commerce systems. It allows only the variation of e-commerce data set in the framework, which was planned ahead by the sales management and is realised by using the links to the data sources as well as the media-neutral data base for the multichannel e-commerce.

The effort to offer customers a wide range of options, alternative choices, and related short lead times usually requires bidirectional relations between the data source systems of the enterprise, the PM and PDM systems, and the multichannel e-commerce system including special applications such as product configurators. The task is very complex but enables marketing of the enterprise to achieve better competitiveness in customer-dominated markets.

STATE-OF-THE-ART

Recently, SME have usually neither a multichannel e-commerce system nor an integrated enterprise information system and more than ever no complex solution for the bidirectional, dynamic data exchange between the data source and sales system. The result of analysing the state-of-the-art on behalf of a company as a European market player was the description of the important e-commerce channels. Data source systems including CAD and ERP were located. The current status was listed in a data source and flow map. This is the prerequisite for the replacement of the auxiliary solution by implementing a PM and PDM system. The existing inhomogeneous classic as well as IT-solutions for the distribution have been replaced by an integrated system as follows. (Dippold 2005; Schumann and Rutsch 2007)

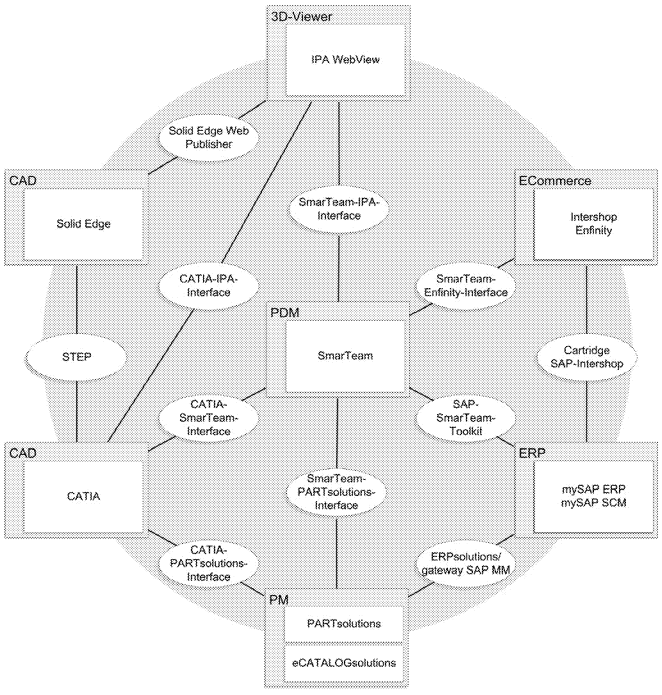


Figure 3: Laboratory architecture

The knowledge about the different internal and external data requirements, layers, systems, and sources was the basis for developing a model for an integrated approach. Several related functions and systems were involved in the integrated information system. The basic layer is dominated by the data source systems such as CAD and ERP, the intermediate layer is shaped by the PM and PDM systems, and the outlet layer will be characterised by the multichannel e-commerce system. (Vernadat 1996)

Based on the international state-of-the-art a first step, by using standard software and data base systems, towards an integrated corporate information system has been made. A laboratory at the Zwickau Institute of Management and Information (Figure 3) has been set up.

While a nearly complete set of corporate information systems is implemented, one of the first conclusions of research was just to use a subset to reduce the complexity of the system. The ones not taken into account in the first place will be included within the future steps of research.

STEPS OF RESEARCH AND IMPLEMENTATION

The Inner Loop (First Step)

The elements of the inner loop (PDM, ERP, and PM) are used in the primary step (Figure 4), because they include the main functions needed to design a proprietary feedback loop.

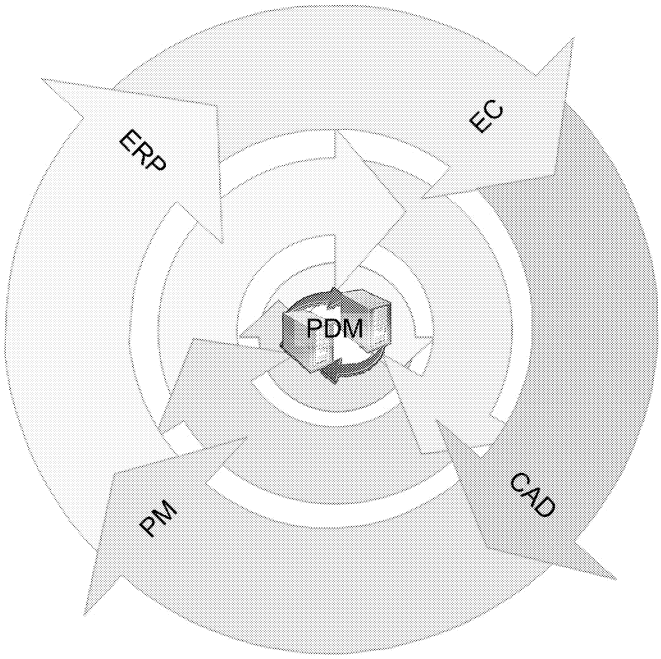


Figure 4: Steps of research and implementation

The catalogue software eCATALOGsolutions from CADENAS was used to create a Sales Product Data Model (SPDM) which includes geometric as well as product information data. The model is packed in a configurator to give the customer some scope for design. Twin wheel castors are used as samples (Figure 1) which could be selected from different default dimensions. After the SPDM is defined it is

made available for the customer within PARTsolutions. The customer can now create a configuration which satisfies his requirements. While doing so a bill of materials (BOM) is generated immediately.

After the customer deployed the order within the PM-system a routine saves the BOM customer-related into the PDM data base. Normally in this procedure an e-commerce or shop-system should be involved, for instance to handle customer information and shop access. This should not be the case in our primary step of evaluation. The evaluated PM system offers also the opportunity to save the configuration being made in different CAD-formats via native data export. The drawings now available for production enable thinking about more complex products to-be defined by the customer for future configurators.

Primarily based on the BOM and the order quantity the ERP-system initiates the costumer-based job execution in the production. Also the customer data passes the interface and is ready for accounting. Job execution in an automated way is possible, because the ERP system knows all material positions. Those are directly linked with the corresponding resources. This static chaining should be replaced later on.

The first part of the inner "market-driven optimization of product ranges"-feedback loop has been created. The next challenge is in getting a direct feedback from the ERP system into the SPDM, initially defined at the setup of the system. Economic metrics have to be evaluated or defined, which affect the loop primarily in relation to cost optimization of the production involved. Closing the first or inner loop by implementing bidirectional interfaces provides the basis for the next step.

Second Step

The second step contains the implementation of a new customer interface in the form of an EC system. The e-commerce system offers the opportunity to gain more customer-metrics beyond the SPDM. Those metrics are known as e-metrics. (Sterne 2002)
E-metrics are various measures describing the behaviour of visitors especially on web pages. They are used to review efforts made in a special context, e.g. to sell products (Figure 5).
Using those measures and the technology behind to optimise the SPDM with respect to customer requirements is the new stated context. The overall goal is the same: selling products. According to the approach of Jim Sterne there are several stages to the point where the purchase is being placed. After log in or register, the potential customer starts exploring and browsing the SPDM, maybe focusing on details, selecting and engaging some options, and finally purchases, signs-up or agrees with the choice made. (Sterne 2002) If a customer is focused and skips stages he shall be deemed to be determined or convinced. If he is leaving without ordering, the customer could possibly be unfocused, uncommitted or a researcher.

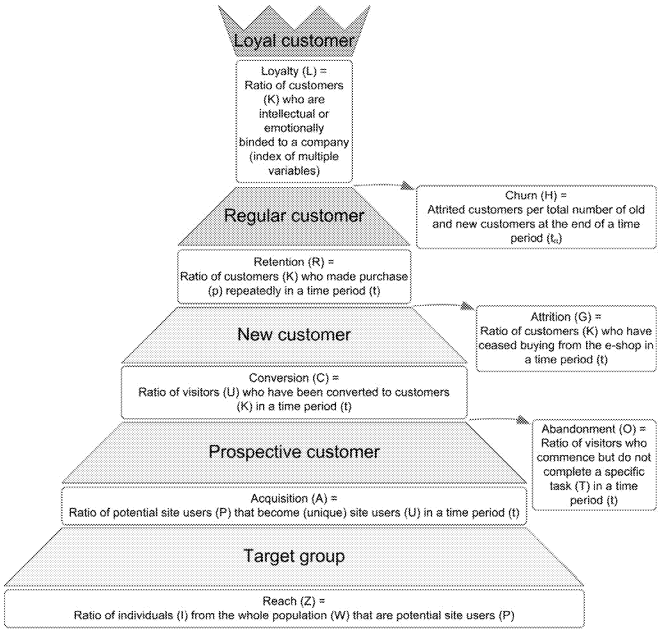


Figure 5: Success e-measures
(Reinecke and Janz 2007; Sterne 2002)

The SPDM therefore could be deeply optimized with respect to customer requirements using SPDM-customer data (a selected order) and e-measures in the first stage of expansion of the inner loop or system. Therefore the SPDM needs to be published in the context of an e-commerce channel.

To achieve proper feedback on market or customer demand rules based on e-metrics have to be developed and evaluated. At this stage the new customer interface (Figure 6) offers the chance to measure customer activities.

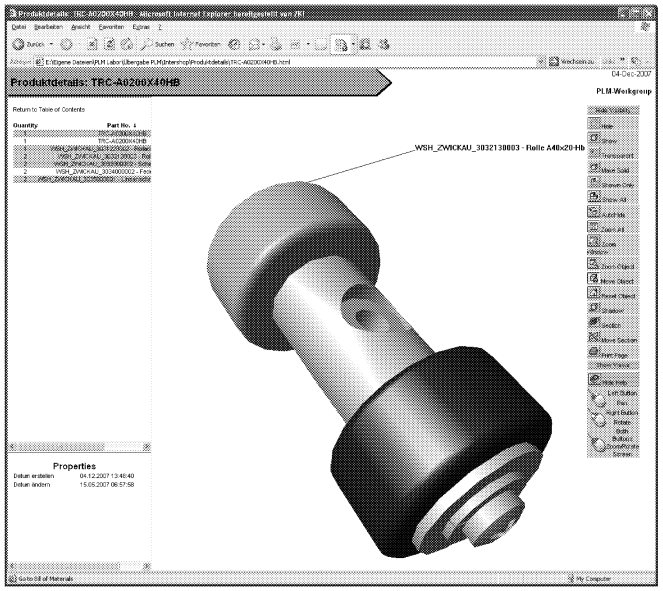


Figure 6: SPDM published via hypertext markup language

Based on those metrics (e.g. product-related acquisition, conversion, or retention of customers) it should be possible to near-term estimate the changing product demand. The adoption of an e-commerce system also enables the access to

the different kinds of EC sales channels to expand business and to open new markets.

Last Step (Outer Loop)

The last to-be implemented elements are the multiple CAD systems as the real current source of engineering design. This seems to be the main challenge due to the differences within the multiple CAD applications. The substitution of the product configurator is the aim of implementing CAD systems. It is foreseen that the system will collapse in multi-CAD usage. Therefore a middleware is needed. A PM system could be used as a repository for the definition of standard features, as recommended in ISO 10303-224:2006, for designing new products or defining orders.

CONCLUSSION

The e-commerce scene of SME will be dominated by integrated enterprise information systems in the near future. It is necessary to analyse and describe the business processes and models of the future e-commerce scenarios for SME. Their requirements influence the planning and design of the information flows and data support systems strongly. Especially, the use of PM and PDM systems as intermediate connectors in integrated enterprise information solutions will enormously increase within the next years.

This is the only way for SME to preserve the competitiveness in the procedure of dynamisation and globalisation. The unidirectional solutions are implementable by the existing standard software and interfaces. The bidirectional, dynamic data flow requires add on activities for completion of the available systems enabling the directed access of multichannel e-commerce systems to the enterprise data sources in integrated information systems of SME.

REFERENCES

Dippold, R. 2005. *Unternehmensweites Datenmanagement: von der Datenbankadministration bis zum Informationsmanagement*. Vieweg, Braunschweig.

Dunning, J.H. 1998. "Globalization, Technological Change and the Spatial Organization of Economic Activity". In *The dynamic firm*, A.D. Chandler, P. Hagström and Ö. Sölvell (Eds.). Oxford University Press, Oxford.

Reinecke, S. and S. Janz. 2007. *Marketingcontrolling*. Kohlhammer, Stuttgart.

Schumann, C.-A.; S. Kassel; C. Tittmann; and G. Kay. 2004. "AI-based Integration of Business Intelligence and Knowledge Management in Enterprises". In *The Symposium on Professional Practice in AI: 1st IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI2004) as part of the 18th IFIP World Conference*, E. Mercier-Laurent (Ed.). Toulouse.

Schumann, C.-A. and A. Rutsch. 2006. "Multichannel E-Commerce System for SME as European Market Player". In *Proceedings of the 41st Annual Assembly of the Latin American Council of Management Schools CLADEA 2006*. Montpellier.

Schumann, C.-A. and A. Rutsch. 2007. "Directed Access of Multichannel E-Commerce Systems to Data Sources of Integrated SME Information Systems". In *Proceedings of the 14th European Concurrent Engineering Conference (ECEC) 2007*. Delft.

Sterne, J. 2002. *Web metrics: proven methods for measuring Web site success*. Wiley, New York.

Vernadat, F. 1996. *Enterprise modeling and integration*. Chapman & Hall, London.

Wirtz, B.W. 2001. *Electronic Business*. Gabler, Wiesbaden.

AUTHOR BIOGRAPHY

CHRISTIAN-ANDREAS SCHUMANN, born 1957 in Chemnitz (Germany), studied Industrial Engineering at the 'Chemnitz University of Technology' (CUT), doing his first doctor's degree in 1984 and second doctor's degree in 1987. He was appointed associate professor for plant planning and information processes at CUT in 1988. In 1994 he became professor for business and engineering information systems at the 'University of Applied Sciences Zwickau'. Since March 2003 he is dean of the faculty 'Business and Management Sciences' at Zwickau. Currently he is also director of the 'Centre for New Forms of Education' and director of the 'Central German Academy of Further Education'.

AN ARCHITECTURE FOR COLLABORATIVE WORKING ENVIRONMENTS IN MANUFACTURING INDUSTRY

Ana Correia
Cristina Grama
Dragan Stokic
ATB-Bremen
Wiener Strasse 1

28359 Bremen
Germany

Ana Campos
Rui Neves Silva

Uninova
FCT Campus

2829-516 Caparica
Portugal

Jon Aguirre

Fundacion Fatronik
Paseo Mikeletegi 7, Parque Tecnológico
20009 Donostia
Spain

KEYWORDS
Collaborative environment architecture, Service Oriented Architecture, Manufacturing industry

ABSTRACT

The paper proposes a new architecture for collaborative working environments (CWE) in manufacturing industry. The intention is to contribute to a definition of a reference architecture for CWE. In order to define such an architecture, specific requirements regarding CWE in manufacturing industry are analysed. Three views upon such an architecture are elaborated: User level reference architecture, Building block level reference architecture, and Software interface level reference architecture. The novel proposed architecture follows principles of Service Oriented architecture (SOA) and WEB 2.0 approach. Implementation issues are discussed as well.

INTRODUCTION

In order to enable efficient design of different systems such as Collaborative Working Environments (CWE), automation systems, processes, it is necessary to establish a consistent development methodology. A key tool of such a methodology is the definition of a Reference Architecture (RA) for these systems, which should provide a unified representation of essential features. The basic objectives are to (Dornier 1992):

- provide a simple, unified architecture for systems, enabling traceability between the solution independent requirements and final implementations,
- achieve minimum complexity and design simplification,
- support for interfacing, and
- support re-usability of modules, etc.

A RA may serve as a tool for the comparison of different concepts/implementations, their evaluation and an integration of modules developed within different approaches. It shall enable and efficiently support the communication process between people involved in the development process. This means that it shall serve as a means to ensure a common unified, unambiguous and widely understood terminology between people interested, from a utilisation point of view, in the behaviour of the system, and people, being experts in different technical domains, realising a specific functionality of the system to achieve the required behaviour and performance. To reach this objective, a RA shall be easily interpretable and applicable and, therefore, it shall represent a structure of limited complexity (Stokic et al. 1994).
Specifically the need for RA for CWE comes from the need to achieve a global agreement on how collaborative environments interoperate in order to create a critical mass of developers and users – similar to the World Wide Web. A way to specify this agreement is to define a RA for collaborative environments (Decker 2006).
The objective is to seamlessly and ubiquitously provide a CWE to the users, thanks to a collaborative service infrastructure and end-user application-specific interfaces. For this, collaborative service interoperability is a crucial goal (Ralli 2006).
The paper intends to contribute to a definition of RA for CWE specifically applicable for collaborative work in manufacturing industry.

STATE-OF-THE-ART

Since a RA is the result of an agreement and dissemination act, creating a standardisation group is a necessity for achieving progress in this direction (Decker 2006). Such a group exists and is known as Open Collaborative Architecture Working Group or OCA WG (Ralli 2006).
The mission of the OCA WG is to create a forum of discussion and contributions to find out the rules of the CWE architecture (meta-architecture) in order to enable CWE service interoperability.
High-level collaboration middleware, or E-collaboration UpperWare, is a software generic infrastructure where the generic model of collaborative infrastructure is mapped to the services; furthermore, to alleviate the tasks of the developers, the basic collaborative functions, locking, presentation control, user presence management, organisation management and communication control are gathered into a collaborative infrastructure and made available to the applications. Also a collaborative service can be built by composing or by orchestrating the (distributed)

collaborative functions provided by the collaborative infrastructure. Current Middleware definition as a "glue" to patch machines and their software functions together is not appropriate anymore to distinguish this new conceptual approach. Next to a fundamental architectural blueprint for RA it seems useful to include vertical "plug-ins" as extensions for various, often changing, industrial requirements which are not an integral part of the fundamental layer (ExpertGroup 2006).
Most of the architectures for CWE proposed so far share three-layer models where basic technologies are at the bottom, collaborative tools stand for the middleware and user interfaces complete the top level. Different capacities, functions and features are considered in several ways, depending on each specific initial approach and therefore a general layered meta-architecture is to be worked out.
Several architectures are currently available. The one proposed by C@R project (C@R 2007) served also as an initial OCA-WG RA for CWE.
The EC Expert Group on CWE, although not intending to specify a RA, has also defined what could be the key elements of the CWE infrastructure. The emphasis is that all functions will be offered as services (generic services, domain-specific services and context-specific services) to the user or application developer. The combination of generic cross-domain services and tools with domain-specific ones will build application-specific collaboration environments, and a combination of generic and domain specific services with context-specific services will build situational specific collaborative applications. This reflects two key aspects of CWE – a key dual approach: generic cross-domain aspects and domain, application specific aspects (ExpertGroup 2006).
The objective of the project CoSpaces (CoSpaces 2007) is to develop collaboration models and distributed technologies to support collaborative workspaces in distributed virtual manufacturing enterprises. The development of CoSpaces architecture is based on the Open Group Architecture Framework (TOGAF 2007) methodology. TOGAF is a framework for developing architectures. It provides tools to design, evaluate and build architectures for different purposes.
The ECOSPACE project (ECOSPACE 2007) aims to build an infrastructure for the so-called Collaborative Working Environment for e-Professionals, starting from such concepts as SOA, Web 2.0 and the Semantic Web.
As it is well known, there is no RA for collaborative work in the manufacturing industry, and the proposed paper intends to contribute to the specification of such a RA.

RESEARCH APPROACH

Main challenges in the definition of reference architectures for CWE

According to (ExpertGroup 2006), research on CWE should focus on reference models for semantic collaborative working, and on seamlessly integrated activity-oriented, context-aware collaboration functions. The key emphasis is on:

- **context modelling and awareness and activity oriented functions**. The objective is to support **semantic interoperability** and ensure flexible cross-domain working environments in which functions could be effectively re-used. Semantic interoperability could be defined in this context as the ability to integrate resources (elements) that were developed using different vocabularies and different perspectives on the data. The objective is to stress system thinking in defining 'upperware'.

- **evolving** (dynamically changing) **collaborative patterns**, aiming again to support integration of synchronous and asynchronous work.

The reference models should support **'automatic' composition and integration** of basic functions exposed as services, taking into account the WEB 2.0 paradigm.
Explicit RTD challenges arise from the need to empower people to work both in unstructured teams and in well-defined and stable settings through evolving CWE reference models for seamlessly integrated, activity-oriented, context-aware collaboration functions that provide enhanced knowledge sharing mechanisms, better decision making process and less burdensome group processes support. The security and Intellectual Property Rights (IPR) issues need also to be considered.
The RTD challenges to be addressed are, therefore: context modelling, activity-oriented CWE, evolving collaboration patterns and integration of synchronous and asynchronous work, WEB 2.0 issues.

Requirements upon the CWE architecture in manufacturing industry

The paper focuses on CWE in manufacturing industry. Therefore, the specific requirements for CWE in manufacturing industry are taken into account.
The key requirements on the architectures for CWE in manufacturing industry can be summarised as follows (Stokic 2006):

1. Needs for generic collaborative services supporting different collaborative patterns in the manufacturing industry and focusing primarily on organised groups.

2. Needs for tools to support generation/update of collaborative services. The objective is not to have a straightforward possibility of developing/updating services by the (non-ICT specialists) users themselves, but to allow the ICT experts in manufacturing industry (non experts for CWE technology) to easily update/maintain the collaborative services.

3. Since for collaboration in industry the work with real and actual data (e.g. on-line data from shop-floor processes) is of crucial importance, an information middleware for collaboration and virtualised resources is a key element of an architecture for a CWE.

4. Security and safety are key issues for CWE in manufacturing industry. Therefore, services/tools for

CWE in industry must strongly take into account access capabilities in industrial companies.

5. Although the prime objective of CWE in manufacturing industry is to focus on organised teams, the manufacturing industry has to be open (on longer-term) for collaboration with a wider community. Subsequently, pro-active, culturally-aware services allowing access to virtualised resources and knowledge to support creativity by involving teams within wider communities in generation of new ideas on products and processes and solving problems are needed.

6. Traceability and repeatability of work is required in industry. Therefore all collaborative work must be fully traceable and appropriately documented.

7. Multi-criteria decision making in industrial collaborative environments, with main emphasis on responsibility of the decision, traceability, selection and weight of the decision criteria, representation and handling of uncertainty has to be supported by the CWE services.

8. The collaborative services and their design/updates must be under clear control of one or more parties where IPR issues are clearly defined.

9. Formalised collaboration is important in industrial environments, to enable repetition of processes and a faster learning curve.

10. Definition of the appropriate targets and timings for collaboration.

11. Collaboration in manufacturing has a close link to information sharing and creation, and should have a well-defined application.

Other requirements include individual characteristics of actors, regarding position, movement, patterns, preferences, etc. In an eCollaborative working environment, a certain degree of customisation is mandatory.

FINDINGS

Proposed architecture

The proposed architecture follows the SOA principles and WEB 2.0 approach (Campos et al. 2007). RA is composed of a collection of views at various levels. This approach has been applied employing three views:

- User level reference architecture,
- Building block level reference architecture, and
- Software interface level reference architecture

The following text details these three views, showing how the proposed architecture conforms to them.

User Level Reference Architecture

The view at the conceptual level focuses on the functions offered to the human users. Only externally observable behaviour is highlighted in the diagram of Figure 1, so the low-level details or implementation issues are not depicted.

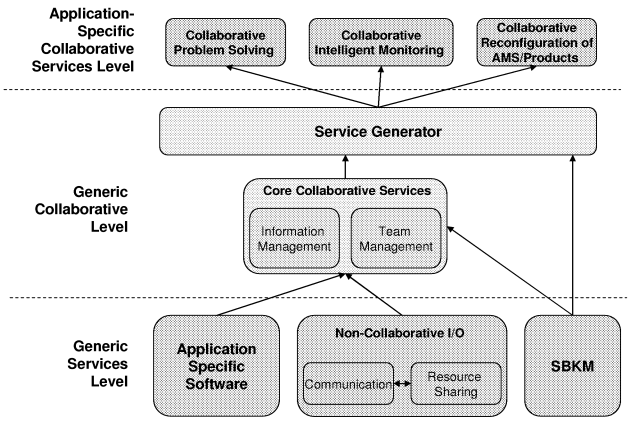


Figure 1: Proposed User level architecture.

Three levels can be distinguished:

- generic services level for non-collaborative functionality/services,
  - Generic, non-collaborative I/O functions (communication using chat programs, e-mails, video conferences and so on, and resource sharing, involving more complex forms of representation, such as diagrams and drawings)
  - Semantic-based knowledge management (SBKM) tools for dealing with ontologies, reasoning (for instance, search functionality, case-based reasoning, rule-based reasoning).
  - Application-specific software.
- generic collaborative level for functionality/services fulfilling generic and elementary collaborative tasks (such as resources discovery, tracing collaboration etc.)
  - Core collaborative services (CCS): these are atomic collaborative services (for instance, collaboration call) which are used by the Service Generator in order to compose application-specific collaborative services (see the text to follow on further examples for CCS).
  - Service generator: used for combining the atomic collaborative services described above into application-specific collaborative services (in line with WEB 2.0 approach).
- collaborative application-specific services (CAS) level, which addresses complex collaborative tasks which can be decomposed into elementary collaborative tasks (supported by the functionality/services at the generic collaborative services and/or at the generic services level). Examples of such services are (Campos et al. 2007):
  - collaborative problem solving,
  - collaborative intelligent monitoring,
  - collaborative reconfiguration of assembly and manufacturing systems (AMS)/products.

The classification of services is made taking into account the 'collaborativity' degree, or the notion of complexity of collaborative task.

Building Block Level Reference Architecture
This view refines what was previously presented at user level view. Like the user view, it is conceptual, but it details the structure of the architecture, including some high-level internal components (see Figure 2):

- Service Management Module, which offers an environment for service execution.
- UpperWare, or Orchestration Module: this coordinates the orchestration of CCS into CAS.
- Middleware: takes care of data arrived from legacy systems, AMS or application-specific software, making sure to format it according to the CCS needs.

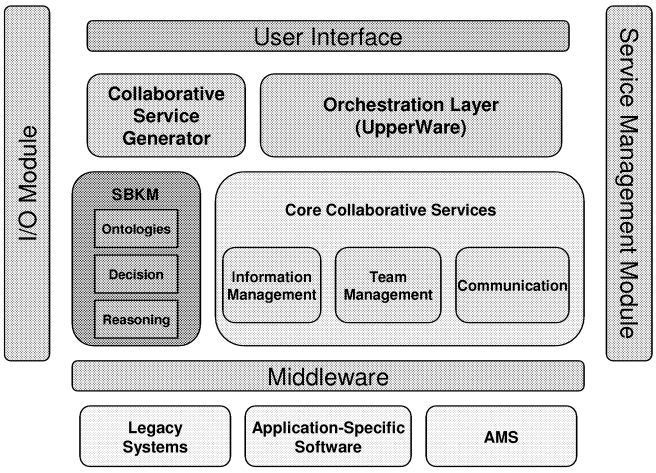


Figure 2: Building block reference architecture.

In order to maximise the re-usability of CCS, they have been grouped by elementary collaborative functionality:

- Information management group, which includes:
  - communication,
  - knowledge sharing.
- Team management group, for instance:
  - resource discovery,
  - team building,
  - collaboration traceability.
- Communication group.

Software Interface Level Architecture
This is the lowest level view in the RA. It refines the previous two views and helps identify the implementation building blocks. As this view concentrates on the "how" aspect, some components are ubiquitous (for instance, I/O occurs both at low level, delivering data used by the CCS, and at the high level, through GUIs which expose the CAS). Figure 3 presents this view detailing the Service Management module, the Team Management Services, the Information Management services as well as the SBKM Services.

Possible implementation technologies

The services present in the proposed platform will be described next, together with some details regarding potential technologies/tools to be used in their implementation. The complete system is developed using Java, and is deployed as an application in an application server (JBoss AS 4.0.5, open source software, licensed under LGPL).

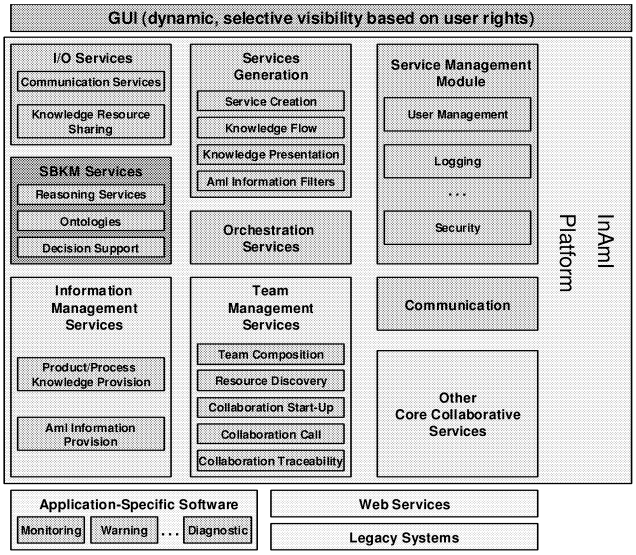


Figure 3: Proposed Software interface level reference architecture.

I/O Services
The I/O Services module encompasses, as noted before, I/O services without a collaborative aspect. These are basic services for communication and knowledge sharing, relying on established tools such as email, chat programs, video conferencing tools, etc. The proposed architecture uses existing open source implementations of such tools, developed using the Java programming language for easier adaptation and connection with the other modules in the system.

SBKM Services
These services offer knowledge management functionality. As such, they often require specialised tools for a certain aspect of the knowledge management process.
The reasoning services rely on case-based reasoning. For this purpose, during implementation an existing open source tool was used: jColibri, a Java implementation of a CBR tool. This was extended with an algorithm for estimating problem similarity.
Another used tool is Protégé, for ontology definition and extension.

Core Collaborative Services
The AmI Information Provision service maintains, for the purposes of enriching information at request, an ontology. This ontology is based on the SOUPA ontology; the tool used to define it is Protégé. Other technologies to be used by the AmI Information Provision service are JADE agents.
The Core Collaborative Services will be exposed as web services so that they can be composed into Collaborative

Application Services. These web services are SOAP compliant and use the WSDL (Web Service Description Language) standard.

Orchestration Services for Service Generation
Within the proposed architecture Collaborative Application Services are composed from CCS through web service orchestration. The BPEL standard is used, being one of the most important service orchestration language specification. It is based on XML, and specifies composite services in terms of atomic services (named activities) and the appropriate control dependencies in order to achieve the desired control flow. BPEL defines a notation for specifying business process behaviour which is based on existing functionalities exposed as web services. When defining a BPEL process, the process developer basically defines a web service which is a composition of existing web services.
A BPEL process basically consists of steps, also called *activities*. Both primitive (atomic) activities and composed activities are supported. Using BPEL, the process developer formally defines a business process, the output being the *process definition*.

Service Management Module
This module offers a frame within which various auxiliary functionalities are being dealt with. Like the rest of the system, it is implemented in Java, and offers user management functionality, logging of activity within the system, security in the form of user authentication and various user access rights, as well as an environment for the actual execution of services.
First, the module allows coping with user rights, allowing or disallowing a certain user access to a particular part of the system. It allows to integrate into the application a granular, complete and manageable permission scheme. The authorisation and authentication of users as done by this module enables the following:

- management of users, groups and entities,
- definition of a fine-grained permission
- associate a user or group with an entity, by assigning each pair of user/group and entity a role.

The module also supports the *Collaboration Traceability* CCS, by offering logging of user activity and interaction. Although the interesting for the Collaboration Traceability CCS are the user actions regarding collaboration, the logging in this module is more detailed and it logs every user interaction with the system.
On the other hand, the Service Management Module also involves the actual execution environment required for services, as seen from the lowest level – e.g. the services are deployed in an application server (JBoss) and developed in Java.

APPLICATIONS

The proposed architecture is applied in several manufacturing companies to support different aspects of collaborative work within an extended enterprise concept. Specifically it is being applied in a collaborative platform and it is being tested within three Business Cases (BC) as shortly described bellow.

Business Case 1

A company designs, manufactures and trades air conditioning equipment from 6 kW up to 600 kW under its own brand name and also as an original equipment manufacturer for other brands. The company needs to continuously find the best compromise between product quality, market attractiveness and price. Maintenance, after-sales services and production efficiency of provided equipments are key issues for the company. The company has clearly identified that the collaboration with clients must be strengthened to understand their needs, and to acquire information from the unit in its working environment since this information is vital to offer new services required by clients.
This company has identified two different scenarios, or use cases, for the direct application of the system implementing the RA. In the first use case, oriented on the support of acclimatisation units, in distinct geographical locations interact simultaneously to solve a problem in the machine. In the second use case, oriented on product design and reconfiguration, the product designer accesses information and may request further inputs from different sources (distributed inside the manufacturer's premises and external – the client) to obtain knowledge about the product.

Business Case 2

This business case involves a company that designs and delivers AMS for the automotive industry worldwide and another company who provides network service solutions. The objective is to radically improve collaborative work among the lines vendor and its end-users. These companies have established different forms of collaborative work with their customers, but the intention is to set up a much tighter and human-centric collaboration. They need a highly adaptive and powerful platform for dynamic improvements of services supporting collaborative work with their end-users.
These two companies have identified three application scenarios for the direct use of the developed platform. One of the application scenarios addresses an application service for problem solving within complex assembly lines. The target platform is applied to support collaboration among the maintenance services provider team at the shop-floor, service team at the provider of AMS and team(s) at the provider of the control system of AMS. The core collaborative services provided are combined with the maintenance tools currently used by the vendor of AMS. Another use case approaches the intelligent monitoring of a machine requested by the maintenance expert in collaboration with the machine vendor and the controller vendor. The process involves the collaborative definition of warning levels and collaborative analysis of data and trends, and results in a maintenance action. Finally the third use case addresses the reprogramming of production units within the AMS. The core collaborative services are used to support the collaboration of actors in the design and implementation of a new feature or the extension of an existing one in a certain production unit.

Business Case 3

This business case involves a company that manufactures cylindrical and flat surface grinding machines in the aeronautical, iron and steel, oil and general machining sectors. The objective of this business case is to improve intelligent monitoring of the critical elements, components or subsystems and provide a better collaborative maintenance services embedded in the machines, oriented to collaborative problem solving.
This company has identified two use cases. The first one addresses collaborative intelligent monitoring and the second collaborative problem solving. The first application service aims at defining rules for an intelligent monitoring system that will be installed in several machines in order to optimize monitoring of machines, discovering the necessity of a flexible and easy collaboration approach between all actors, and supply a competitive advantage over the competitors. In the second application scenario the aim is to define a collaborative procedure oriented to minimize maintenance cost and time spent on maintenance tasks and the shut-down cost and time.

Application scenarios

Summarising, the application scenarios identified in the above described business cases were the following:

- collaborative problem solving around AMS (e.g. searching for reasons of problems within manufacturing processes)
- intelligent monitoring of processes (collaborative definition of parameters, warning levels, rules to monitor processes)
- reconfiguration/reprogramming of AMS/product.

Within all these applications, the developed platform effectively supports collaborative work of different teams: shop-floor workers around AMS, maintenance teams, designers of AMS, control system providers etc.
In particular, the scenario for reprogramming of an AMS provides support for the user to:

1. identify and describe the need for a reprogramming of an AMS.
2. Identify and suggest the most suitable resources to support in reprogramming of the AMS.
3. Suggest a team to make the reconfiguration/ reprogramming of the AMS/product, if applicable.
4. Initiate the collaboration process to reconfigure/ reprogram the AMS/product.
5. Functionality to provide statistical information on the most frequent problems occurred in the production unit that is being reprogrammed.
6. Reporting on the reprogramming situation, where the collaborative work is documented for later traceability and future reference.

Middleware solutions

Middleware solutions for the current BCs include interfaces to AMS and other systems to collect information on AMS, products and processes. Different solutions for the middleware are under testing starting from different customised company–specific solutions (BC1) and web based solutions (such as Siemens' ePS platform in BC2, EPS 2005), agent-based solution (BC3), up to generic solution - so-called Generic Data Server providing data flow between the lower and higher control levels of an AMS (InAmI 2005).
A prototype of the targeted platform is currently being tested indicating that a very effective support of collaborative work among the teams involved in life cycle management of AMS can be provided by the targeted solution. The tests will be carried out extensively using real data from the three business cases presented.

BENEFITS

The advantages of complying to a unified RA become evident in all stages of application development. First, it makes for a more efficient design of the system, having already a logical structure within which to work. It also allows easier re-use of the same software solutions in different application scenarios or industrial environments, as demonstrated in the above description of the targeted business cases. The application's degree of complexity can be minimised, supporting at the same time reusability of modules (due to their standard, RA-complying interface) in similar applications even in different industrial sectors such as construction industry, for instance. The architecture of the system becomes more transparent and easier to understand and maintain, which helps with interoperability of modules developed in different applications complying with the RA. The interaction between the application's modules is clearer and better structured, enabling higher traceability between the abstract, initial requirements and the final implemented application.
All the listed technical benefits have the effect of reducing design costs and time, making for a more efficient design of the application and allowing for expenses reduction due to the reusability of the modules.

CONCLUSIONS

The paper represents one of the first comprehensive attempts to propose RA for CWE in manufacturing industry. A new architecture for CWE is developed and effectively applied in industry.
The technical requirements for a RA to support eCollaboration in the manufacturing industry have been studied, along with the current work being done in this direction. The study of several current initiatives working in the direction of a potential RA has helped to identify common features that a RA should present. The architectures used for comparison are the one proposed by the C@R project (initial OCA WG architecture), the architecture elaborated in the CoSpaces project and the ECOSPACE

project architecture, as well as Expert Group proposal for Collaborative Infrastructure.

The contributions to a potential RA for eCollaboration have been partially generated by the described attempts to reduce the proposed architecture elements to their common abstract denominator. The three RA views have been used for this purpose, illustrating the proposed architecture as seen through them.

Further contributions to the elaboration of a RA for CWE are represented by the identified requirements on CWE architectures from the point of view of the manufacturing industry.

The implementation of the proposed architecture has also been described in general lines, describing the overall functionality of the main subsystems, together with brief references to potential technologies to be used.

The proposed architecture can be effectively re-used for collaborative work not only in manufacturing industry. Especially the developed CCS can be applied in different domains. In fact it is likely that the above described RA and CCS will be reused in the scope of several projects aiming at small companies in the area of process control (e.g. control of weighing systems) and in the construction area as well as in collaborative design processes in automotive and aerospace industry.

ACKNOWLEDGEMENT

The work presented is partly carried out in the scope of the current RTD project InAmI, which is funded under the Sixth Research Framework Program of the European Union (contract FP6-2004-IST-NMP-2-16788).

REFERENCES

Campos A., Stokic D., Correia A., Aguirre-Ibarbia J., Pina P., Neves-Silva R. 2007, Supporting eCollaboration in Manufacturing Industry, eChallenges2007.

CoSpaces 2007, http://www.cospaces.org/

C@R 2007, http://c-rural.eu/

Decker S., 2006, "Towards a Reference Architecture for Collaborative Environments", ECOSPACE Newsletter, No.1, (http://www.ip-ecospace.org), (October).

Dornier GmbH, 1992, A&R Control Development Methodology Definition Report, (Doc. Nr. CT2/CDR/DO Issue 2.0, Sept. 1, 1992).

ECOSPACE 2007, http://www.ip-ecospace.org/

EPS 2005, EPS Siemens Network Services: http://www.automation.siemens.com/mc/mc-sol/en/eb90ca71-fe86-4045-b1f2-4653c2c30263/index.aspx.

Expert group, 2006, "Towards Activity-oriented Collaborative Working Environments – A Research Roadmap 2007 – 2020". Responsible EC officer and editor: Isidro LASO-BALLESTEROS. DG Information Society. European Commission. Rapporteurs: Nikolay Mehandjiev and Dragan Stokic, (June).

InAmI 2005, Presentation of InAmI Project. http://www.inami.eu

Ralli C., 2006, Open Collaborative Architecture Working Group (OCA WG), Networking workshop Collaboration Upperware for New Working Environments, IST Event, Helsinki.

Stokic D., U. Kirchhoff, J. Cordes, A. Elfving, P. Putz, 1994, Definition of reference architecture for control systems of mobile robots, The 10th ISPE/IFAC International Conference on CAD/CAM, Robotics and Factories of the Future, Ottawa (August).

Stokic D., 2006, "A New Collaborative Working Environment for Concurrent Engineering in Manufacturing Industry", CE 2006, 13th ISPE International Conference on Concurrent Engineering, Juan-les-Pins, France.

TOGAF 2007, http://www.opengroup.org/togaf/

WORKFLOW SIMULATION

PARTIAL RECOGNIZING ALGORITHM FOR VERIFICATION OF WORKFLOW PROCESSES

Armen Kostanyan
Anna Varosyan

Educational and Research Center of Information Technologies
Yerevan State University
1 A.Manoogian Str., Yerevan, Armenia, 0049
Email: armko@ysu.am, annavarosyan@yahoo.com

KEYWORDS
Model analysis, verification, approximation techniques.

ABSTRACT

The paper presents the formal definition of workflow process, its semantics and correctness assertion. A partial recognizing algorithm for formal verification of acyclic workflow processes is suggested. The algorithm uses molds that represent upper and lower approximations of environment state sets. The consideration of mold transformations on all branching paths allows to avoid the process execution and examination per each input. At the same time the use of molds makes the verification algorithm as a partially recognizing one: there is a possibility that the assertion about the workflow process correctness is neither confirmed nor rejected. The algorithm is applied to verify the workflow library processes provided by IBM for IT Infrastructure.

INTRODUCTION

Business Process Management is a key technology in the overall enterprise strategy of an organization (Leymann and Roller 2000). The business processes have to be specified by a language called workflow specification language (IBM 2001, MQSeries Workflow). It is very important to verify the business process correctness prior to the operation (Khoshafian 2004). The business process correctness can be checked by the process graph structural verification, such as cycle analysis, branching path synchronization and so on. The formal verification can be considered as another case of verification to verify the process semantic correctness. The formal verification requires knowledge about the underlying business process (internal structure of tasks, data flow, etc.). Two invariant conditions have to be specified for the process. First, the process precondition has to be satisfied prior to the process execution. Second, the postcondition has to be checked on each process execution end. The process is considered correct if the postcondition is of "true" value for all possible executions.

Currently, the most approaches to validate workflow systems have restrictions on workflow control structures that reduce the verification complexity. An idea of limiting control flow graph size gives linear complexity for the verification paradigm on a transactional logic based workflow (Davulcu et al. 1998). Another idea is proposed for formally modeling workflows by Petri nets. It is shown that when verifying certain soundness properties a well structured or free-choice (Van der Aalst et al. 2000) control flow leads to a polynomial verification complexity. There were many results in the area of workflow graph structural verification, such as the lack of synchronization, presence of deadlocks and cycles without exit. Most of results have used verification methods based on calculus (Xu and Yu 2007), graph search techniques (Perumal and Mahanti 2007), model checking (Pfeiffer et al. 2004) and propositional logic (Bi and Zhao 2004). However, the main interest was in general control structure properties, while the semantics of each specific task was not considered. No researches have been done by now on the verification to cover every aspect of workflow systems.

In this paper a method of partial verification of acyclic workflow processes is suggested that is based on molds (primitives for the description of environment state subsets). The idea of the method was applied to the verification of linear programs processing Boolean data (Kostanian 1995.) and tests for computational system components (Shoukourian et al. 1996, Shoukourian and Kostanian 1997). The mold technique also has been used for the verification of special type object-oriented programs (Kostanyan and Asatryan 2001).

The paper presents the formal definition of workflow process, its semantics and correctness assertion. Also, an extension of workflow process semantics is constructed that allows to use an acyclic workflow process control graph for the stepwise transformation of an initial mold into a final one. A conclusion about the workflow process partial correctness is made based on the match of final mold to the postcondition.

W-PROCESS

Definitions

The short description of formal workflow/BPM model which underlies a model of IBM's MQSeries Workflow (Leymann and Roller 2000), is presented in this section.

The main model components are ***activities*** and ***connectors***. The activities are associated with a context being defined

as data passing to an activity. It is called ***input container***. An activity also returns data called ***output container***. Some output container elements of activity can be passed to the input container elements of other activities or to the ***external memory***. All data elements are collected in the set **V.** Control and data connectors provide connections between activities. A control connector has an associated Boolean predicate called ***transition condition***.

A directed graph based on sets of activities and control connectors is called ***control flow*** of workflow process (***W-process***). The execution of W-process can be described in following general stages. At first, initial activities have to be executed and transition conditions of their outgoing control connectors should be calculated. Then, the set of ready activities have to be determined by the selecting activities whose all incoming transition conditions have been calculated. The most priority activity should be selected from this set, its join condition has to be calculated and the activity must be executed in the case of join condition truth-value. The execution of W-process will be stopped if the set of ready activities is empty.

Let give the formal definition of non-hierarchical workflow processes using following designations:

- If $\mathbf{x}=<\mathbf{x}_1,\ldots,\mathbf{x}_n>$ is a tuple, and $\mathbf{1}\le\mathbf{i}_1,\ldots,\mathbf{i}_k\le\mathbf{n}$ is a set of indices, then $\pi_{i1,\ldots,ik}(\mathbf{x})$ denotes the tuple $<\mathbf{x}_{i1},\ldots,\mathbf{x}_{ik}>$.
- If **X** is a set, the $\wp(\mathbf{X})$ denotes the set of all subsets of **X.**

Definition(W-process) A tuple $\mathbf{P}=<\mathbf{T}, \mathbf{N}, \mathcal{C}, \mathbf{E}, \Phi, i, o, \Psi, \Delta, \mathcal{M}>$ is called ***W-process***, if

1. **T** is a finite set of ***types*** having own set of values. The set of values for $\mathbf{t}\in\mathbf{T}$ is denoted by **DOM(t)** and is called the ***domain*** of **t**. Each variable used in the model belongs to any type from **T**. Let designate the set of **v** variable values by **DOM(v)** or by **DOM(t)** if **v** belongs to the type **t**.
2. **N** is a finite set of ***activities*** having some priorities.
3. $\mathcal{C}$ is the finite set of ***transition conditions***.
4. The set $\mathbf{E}\subseteq\mathbf{N}\times\mathbf{N}\times\mathcal{C}$ is the set of ***connectors*** with the following restrictions
   - **E** is *unified*: $\forall\mathbf{e},\mathbf{e'}\in\mathbf{E}: \pi_{1,2}(\mathbf{e})=\pi_{1,2}(\mathbf{e'})\Rightarrow \pi_3(\mathbf{e})=\pi_3(\mathbf{e'})$.
   - The **G=(N, E)** graph is acyclic.
5. $\Phi: \mathbf{N}\to\mathcal{F}$, where $\mathcal{F}$ is the set of all activity join conditions. The ***join condition of activity*** $\mathbf{A}\in\mathbf{N}$ is defined as a Boolean expression $\Phi(\mathbf{A})$ depending on elements of set $\mathbf{A}^{\leftarrow}=\{\mathbf{e}\in\mathbf{E} \mid \pi_2(\mathbf{e})=\mathbf{A}\}$. Suppose $\Phi(\mathbf{A})\equiv\mathbf{1}$, if $\mathbf{A}^{\leftarrow}=\varnothing$.
6. $i: \mathbf{N}\cup\mathcal{C}\to\mathcal{IV}$ is ***input container*** map where $\mathcal{IV}$ is the set of input variable tuples of activities and conditions. Each input variable is unique and belongs to any type from **T**.
7. $o: \mathbf{N}\to\mathcal{OV}$ is ***output container*** map where $\mathcal{OV}$ is the set of output variable tuples of activities. Each output variable is unique and belongs to any type from **T**.
8. $\Psi: \mathbf{N}\cup\mathcal{C}\to\mathcal{E}$, where $\mathcal{E}$ is the set of activities and conditions of all possible implementations.
   The ***implementation of activity*** $\mathbf{A}\in\mathbf{N}$ is defined as a map $\Psi(\mathbf{A}): \times_{\mathbf{v}\in i(\mathbf{A})}\mathbf{DOM(v)}\to\times_{\mathbf{v}\in o(\mathbf{A})}\mathbf{DOM(v)}$. The ***implementation of condition*** $\mathbf{p}\in\mathcal{C}$ is defined as a map $\Psi(\mathbf{p}): \times_{\mathbf{v}\in i(\mathbf{p})}\mathbf{DOM(v)}\to\{\mathbf{0},\mathbf{1}\}$.
9. $\Delta: \mathbf{N}\times(\mathbf{N}\cup\mathcal{C}\cup\{\mathcal{M}\})\to\cup_{\mathbf{A}\in\mathbf{N},\ \mathbf{B}\in\mathbf{N}\cup\mathcal{C}}\wp(o(\mathbf{A})\times(i(\mathbf{B})\cup\mathcal{M}))$ is ***data connector map*** with following restrictions:
   a. $[\mathbf{A}\in\mathbf{N}, \mathbf{B}\in\mathbf{N}\cup\mathcal{C}]\Rightarrow\Delta(\mathbf{A},\mathbf{B})\subseteq o(\mathbf{A})\times i(\mathbf{B})$.
   b. $\mathbf{A}\in\mathbf{N}\Rightarrow\Delta(\mathbf{A},\mathcal{M})\subseteq o(\mathbf{A})\times\mathcal{M}$.
   c. $[\mathbf{A}\in\mathbf{N}, \mathbf{B}\in\mathbf{N}\cup\mathcal{C}\cup\{\mathcal{M}\}, <\mathbf{v1},\mathbf{v'}>\in\Delta(\mathbf{A},\mathbf{B}), <\mathbf{v2},\mathbf{v'}>\in\Delta(\mathbf{A},\mathbf{B})]\Rightarrow\mathbf{v1}=\mathbf{v2}$.
   d. $\forall<\mathbf{v},\mathbf{v'}>\in\Delta(\mathbf{X},\mathbf{Y}), \mathbf{DOM(v)}=\mathbf{DOM(v')}$.
10. $\mathcal{M}$ is an ***external memory***. Each variable from $\mathcal{M}$ is unique and belongs to any type from **T.**

Let $\mathbf{A}^{\rightarrow}=\{\mathbf{e}\in\mathbf{E} \mid \pi_1(\mathbf{e})=\mathbf{A}\}$ be the set of outgoing transition conditions from the activity **A.**

Suppose **P** is a workflow process and $\mathbf{X}\in\mathbf{N}\cup\mathcal{C}$. Let define the ***set of all states for*** **X** as $\mathbf{States(X)}=\times_{\mathbf{v}\in i(\mathbf{X})}\mathbf{DOM(v)}$, and the ***set of external memory states*** as $\mathbf{States}(\mathcal{M})=\times_{\mathbf{v}\in\mathcal{M}}\mathbf{DOM(v)}$. Let call $\mathbf{B}=\times_{\mathbf{X}\in\mathbf{N}\cup\mathcal{C}\cup\{\mathcal{M}\}}\mathbf{States(X)}$ as a ***set of states for*** **P** and designate as $\mathbf{b(X)}$ the state of $\mathbf{X}\in\mathbf{N}\cup\mathcal{C}\cup\{\mathcal{M}\}$ for $\mathbf{b}\in\mathbf{B}$.

Suppose $\mathbf{b}\in\mathbf{B}$, $\mathbf{A}\in\mathbf{N}$. Let define the result of applying activity **A** to the state **b** as a state **bA=b'** which can be presented as following:

$$(\mathbf{b}'(\mathbf{X}))_{\mathbf{v'}}=\begin{cases}(\mathbf{b}(\mathbf{X}))_{\mathbf{v'}}, \text{ if } \forall\mathbf{v}\in o(\mathbf{A})\ [<\mathbf{v},\mathbf{v'}>\notin\Delta(\mathbf{A},\mathbf{X})],\\ (\Psi(\mathbf{A})(\mathbf{b}(\mathbf{A})))_{\mathbf{v}}, \text{ if } <\mathbf{v},\mathbf{v'}>\in\Delta(\mathbf{A},\mathbf{X}),\end{cases}$$

for all $\mathbf{X}\in\mathbf{N}\cup\mathcal{C}\cup\{\mathcal{M}\}$.

Let present an algorithm that describes the recursive semantics of W-process assuming that each activity has a unique priority. In general, the W-process semantics might be based on the nondeterministic choice of an activity from a set of activities having the same priority. It will make the semantic determination algorithm as nondeterministic. After some modifications the suggested approach of the process verification can be adjusted to this case.

Semantics of W-Process

Each W-process **P** calculates a mapping $\Psi(\mathbf{P}): \mathbf{B}\to\mathbf{B}$, defining by the following algorithm.

Algorithm S. Semantics of W-process.
Input: Initial state $\mathbf{initState}\in\mathbf{B}$.
Output: Final state $\mathbf{finState}\in\mathbf{B}$.
Remark: Creates an initialization and calls **SR**.
Method:

S1. For all $\mathbf{A}\in\mathbf{N}$ do
   - $\mathbf{counter(A)}\leftarrow|\mathbf{A}^{\leftarrow}|$.
   - If $\mathbf{counter(A)=0}$, then insert **A** into the activities priority queue **front** that is sorted in non-decreasing order of priorities.

S2. For all $\mathbf{e}\in\mathbf{E}$ assume that $\mu(\mathbf{e})$ is undefined.
S3. Return $\mathbf{SR(initState, front, counter}, \mu)$.

Algorithm SR.
Input:

- Current state $\mathbf{currState}\in\mathbf{B}$,
- priority queue **front** of activities,

- mapping **counter**: $\mathbf{N}\to\{\mathbf{0}, \mathbf{1}, \mathbf{2}, \ldots\}$,
- partially defined mapping $\mu\subseteq \mathbf{E}\to\{\mathbf{0}, \mathbf{1}\}$.

Output: Final state **finState**$\in$**B**.
Method:

SR1. If **front** is empty, then return **currState**.
SR2. Select most priority activity **A**$\in$**front** and remove **A** from **front.**
SR3. If $\Phi(A)(\langle\mu(e)|\ e\in A^{\leftarrow}\rangle)=0$, then
- nextState$\leftarrow$ **currState**.
SR4. Else, if $\Phi(A)(\langle\mu(e)|\ e\in A^{\leftarrow}\rangle)=1$, then
- nextState$\leftarrow$ **(currState) A**.
- For all $e\in A^{\rightarrow}$, **e=(A, A', p)** do a)-c)
a) $\mu(e)\leftarrow\Psi(p)$**(nextState)**,
b) **counter(A')** $\leftarrow$ **counter(A')-1**,
c) If **counter(A')=0**, then
insert **A'** into **front.**
SR5. Return **SR(nextState**, **front**, **counter,** μ**)**.

Let designate Ψ**(P)(b)=bP** and call the W-processes $\mathbf{P_1}$ and $\mathbf{P_2}$ equivalent, if $(\mathbf{bP_1})(\mathcal{M})=(\mathbf{bP_2})(\mathcal{M})$.

Correctness of W-Process

Definition Let α, β are two predicates depending on variables from $\mathcal{M}$. Let call α and β a precondition and a postcondition correspondingly. In this case, **P** is correct with respect to α and β, if

$$(\forall \mathbf{b}\in\mathbf{B})\ [\ \alpha(\mathbf{b}(\mathcal{M}))=1\ \Rightarrow\ \beta((\mathbf{bP})(\mathcal{M}))=1].$$

EXTENSION OF SEMANTICS OF W-PROCESS

Definitions

Let extend semantics of W-processes in following way. Suppose $\mathbf{s}\subseteq\mathbf{B}$, $\mathbf{A}\in\mathbf{N}$ and define an action of activity **A** applied to **s** as a set $\mathbf{sA=s'}\subseteq\mathbf{B}$ such that

$$(\mathbf{s'}(\mathbf{X}))_{\mathbf{v'}}=\begin{cases}(\mathbf{s}(\mathbf{X}))_{\mathbf{v'}}, \text{ if } \forall \mathbf{v}\in o(\mathbf{A})\ [<\mathbf{v}, \mathbf{v'}>\notin\Delta(\mathbf{A}, \mathbf{X})],\\ (\Psi(\mathbf{A})(\mathbf{s}(\mathbf{A})))_{\mathbf{v}}, \text{ if } <\mathbf{v}, \mathbf{v'}>\in\Delta(\mathbf{A}, \mathbf{X}),\end{cases}$$

for all $\mathbf{X}\in\mathbf{N}\cup\mathcal{C}\cup\{\mathcal{M}\}$.

Let introduce some designations. Suppose $\mathbf{p}\in\mathcal{C}$ and $\mathbf{supp(p)}=\{\mathbf{b}\in\mathbf{B}|\ \Psi(\mathbf{p})(\mathbf{b})=1\}$. Let $\mathbf{F}(\mathbf{x_1}, \ldots, \mathbf{x_n})$ be a Boolean expression depending on $\mathbf{x_1}, \ldots, \mathbf{x_n}$ and $\langle\mathbf{s_1}, \ldots, \mathbf{s_n}\rangle$ is a tuple of subsets of **B**. Let designate

$$F(s_1,\ldots,s_n)=\bigcup_{\{(\alpha_1,\ldots,\alpha_n)|F(\alpha_1,\ldots,\alpha_n)=1\}} s_1^{\alpha_1}\cap\ldots\cap s_n^{\alpha_n},$$

where $\mathbf{s^1=s}$, $\mathbf{s^0=B}\setminus\mathbf{s}$.

Let extend semantics of W-process that calculates a mapping $\mathbf{2^B\to 2^B}$ defining through the following algorithm.

Algorithm E. Extended Semantics of W-process.
Input: Initial subset **initSet**$\subseteq$ **B.**
Output: Final subset **finSet**$\subseteq$ **B.**
Remark: Creates an initialization and calls **ER**.
Method:

E1. For all $\mathbf{A}\in\mathbf{N}$ do
- **counter (A)** $\leftarrow|\mathbf{A}^{\leftarrow}|$.
- If **counter(A)=0**, then insert **A** into the activities priority queue **front** that is sorted in non-decreasing order of priorities.

E2. For all $\mathbf{e}\in\mathbf{E}$ assume that μ**(e)** is undefined.
E3. Return **ER(initSet**, **front**, **counter,** μ**)**.

Algorithm ER.
Input:
- Current subset **currSet**$\in$**B,**
- priority queue **front** of activities,
- mapping **counter**: $\mathbf{N}\to\{\mathbf{0, 1, 2, \ldots}\}$,
- partially defined mapping $\mu\subseteq\mathbf{E}\to\mathbf{2^B}$.

Output: Final subset **finSet**$\subseteq$**B**.
Method:

ER1. If **front** is empty, then return **currSet**.
ER2. Select most priority activity **A**$\in$**front** and remove **A** from **front**.
ER3. $\mathbf{s_1}=\Phi(\mathbf{A})(\langle\mu(\mathbf{e})|\ \mathbf{e}\in\mathbf{A}^{\leftarrow}\rangle)$, $\mathbf{s_0=B}\setminus\mathbf{s_1}$.
ER4. If $\mathbf{s_0}\neq\varnothing$, then do
- front'$\leftarrow$**front, counter'**$\leftarrow$**counter**, $\mu'\leftarrow\mu$.
- finSet$_0\leftarrow$ **ER(currSet, front', counter',** μ**')**.
Else, if $\mathbf{s_0}=\varnothing$, then $\mathbf{finSet_0}\leftarrow\varnothing$.
ER5.If $\mathbf{s_1}\neq\varnothing$, then do
- nextSet$\leftarrow$**(currSet) A**.
- For all $\mathbf{e}\in\mathbf{A}^{\rightarrow}$, **e=(A, A', p)** do a-c)
a) μ**(e)** $\leftarrow$ **nextSet** $\cap$ **supp**(**p**).
b) **counter(A')** $\leftarrow$ **counter(A')-1**.
c) If **counter(A')=0**, then
- insert **A'** into **front**.
- finSet$_1\leftarrow$**ER(nextSet, front, counter,** μ**)**.
Else, if $\mathbf{s_1}=\varnothing$, then $\mathbf{finSet_1}\leftarrow\varnothing$.
ER6. Return $\mathbf{finSet_0}\cup\mathbf{finSet_1}$.

Extended Semantics and Correctness of W-Process

Let α and β are precondition and postcondition for a W-process **P**. Let $\mathbf{B}(\alpha)=\{\mathbf{b}\in\mathbf{B}|\ \alpha(\mathbf{b})=1\}$, $\mathbf{B}(\beta)=\{\mathbf{b}\in\mathbf{B}|\ \beta(\mathbf{b})=1\}$. The following assertion holds.
Theorem 1
P is correct with respect to α and $\beta\Leftrightarrow\mathbf{P}(\mathbf{B}(\alpha))\subseteq\mathbf{B}(\beta)$.

APPROXIMATIVE CALCULATIONS VIA MOLDS

Definitions

Let Σ be a predefined set of **B** subsets ordered by the inclusion relation. Suppose that
- $\varnothing, \mathbf{B}\in\Sigma$,
- Σ is close with respect to intersection operations.

Suppose $\mathbf{s}\subseteq\mathbf{B}$. Let denote by $\lceil\mathbf{s}\rceil$ the ***unique minimum element of*** Σ ***containing*** **s**, and by $\lfloor\mathbf{s}\rfloor$ any ***maximum element of*** Σ ***containing in*** **s**. Let call a pair $\mathbf{mold}=\langle\sigma_1, \sigma_2\rangle$ so that $\sigma_1, \sigma_2\in\Sigma$, a ***mold*** and denote the set of all molds by **M**. Let use the designation Λ for mold $\langle\sigma_1, \sigma_2\rangle$, if $\sigma_1\supset\sigma_2$, and identify such a mold as $\langle\varnothing, \varnothing\rangle$.

Let $\mathbf{A}\in\mathbf{N}$. Define an action $\mathbf{A: M\to M}$ so that $(\mathbf{mold})\mathbf{A}=\langle\lfloor(\pi_1(\mathbf{mold}))\mathbf{A}\rfloor, \lceil(\pi_2(\mathbf{mold}))\mathbf{A}\rceil\rangle$ for all $\mathbf{mold}\in\mathbf{M}$. Similarly, for $\mathbf{p}\in\mathcal{C}$ define an action $\mathbf{p: M\to M}$

so that $(\mathbf{mold})p=\langle\lfloor\pi_1(\mathbf{mold})\cap\mathbf{supp}(p)\rfloor, \lceil\pi_2(\mathbf{mold})\cap\mathbf{supp}(p)\rceil\rangle$ for all $\mathbf{mold}\in\mathbf{M}$.

Finally, if $\mathbf{F}(x_1, \ldots, x_n)$ is a Boolean expression depending on $x_1, \ldots, x_n$ and $\langle\mathbf{mold}_1, , \ldots, \mathbf{mold}_n\rangle$ is a tuple of molds, then designate

$$F(mold_1,...,mold_n)=\bigvee_{\{(\alpha_1,...,\alpha_n)|F(\alpha_1,...,\alpha_n)=1\}} mold_1^{\alpha_1}\wedge...\wedge mold_n^{\alpha_n},$$

Suppose the operations of ***disjunction***, ***conjunction*** and ***negation*** over molds are defined in such way:

$\mathbf{mold'}\vee\mathbf{mold''}=\langle\lfloor\pi_1(\mathbf{mold'})\cup\pi_1(\mathbf{mold''})\rfloor, \lceil\pi_2(\mathbf{mold'})\cup\pi_2(\mathbf{mold''})\rceil\rangle$,

$\mathbf{mold'}\wedge\mathbf{mold''}=\langle\lfloor\pi_2(\mathbf{mold'})\cap\pi_2(\mathbf{mold''})\rfloor, \lceil\pi_2(\mathbf{mold'})\cap\pi_2(\mathbf{mold''})\rceil\rangle$,

$\neg\mathbf{mold}=\langle\lfloor\mathbf{B}\setminus\pi_2(\mathbf{mold})\rfloor, \lceil\mathbf{B}\setminus\pi_1(\mathbf{mold})\rceil\rangle$.

Let define a mapping $\Psi(\mathbf{P})$: $\mathbf{M}\to\mathbf{M}$ for W-process **P** that can be described by following algorithm.

Algorithm W. Approximative Calculations via Molds.

Input: Initial mold $\mathbf{initMold}\in\mathbf{M}$.

Output: Final mold $\mathbf{finMold}\in\mathbf{M}$.

Remark: Creates an initialization and calls **WR**.

Method:

W1. For all $\mathbf{A}\in\mathbf{N}$ do

- $\mathbf{counter\ (A)}\leftarrow|\mathbf{A}^{\leftarrow}|$.
- If $\mathbf{counter(A)=0}$, then insert **A** into the priority queue **front** of activities prioritized in non-decreasing order of priorities.

W2. For all $\mathbf{e}\in\mathbf{E}$ assume that $\mu(\mathbf{e})$ is undefined.

W3. Return $\mathbf{WR(initMold, front, counter}, \mu)$.

Algorithm WR.

Input:

- Current mold $\mathbf{currMold}\in\mathbf{M}$,
- priority queue **front** of activities,
- mapping $\mathbf{counter}$: $\mathbf{N}\to\{\mathbf{0, 1, 2, \ldots}\}$,
- partially defined mapping $\mu\subseteq\mathbf{E}\to\mathbf{M}$.

Output: Final mold $\mathbf{finMold}\in\mathbf{M}$.

Method:

WR1. If **front** is empty, then return **currMold**.

WR2. Select most priority activity $\mathbf{A}\in\mathbf{front}$ and remove **A** from **front**.

WR3. $\mathbf{mold_1}=\Phi(\mathbf{A})(\langle\mu(\mathbf{e})|\ \mathbf{e}\in\mathbf{A}^{\leftarrow}\rangle)$, $\mathbf{mold_0}=\neg\ \mathbf{mold_1}$

WR4. If $\mathbf{mold_0}\neq\Lambda$, then do

-front'$\leftarrow$**front, counter'**$\leftarrow$**counter**, $\mu'\leftarrow\mu$.

-finMold$_0\leftarrow$**WR(currMold,front',counter',**μ**')**.

Else, if $\mathbf{mold_0}=\Lambda$, then

- finMold$_0\leftarrow\Lambda$.

WR5. If $\mathbf{mold_1}\neq\Lambda$, then do

- nextMold$\leftarrow$**(currMold) A**.

- For all $\mathbf{e}\in\mathbf{A}^{\rightarrow}$, $\mathbf{e=(A, A', p)}$ do a-c)

a) $\mu(\mathbf{e})\leftarrow(\mathbf{nextMold})\ \mathbf{p}$.

b) $\mathbf{counter(A')}\leftarrow\mathbf{counter(A')-1}$.

c) If $\mathbf{counter(A')=0}$, then

- insert **A'** into **front'**.

- finMold$_1\leftarrow$**WR(nextMold, front, counter,** μ**)**.

Else, if $\mathbf{mold_1}=\Lambda$, then $\mathbf{finMold_1}\leftarrow\Lambda$.

WR6. Return $\mathbf{finMold_0}\vee\mathbf{finMold_1}$.

Let use the designation **(mold) P** for the value $\Psi(\mathbf{P})(\mathbf{mold})$.

Approximative Calculations and Correctness of W-Process

Let α and β are precondition and postcondition of W-process **P**, $\mathbf{supp}(\alpha)=\{\mathbf{b}|\ \alpha(\mathbf{b})=\mathbf{1}\}$, $\mathbf{supp}(\beta)=\{\mathbf{b}|\ \beta(\mathbf{b})=\mathbf{1}\}$.

Let $\mathbf{mold}(\alpha)=\langle\lfloor\mathbf{supp}(\alpha)\rfloor,\lceil\mathbf{supp}(\alpha)\rceil\rangle$,
$\mathbf{finalMold=mold}(\alpha)\ \mathbf{P}$,
$\mathbf{mold}(\beta)=\langle\lfloor\mathbf{supp}(\beta)\rfloor,\lceil\mathbf{supp}(\beta)\rceil\rangle$.

The following statement about the correctness of **P** with respect to α and β, based on comparison of basic sets, holds.

Theorem 2

a) If $\pi_2(\mathbf{finalMold})\subseteq\pi_1(\mathbf{mold}(\beta))$, then **P** is correct with respect to α and β;

b) If $\neg\ (\pi_1(\mathbf{finalMold})\subseteq\pi_2(\mathbf{mold}(\beta))$, then **P** is incorrect with respect to α and β;

c) Otherwise, the assertion about correctness of **P** with respect to α and β is undefined.

SPLITTING MOLDS

Let describe a possible determination of basis having increased effectiveness of Boolean operators on the molds. For this purpose a basic set $\mathbf{s}\in\Sigma$ has to be defined as a Cartesian product of variable value subsets generating the state of environment, in such way $\mathbf{s}\in\Sigma\ \Rightarrow\ \mathbf{s}=\underset{v\in\mathcal{V}}{\times}s(v)$, where

- $\mathcal{V}$ is an ordered set of *activity input variables*, *transition condition input variables* and *external memory* variables. Suppose the variables of $\mathcal{V}$ are ordered in some predefined order.
- $\mathbf{s}(v)\subseteq\mathbf{DOM}(v)$.

Notice that $\mathbf{s}=\varnothing$ if $\mathbf{s}(v)=\varnothing$ for some $v\in\mathcal{V}$.

Let call above described subsets of **B** ***splitting sets*** and molds constructed on the base of splitting sets, as ***splitting molds***. Let introduce an ordering relationship for the set of splitting subsets as: $\mathbf{s}\leq\mathbf{s'}$, if $\mathbf{s}(v)'\subseteq\mathbf{s}(v)$ for all $v\in\mathcal{V}$. Let **s'** and **s''** are splitting subsets, $v\in\mathcal{V}$. Let consider following operations:

- $s'\oplus s''=\underset{v\in\mathcal{V}}{\times}\left(s'(v)\cup s''(v)\right)$,
- $s'\otimes s''=\underset{v\in\mathcal{V}}{\times}\left(s'(v)\cap s''(v)\right)$,
- $s'\cup_p s''=\underset{v\in\mathcal{V}}{\times}s(v)$, where $s(v)=\begin{cases}s'(v)\cap s''(v), & v\neq p\\ s'(v)\cup s''(v), & v=p\end{cases}$
- $s'\setminus_p s''=\underset{v\in\mathcal{V}}{\times}s(v)$, where $s(v)=\begin{cases}s'(v), & v\neq p\\ s'(v)\setminus s''(v), & v=p\end{cases}$

Theorem 3 Let **s'** and **s"** are splitting sets, $s'=\underset{v\in\mathcal{V}}{\times}s'(v)$, $s''=\underset{v\in\mathcal{V}}{\times}s''(v)$.

The following assertions hold:

a) $\lfloor\mathbf{s'}\cap\mathbf{s"}\rfloor=\lceil\mathbf{s'}\cap\mathbf{s"}\rceil=\mathbf{s'}\otimes\mathbf{s"}$.

b) $\lfloor\mathbf{s'}\cup\mathbf{s"}\rfloor$ could be defined as

- **s' $\cup_p$ s"**, if $(\exists p \in \mathcal{V})$ **[s'(*p*)\s"(*p*)≠∅, s"(*p*)\s'(*p*)≠∅]**,
- **s'**, if **s"≤s'**, or
  $(\exists p \in \mathcal{V})$**[s"(*p*) ⊂s'(*p*) &**$(\forall v \in \mathcal{V}, v \neq p)$**[s'(*v*)⊆s"(*v*)]]**,
  or
  $(\exists p, q \in \mathcal{V}, p \neq q)$ **[s'(*p*)\s"(*p*)≠∅ & s'(*q*)\s"(*q*)≠∅]**,
- **s''**, if **s'≤s"**, or
  $(\exists p \in \mathcal{V})$ **[s'(*p*) ⊂s"(*p*) &**
  $(\forall v \in \mathcal{V}, v \neq p)$ **[s"(*v*)⊆s' (*v*)]]**,
  or
  $(\exists p, q \in \mathcal{V}, p \neq q)$ **[s"(*p*)\s'(*p*)≠∅ & s"(*q*)\s'(*q*)≠∅].**

c) $\lceil$**s' ∪ s"**$\rceil$ = **s' ⊕ s"**.

d) $\lfloor$**s' \ s"**$\rfloor$ could be defined as
- **s' \ $_p$ s"**, if $(\forall v \in \mathcal{V})$**[s'(*v*)∩s"(*v*)≠∅] &**
  $(\exists p \in \mathcal{V})$ **[s'(*p*)\s"(*p*)≠∅]**,
- **s'**, if a $(\exists p \in \mathcal{V})$ **[s'(*p*)∩s"(*p*)=∅]**,
- ∅, otherwise.

e) $\lceil$**s' \ s"**$\rceil$ is defined as
- **s'**, if $(\exists p, q \in \mathcal{V}, p \neq q)$ **[s'(*p*)\s"(*p*)≠∅,**
  s'(*q*)\s"(*q*)≠∅],
- $\lfloor$**s' \ s'**$\rfloor$, otherwise.

Thus, the manipulations over splitting sets that are implied by Boolean operators being applied to splitting molds (that is the approximations of union, intersection and subtraction), are reduced to the set-theoretic operations on the type domain subsets. In such way the cardinal number of considered sets is essentially decreased and the effectiveness of operator performing is increased correspondingly. Actually, instead of dealing with the sets having $\prod_{v \in \mathcal{V}} |s'(v)|$ and $\prod_{v \in \mathcal{V}} |s''(v)|$ cardinal numbers, we $|\mathcal{V}|$ times deal with the sets having cardinal numbers at most $\max_{v \in \mathcal{V}}(|s'(v)|, |s''(v)|)$.

The next question is: *How to increase the effectiveness of performing actions of activities and predicates on splitting molds?* For this purpose it is possible to use such implementations of activities and transition conditions that produce splitting molds to be applied to other splitting molds.

APPLICATION OF THE ALGORITHM

One of more popular frameworks is the IT Infrastructure Library (ITIL). IBM provides Process Reference Model for IT (PRM-IT) that supplements the content of ITIL based on IBM's extensive IT management experience (IBM 2007). IBM has developed a library with detailed (task-level) process workflows based on PRM-IT. It is also possible to build new processes based on given processes.

Let consider one of these processes and show the result of the algorithm application to the part of it (See Figure 1). Some changes have been made on it by replacing some part of the original process with "*Cancel Incident*". The correctness of template logic has to be verified. Let all input variables, except for external variables, have the value "*Unknown*". Assume all external input variables have value "*Known*". Suppose by implementing each activity the "*Known*" value is just passed for its outputs and <*yes*, *no*> for branching activities predicate variables are passed as well. The postcondition of this process part can be formulated by the condition **PostCon = (bCanceled OR (bWorkaroundOrFixProvided AND bIncidentResolutionPlanCreated)) AND bResolved** .

Applying the algorithm to the given process a result will be that the *activity 7* cannot be activated and the value of the variable *bResolved* for the *activity 7* and memory will remain "*Unknown*". As a result, the condition for incorrect processes will be satisfied (See Theorem 2). This sample illustrates a very simple model of the discussed flow with some shortening so the incorrect design (2 mutually exclusive execution paths with one synchronization point) can be easily found. The control connector between *activities 4* and *7* has to be removed to correct the modified template logic. The postcondition has to be also changed to the **PostCon = bCanceled OR (bWorkaroundOrFixProvided AND bIncidentResolutionPlanCreated AND bResolved)**. Applying the algorithm to the corrected process will be resulted in the satisfaction of the correct process condition (See Theorem 2).

At present, the works on the tool for applying the algorithm to all library processes are under development.

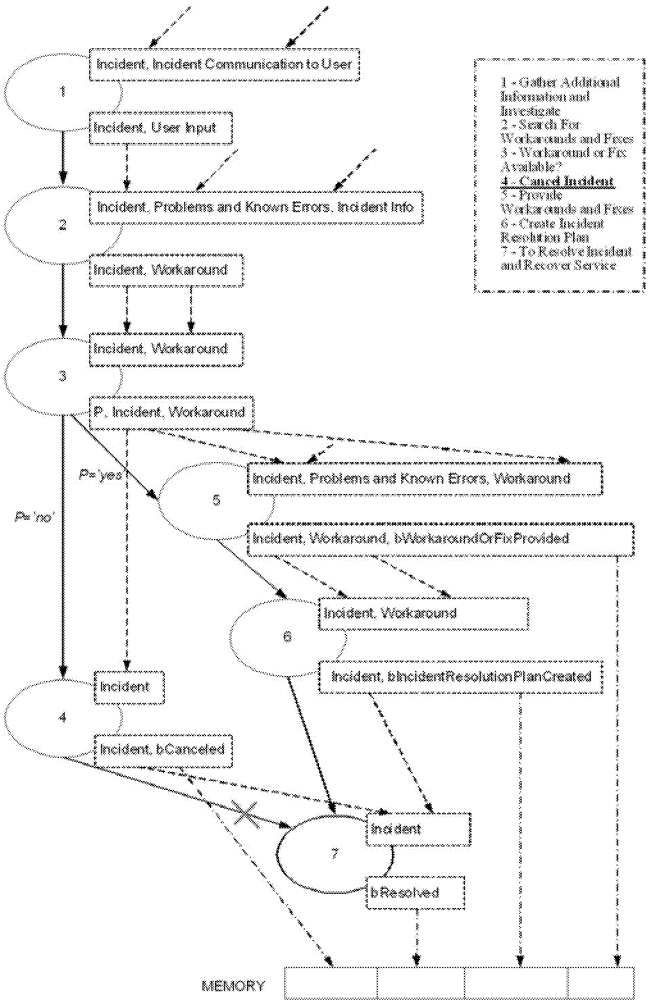


Figure 1: Investigate and Diagnose Incident

CONCLUSION

A partial recognizing algorithm for the formal verification of acyclic processes, which underlies IBM/Workflow model, is suggested. The algorithm uses the molds for upper and lower approximation of environment possible state sets. The consideration of mold transformations on all branching paths allows to avoid the process execution and examination per each input. The algorithm instead examines all branching paths through the molds. The weakness of the algorithm is the possibility of getting uncertain answer to the question regarding the program correctness. Additional investigations on specific cases should be performed to estimate the probability of such a result.

The presented approach is now used as a basis for the tool development. The tool is suggested to verify the processes for IMB PRM-IT library.

Works on analyzing the algorithm effectiveness through the applying the algorithm to some popular libraries and building more accurate approximations are in progress.

REFERENCES

Bi H.H.1 and Zhao J.L.2. 2004. "Applying Propositional Logic to Workflow Verification", *Information Technology and Management,* Volume 5, Numbers 3-4, July 2004 , pp. 293-318(26)

Davulcu, H.; M.Kifer; C.R.Ramakrishnan and I.V.Ramakrishnan. 1998. "Logic based modeling and analysis of workflows", *In ACM Symposium on Principles of Database Systems*, June 1998

IBM 2001. "IBM MQSeries Workflow: Concepts and Architecture", Version 3.3, GH12-6285-03, Product No. 5697-FM3, Mar. 2001, pp. iii-58.

IBM 2007. "IBM Process Reference Model For IT (PRMIT)", version 2.0, January 2007.

Leymann, F. and D. Roller. 2000. "Production Workflow: Concepts and Techniques", Prentice Hall Press (2000)

Khoshafian, S. 2004. "Intelligent Business Process Management for Service-Oriented Enterprises," *Business Integration Journal*, August 2004.

Kostanian, A. 1995. "A Problem of Verification for Linear Programs with Boolean Arrays", *Kibernetika i sistemny analiz*, Kiev, Ukraine, 1995, #5, pp. 87-94.

Kostanyan, A. and A. Asatryan. 2001. "Usage of Invariant Molds for Verification of a Particular Class of Object-oriented Programs", *Proc. of the Conf. "Computer Science & Information Technologies"*, Yerevan, Armenia, 2001, pp. 141-143.

Perumal, S. and A. Mahanti. 2007. "Applying Graph Search Techniques for Workflow Verification," *hicss*, p. 48c, *40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 2007

Pfeiffer, J.-H.; W. R. Rossak and A. Speck. 2004. "Applying Model Checking to Workflow Verification," *ecbs*, p. 144, *11th IEEE International Conference and Workshop on the Engineering of Computer-Based Systems (ECBS'04)*, 2004

Shoukourian, S.; A. Kostanian; V.Margarian and A.Ashour. 1995. "An Approach for System Tests Design and Its Applications", *Proc. of the 13-th IEEE VLSI Test Symposium*, NJ, Princeton, USA, 1995, pp. 448-453.

Shoukourian, S. and A.Kostanian. 1997. "Results on Justification of a Unified Test Design Technology both for Offline and Online Testing", *Proc. of the Conf. Computer Science & Information Technologies*, Yerevan, Armenia, 1997, pp. 326-329.

Van der Aalst, W. M. P. and A.H.M. ter Hofstede. 2000. "Verification of worfkow task structures: A petri-net-based approach", *Information Systems, 25-1:43-69*, 2000.

Xu,F. and Z.Yu. 2007. "A Workflow Verification Method Based on Calculus", *Proceedings of the First Joint IEEE/IFIP Symposium on Theoretical Aspects of Software Engineering*, pages 188-196, 2007

ACKNOWLEDGEMENTS

The authors express their acknowledgement to Prof. Samvel Shoukourian who inspired them to perform investigations in the area of workflow process formal verification. Authors also appreciate the International Scientific Technology Center- ISTC A1451 Project for supporting this research.

BIOGRAPHY

ARMEN KOSTANYAN is an associate professor of Yerevan State University. He is also a visiting associate professor of American University of Armenia. His main research interests are in the computation system analysis and design area. He participates in the "Development of Scientific Computing Grid on the Base of Armcluster for South Caucasus Region" and "Problems of Analysis and Optimization of Processes in Distributed Environments" research programs.

ANNA VAROSYAN is a lecturer of Yerevan State University. She is also a design engineer in Virage Logic International Corporation. Her research interests include verification and validation of languages and software tools. She participates in the "Development of Scientific Computing Grid on the Base of Armcluster for South Caucasus Region" research program.

MODELING TIME-CONSTRAINED COLLABORATIVE WORKFLOWS BASED ON TIMED RECURSIVE ECATNETS

Kamel Barkaoui
Awatef Hicheur

CEDRIC-CNAM Paris
292, Rue Saint-Martin
Paris 75003, France
E-mail: {Barkaoui | Hicheu}@Cnam.fr

KEYWORDS
Workflow modeling, Recursive Petri Nets, Timed Petri nets.

ABSTRACT

In this paper, we present the Timed Recursive ECATNets (T-RECATNets) formalism for the modeling and the analysis of time constrained flexible and collaborative workflows. Its concurrent semantics is expressed by means of real time rewrite theories. Therefore, we can benefit greatly from the use of the Real-Time Maude simulation tool for purposes of verification and evaluation.

INTRODUCTION

Nowadays, companies to stay competitive are more and more adopting the workflow management systems (WfMSs) technology, for automating, monitoring and improving their critical business processes (Aalst and Hee 2002). A workflow schema is used to represent the structure of a business process in terms of tasks as well as dependencies between tasks. Several workflow systems are available in the current marketplace but some limitations have been encountered in their design. Indeed only few of WfMSs provide efficient ways of managing flexible workflows which require dynamic reconfiguration of their structure at the occurrence of exceptional situations (Halliday and al. 2001). The current trend of business globalisation (where multiple parties belonging to different enterprises collaborate in a business process) brings new challenges regarding flexibility and collaboration of inter-organizational workflow. At one hand, the execution of such processes is distributed over multiple workflow engines (within the same organization or over multiple organisations). On the other hand, the structures of these collaborative processes are extremely dynamic (Sadiq and al. 2006). Another critical aspect to deal with in workflows is the *time* management. The practical need of the workflow designers' states that is useful to represent explicitly time related aspects of business processes (such activity durations, delays between activities or the assignment of deadlines) (Tiplea and Macovei 2005) (Lin and Qu 2004). Indeed, it is crucial to detect the violations of time constraints on workflow execution or to evaluate the overall process duration (Ling and Schmidt 2000). In regards to these previous points, we need a modeling formalism that offers mechanisms allowing us to manage dynamic structural changes and to describe both the dependencies between the process tasks and the distributed execution of the process over collaborative partners, and dealing with time constraints in an appropriate way. For this, we introduce in this paper the new model, namely, Timed Recursive ECATNets (T-RECATNets) as an extension of the Recursive ECATNets model (RECATNets) (Barkaoui and Hicheur 2007) for the modeling and analysis of time constrained flexible workflow systems. The advantage of the RECATNets, is that they offer practical mechanisms for direct and intuitive support of dynamic creation and suppression of processes. The current state of a concurrent system is expressed as a dynamical tree of threads where each one of them has its own execution context. The T-RECATNets inherit of all the modeling abilities of the RECATNets. We attach to each transition a time interval representing the minimum duration and the maximum duration of the action execution corresponding to the firing of transition. The semantics of Timed RECATNets described in terms of rewriting logic (Bruni and Meseguer 2006) by means of real time rewrite theories (Ölveczky and Meseguer 2007) permits the full use of the Real-Time Maude tool for purposes of verification and evaluation on the T-RECATNets while having a solid mathematical foundation.

RECURSIVE ECATNETS

The Recursive ECATNets model (Barkaoui and Hicheur 2007) is defined on the basis of a sound combination of the Extended Concurrent Algebraic Term Nets (Bettaz and Maouche 1992) and the Recursive Petri Nets (Haddad and Poitrenaud 2007). Extended Concurrent Algebraic Term Nets (ECATNets) are a kind of algebraic nets combining the expressive power of Petri nets and of abstract data types where places are marked with multisets of algebraic terms. In an ECATNet, each place p is associated to a sort s. Each input arc of a transition t is labeled by two expressions, respectively, $IC(p, t)$ (input condition) and $DT(p, t)$ (destroyed Tokens). Each output arc of transition t to a place is labeled by an expression $CT(p', t)$ (Created Tokens), and finally each transition t can be labeled by a boolean expression $TC(t)$ (Transition guard). The expression $IC(p, t)$ specifies the partial condition on the marking of the input place p for the enabling of t (Table 1). The expression $DT(p, t)$ specifies the multiset of tokens to be removed from the marking of the input place p when t is fired. The expression $CT(p', t)$ specifies the multiset of tokens to be created in the output place p', when t is fired. Finally, the expression $TC(t)$

is a boolean term which specifies an additional enabling extra condition for the transition t (when $TC(t)$ is omitted, the default value is the term *True*).

Table 1: the Forms of the Expression $IC(p, t)$

IC(p, t)	Enabling condition
α^0	The marking of the input place p must be equal to α ($IC(p, t) = \alpha^0$ means the marking of p must be empty)
α^+	The marking of the place p must include α ($IC(p, t) = \varnothing^+$ means condition is always satisfied)
α^-	The marking of the place p must not include α (with $\alpha \neq \varnothing$)
$\alpha_1 \wedge \alpha_2$	Conditions α_1 and α_2 are both true
$\alpha_1 \vee \alpha_2$	α_1 or α_2 is true

The Recursive ECATNets (RECATNets) inherit from the all the concepts of the ECATNets except that their transitions are partitioned into two categories as in recursive Petri nets (Haddad and al 2007): *abstract* transitions (Figure 2) and *elementary* transitions (Figure 3). To each abstract transition is associated an initial marking represented graphically in a frame. A family of boolean terms Υ is defined and associated to a RECATNet specifying the *termination conditions* (final markings) of the threads.

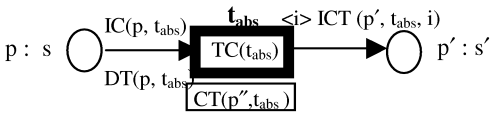


Figure 1: a Generic Abstract Transition

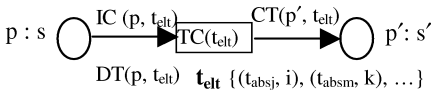


Figure 2: a Generic Elementary Transition

In RECATNets, there is a clear distinction between the firing condition of a given transition t and the tokens which may be destroyed during the firing action of t (respectively specified via the expression $IC(p, t)$ and $DT(p, t)$): a transition t (elementary or abstract) is fireable in a state s if and only if all its input conditions ($IC(p, t)$ $\forall p \in {}^\bullet t$) are satisfied (and $TC(t)$ is true) at this state s. The execution of a RECATNet generates a dynamical tree of threads (denoting the fatherhood relation) where each of these threads has its own activity. When a thread fires an abstract transition t_{abs}, it consumes the multiset of tokens $DT(p, t_{abs})$ from the input place p and simultaneously it creates a new thread child which starts its execution with the initial marking associated to this abstract transition. When a thread reaches a final marking Υi (with $i \in I$, where I is the set of the defined termination indexes), it terminates aborting its whole descent of threads. Then, it produces (in the token game of its father) and for the abstract transition t_{abs} which gave birth to it, the multiset of tokens $ICT(p', t_{abs}, i)$ in the output place p' of t_{abs}. Such a firing is called a cut step and denoted τi. An arc from an abstract transition t_{abs} to its output place p', labelled by an expression $<i>$ $ICT(p', t_{abs}, i)$ means that the tokens $ICT(p', t_{abs}, i)$ are produced in the place p', if the marking Υi is reached in the terminating thread child (the item i is the index of this termination). Such an arc can be omitted if the term $ICT(p', t_{abs}, i)$ is null. Note that if a cut step occurs in the root of the tree of threads, it leads to an empty tree denoted $\perp$. The behaviour of an elementary transition telt depends on a partial function K which associates to it a set of abstract transitions to interrupt and for each of these transitions a termination index. In the graphical representation of a RECATNet, the name of an elementary transition t_{elt} is followed by the set $K(t_{elt})$ when this set is non empty. If a thread fires an elementary transition t_{elt}, $DT(p, t_{elt},)$ is removed from the input place p and simultaneously $CT(p', t_{elt})$ is added to the output place p' of t_{elt}. Moreover, if the function k is defined, the firing of this transition performs the appropriate cut step to each subtree generated by the abstract transitions specified by K and depending of the termination index associated to it the output tokens of these abstract transitions are produced.

Definition 1 (Recursive ECATNets).
A Recursive ECATNet is a tuple RECATNet = *(ECATNet, I, Υ, K, ICT)* where:

- $T = T_{abs} \cup T_{elt}$ ($\cup$ denotes the disjoint union) is the set transitions partitioned into abstract and elementary ones (respectively);
- I is a finite set of indexes; Υ is a family, indexed by I, of Boolean terms defined in order to describe the termination conditions of threads.
- $K : T_{elt} \rightarrow T_{abs} \times I$, is a partial function which associates to an elementary transition the set of interrupted abstract transitions and their associated termination index.
- ICT: $P \times T_{abs} \times I \rightarrow$ CATdas(E, X), (Indexed Created Tokens).

TIMED RECURSIVE ECATNETS

In the extended version of RECATNets model with time specification, called Timed RECATNets (T-RECATNets), each elementary transition t is assigned a time interval $[t_{min}, t_{max}]$ giving the minimum and the maximum duration of the execution of the action corresponding to the firing of this transition. When an elementary transition t becomes enabled, it is immediately taken i.e the action of removing and creating tokens is done instantaneouslsly. However, the created tokens are not available to enable new transitions until they have been in their output place for a time θ ($t_{min} < \theta < t_{max}$) once minimum delay of the transition which created them has passed. The firing of abstract transitions and cut steps is also done instantaneously but no time intervals are assigned to them. T-RECATNets behave exactly like ordinary RECATNets. The enabling conditions of transitions and cut steps are unchanged and the execution of the net generates a *dynamical* tree of threads (denoting the fatherhood relation) where each of these threads is executed in parallel with the others threads of the tree. So, T-RECATNets have a notion of concurrent firing of multisets of transitions (including cut steps). The firing of an abstract transition in a thread of this tree initiates a new thread with the associated marking and the execution of a cut step prunes the subtree of the thread where the firing takes place. Note that the firing of an elementary transition prunes all the subtrees generated by one of the abstract transitions of the set specified in by the function K (if non empty). In the graphical

representation of T-RECATNets, the name of an elementary transition is followed by its duration interval (Figure 3).

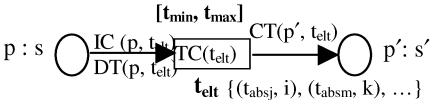


Figure 3: an Elementary Transition in T-RECATNets

Definition 2 (Timed Recursive ECATNets). A Timed Recursive ECATNet is a tuple (*RN*, *IntD*) where:

- *RN* = (*ECATNet, I, ϒ, K, ICT)* is a RECATNet, and
- *IntD*: $T_{elt} \rightarrow Q^+ \times (Q^+ \cup \{\infty\})$ is the duration function which associates to each elementary transition a static time interval (Q^+ is the set of positive rational numbers).

MODELING TIME CONSTRAINED WORKFLOWS USING T-RECATNETS

The T-RECATNets captures main dynamic reconfiguration mechanisms such as ability to create or suppress sub-processes. They permit advanced descriptions of complex data structures and context conditions checking (i.e. read and inhibitor arcs). They allow us to describe timing constraints and to specify the occurrence of exceptional situations dealing with time constraints violation. In our model, the time intervals assigned to elementary transitions describe, tasks duration in workflow processes. Indeed the duration of tasks in a workflow cannot be determined before its implementation. The process designers can only estimate the minimum and the maximum duration of each activity at built time. We propose in our T-RECATNets based approach for modelling collaborative workflows with an adjustable structure to introduce two types of tasks in these processes: *Elementary tasks* (represented by elementary transitions) and a*bstract tasks* (represented by abstract transitions). The execution of an abstract task generates dynamically, as a lower level thread, a new plan of action from the previous one (i.e. plan where the firing of the abstract task takes place). When a plan reaches a final marking, it terminates and the whole descent of action plans generated by it are aborted (i.e. a cut step is executed). Consequently, a dynamical tree of action plans describes the structure of a workflow process. Each action plan is executed independently and in parallel with the other plans of the tree. Note that this feature allows us to describe and to integrate in a simple way both the control-flow and the organizational perspectives of collaborative workflows: The distributed execution of a process over the collaborating process engines and their inter-process interaction can be reflected by the dynamical hierarchical structure of the tree. The event flows that coordinate and link up together these distributed sub-processes are reflected by the firing of abstract transitions (call for a sub-process) or the execution of cut steps (termination of a sub-process and result return to the caller). The exceptional situations that may occur, during the execution of a workflow, can be reflected by the execution of cut steps or by the firings of abstract transitions. So, a workflow process may handle exceptions, respectively, by terminating the current process or generating a new action plan. We illustrate the suitability of our model in modeling time-constrained flexible workflows through a simplified online computer shopping process specified as a T-RECATNet model depicted in Figure 4. The initial state of this net is a tree containing only the root thread with a token (*N*, *code*, *listCmd*) in the place *OrderReceived*. The workflow process starts by the firing of the elementary transition *StartProcess*. Then, the abstract transition "*StartOrder*" initiates the order handling subprocess by creating, dynamically, a new thread with the associated starting marking. In this sub-process the task *VerifCardAndProd* calls two services in parallel, *Check products* and *Validity of credit card* to check respectively the availability of the requested computer components and the customer's credit worthiness.

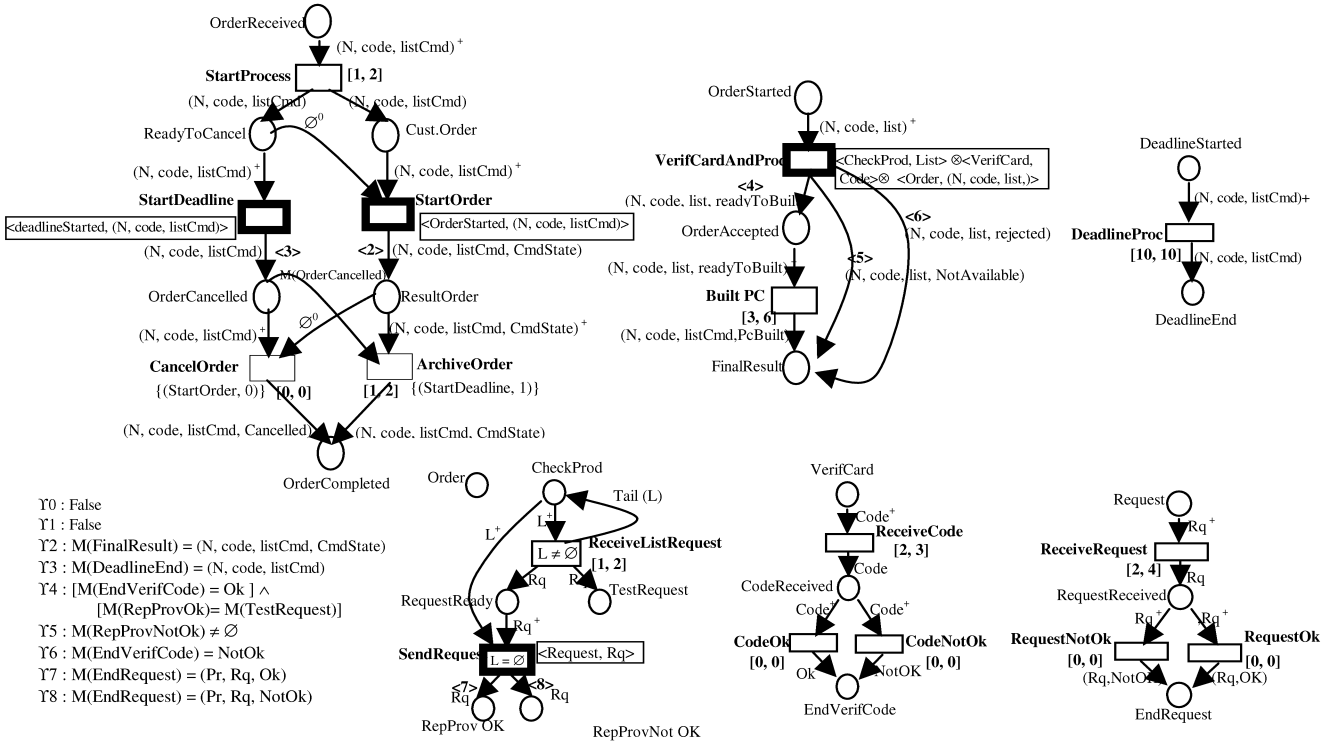


Figure 4: Time Constrained Online computer shopping workflow

The service *Validity of credit card* requires at least 2 units of time and at most 3 units of time to complete (its duration is affected by the duration of the elementary task *ReceiveCode*). In the initiated service to check the availability of the requested products, the abstract task *SendRequest* initialises (at each firing) for a requested product *Rq* a new instance of the service *Research in stock providers,* i.e., a new thread child is created dynamically in the tree of threads. The number of running instances of this invoked service depends on the number of the requested products. Each instance (i.e. thread) represents the local execution of the invoked service in the environment of its associated provider. The completion of one instance is indicated by a token in the place *EndRequest* (see Υ_7 and Υ_8). Depending on the value of the token produced in this place, and after the execution of the appropriate cut step, the place *RqProvOk* or *RqProvNotOk*, respectively, is marked in the previous recursion level. Next, another cut step is enabled at this level of recursion in one of the following cases: (1) if one of the requested computer components is not available (Υ_5 reached), (2) if the credit customer is not OK (Υ_6 reached), or (3) if the credit card and availability of all computer components are both OK (Υ_4 reached). When a cut step is executed, this level of action plan terminates and all the threads generated by it are aborted. Then, depending on the index of this termination (i.e. <4>, <5> or <6>) the outputs of the abstract transition *VerifCardAndProd* are produced. When the place *FinalResult* is marked, a cut step (with the index <2>) is executed which reduces the tree of threads to the root level with a token in the place *ResultOrder*. In the depicted workflow, the whole duration of the order handling sub-process (created by the abstract task *StartOrder*) must not exceed 10 units of times after its initialisation. The control of the complete duration of this sub-process is modelled by the abstract task *StartDeadline*. Its execution initialises another thread which contains a single elementary transition *deadline* to model the duration of the affected deadline (a duration about 10 units). When this sub-process terminates (i.e. a token is produced in the place *EndDeadline* and the appropriate cut step is executed), the *CancelOrder* is enabled and immediately fired only if the order handling sub-process is not yet completed (i.e. there no token in the place *ResultOrder*). In this case, the execution of *CancelOrder* aborts the sub-process generated by *StartOrder* and produces a token *(N, code, listCmd, Cancelled)* in the place *OrderCompleted*. Note that *CancelOrder* interrupts the abstract transition *StartOrder* with the index <0>. In the case where the order handling sub-process terminates before the completion of the deadline, it is the deadline sub-process which is cancelled by the firing of the transition *archiveOrder* and the tokens in *OrderCancelled* are removed. Such a construction with T-RECATNets describes, adequately, the flexible structure of the online shopping workflow process and its distributed execution over collaborative partners. An equivalent modelling of this flexible behavior with other high level Petri nets (e.g. colored Petri nets) would be very complicated. The resulting net, for the modelling of the cancellation of the running processes, for instance, would contain spaghetti-like arcs to remove tokens from all combinations of all places.

T-RECATNETS SEMANTICS IN TERMS OF RL

In what follows, $Spec = (\Sigma, E)$ is an algebraic specification of an abstract data type given by the user and associated to a RECATNet (with E its set of equations and Σ the set of operations and sorts) and where the sorts domains are finite. $T_{\Sigma,E}(X)$ is the Σ-algebra of the equivalence classes of the Σ-terms with variables in X, modulo the equations E. $MT_{\Sigma,E}(X)$ represents the free commutative monoid of the terms $T_{\Sigma,E}(X)$ endowed with the internal operator $\oplus$ and having $\varnothing$ as the identity element. $CATdas(E,X)$ is the structure of equivalence classes formed from the multisets of $MT_{\Sigma,E}(X)$ modulo the associative, commutative and identity axioms for the operator $\oplus$. In rewriting logic (RL), a concurrent system is described by a (generalized) rewrite theory (Σ, E, ϕ, R) such that its static structure is described by the equational theory (Σ, E) whereas its dynamic structure is described by the set of (labeled) rewriting rules R. The function ϕ specifies which arguments of a function symbol f of Σ cannot be rewritten. These arguments are called frozen positions. Each rewrite rule corresponds to an elementary action being able to occur, concurrently, with other actions in the system. A conditional rewrite rule of the form $t \rightarrow t'$ if C, means that a fragment of the system state fitting pattern t can change to a new local state fitting pattern t', if the condition C holds. In (Ölveczky and Meseguer 2007), real time rewrite theories were introduced to specify real time and hybrid systems in rewriting logic. In this type of rewrite theories, some rules called tick rules model time elapse in a system while ordinary rewrite rules model instantaneous change. The underlying membership equational specification of such a rewrite theory should contain an axiomatisation of the sort Time of the time domain and an operator $\{_\}$ which encloses the global state of the system. A tick rule has the form l: $\{t\} \rightarrow \{t'\}$ in time τ_l if C, where the term τ_l of sort *Time* denotes the duration of the rule. In what follows we adopt Real Time Maude notation to present specific equational or rewrite theories. Real Time Maude supports the specification of real time rewrite theories as timed modules. The global distributed state of a Timed RECATNet is described by a dynamical tree *Tr* of threads marking. Each thread (node) *Th* of the tree *Tr* is expressed as a term $[M_{Th} \mid t_{abs} \mid ThreadChilds]$, where: M_{Th} represents the marking of the thread *Th* which is a multiset made up of the marking of each place of this thread. The sub-term t_{abs} represents the name of the abstract transition whose firing (in the thread father) gave birth to the thread *Th*. ThreadChilds represents the threads generated by this thread *Th* in the current tree *Tr*. They are described as a finite multiset of terms of sort *Thread*. The constant *nullThread* represents the empty thread and the operator "_._" is the corresponding multiset union operator which is associative, commutative and has the constant nullThread as the identity element. We associate to each (delayed) created tokens by an elementary transition, a timestamp representing the time where these tokens become available for consumption by other transitions. Delayed tokens (marking) that may become visibles at any time r in an interval $[r_1, r_2]$, from now, are represented by a term $dly(M, r_1, r_2)$ of sort *TimedMarking* (with r_1 and r_2 of sort *Time*). Delayed tokens, will become visible when the minimum delay for these tokens has passed. This happens nondeterministically (i.e. at any time after

expiration of the minimum duration of the transition) by the execution of a rewrite rule of the form:

[*UnDelay*]: **dly**(*M:Marking* , 0 , *R*) → *M:Marking*

We define the function **mte**, which calculates the maximum time elapse possible from a given state of a T-RECATNet before a delayed marking become available. Also, we define the function **delta** to model the effect of passage of time on delayed tokens in thread places by decreasing their minimum and maximum delays according to the time elapsed (Ölveczky and Meseguer 2007). For space limitation, the complete specification of these functions is not given in this paper. The set of rewrite rules associated to a Timed RECATNet consist in a *lazy tick rule* (which model the time elapsing) and for each transition *t* (elementary or abstract) and for each cut step of this T-RECATNet, there is an *eager rewrite rule* describing its behaviour. The application of these eager rules should take precedence over the application of time advancing tick rule. To model the urgency behind the behavior of eager rules, we introduce the function **eagerEnabled** that takes as argument a state of the tree of threads and returns the value *True* if there is an eager rewrite rule enabled for this given state, and returns the value *False* otherwise. The general form of T-RECATNets rewrite rules is given in the following.

```
vars M Mf M1 M: Marking . vars mTh mThf mTh1 mTh2:
Thread . vars T Tf T1 T2: Trans .  var R: Time .
vars mp mp' mp1...mpfinal: Multiset. var TM : TMarking.

***(Elementary rules:For each elementary transition
telt with K(telt)=∅ and Int(telt) = [tmin, tmax])***
crl [telt]:  <p, mp ⊕ DT(p, t)> → <p, mp> ⊗ dly ( <p',
CT(p', t)> ), tmin, tmax ) if InputCond and TC(t).

******** Abstract rules ********************
crl [tabsi]:[TM ⊗<p, mp ⊕ DT(p, t)>, T, mTh
]→[TM ⊗ <p, mp>, T, mTh [<p",CT(p",tabsi)> ⊗ <p1,
Ems> ⊗...⊗ <pn, Ems>,tabsi, nullThread]] if
(InputCond) and (TC(t)) .

******** Pruning rules ****************************
*** (A pruning rule associated to a cut step
occurring in a thread different from the root and
generated by an abstract transition tabsj )***
crl [τi]: [Mf ⊗ <p', mp'>, Tf,[M ⊗ <pfinal, mpfinal>,tabsj,
mTh] mThf] → [Mf ⊗ <p', mp' ⊕ ICT(p',tabsj,i)>, Tf,
mThf] if [ϒi].
*************** the tick rule **************
crl [tick]:  {[ M | nullTrans | Th ]}  →
{delta([ M | nullTrans | Th ], R)} in time R
if (R <= mte([ M | nullTrans | Th ])) and ( not
eagerEnabled(M) ) [nonexec] .
```

In Figure 5, we present an execution of the workflow of Figure 4 using as platform the version 2.3 of the Real-Time Maude system under Linux. The command *tfrew* (timed fair rewrite) of Real-Time Maude simulates one behavior (i.e. one sequence of rewrite steps) of the many possible concurrent behaviors of a system starting from a given initial state and up to a certain duration of the rewrite sequence. Real-Time Maude offers choices of different time sampling techniques so only some moments in time are visited to deal with the problem of dense time (in figure 5, note that we choose to increment time value by amount of 1/2). If one want to know the longest and the shortest time it takes to reach a desired state from an initial one, we can use respectively the commands *find latest* and *find earliest*.

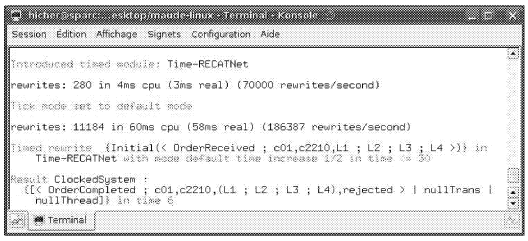


Figure 5: Simulation of the workflow of Figure 4 under the Real Time Maude Tool

CONCLUSION

In this paper, we present the Timed Recursive ECATNets (T-RECATNets) for the modelling of timed–constrained reconfigurable workflow processes operating on distributed execution environments. We define their semantics in rewriting logic by means of real time rewrite theories. In future works, we intend to extend the T-RECATNet model introducing delays on the enabling conditions of transitions. We plan also to enrich the descriptive power of the T-RECATNets model by allowing threads of the same level of hierarchy to communicate results between them and with their father threads before their termination (via shared places). This feature would be particularly relevant in the field of collaborative workflow systems.

REFERENCES

van der Aalst, W.M.P. and K.M. van Hee, 2002 *Workflow Management: Models, Methods, and Systems*. MIT press, Cambridge, MA.

Barkaoui K. and A. Hicheur, 2007. "Towards Analysis of Flexible and Collaborative Workflow Using Recursive ECATNets". *BPM Workshops 2007*, In B. Benatallah, A. ter Hofstede and H. Paik (Eds.) Springer Berlin / Heidelberg. LNCS 4928, 206-217.

Bettaz, M and M. Maouche, 1992. "How to specify non determinism and true concurrency with algebraic terms nets". In LNCS 655, Springer-Verlag, 164-180.

Bruni R. and J. Meseguer, 2006. "Semantic foundations for generalized rewrite theories". In *Theoretical computer science*, Vol 360, No 1-3, 386-414.

Haddad S. and D. Poitrenaud, 2007. "Recursive Petri nets: Theory and application to discrete event systems". In *Acta Informatica*. Springer Berlin / Heidelberg (Eds.).Vol 40 (7-8), 463-508.

Halliday, J.J. ; S.K. Shrivastava and S.M. Wheater, 2001. "Flexible workflow management in the OPENflow system". In *Proceedings 5thIEEE International Enterprise Distributed Object Computing Conference*, Seattle, WA, USA, 82-92.

Ling S. and H. Schmidt, 2000. "Time Petri nets for workflow modelling and analysis". In *Proceedings IEEE International Conference Systems, Man, and Cybernetics*. Vol. 4,3039- 3044.

Lin C. and Qu Y., 2004. "Temporal inference of workflow systems based on time petri nets: Quantitative and qualitative analysis". *International Journal of Intelligent Systems*, Vol 19, 417-442.

Ölveczky P. and J. Meseguer, 2007. "Semantics and pragmatics of Real-Time Maude". In *Higher-Order and Symbolic Computation*. Vol 20 , 161-196.

Sadiq W.; S. Sadiq and K. Schulz, 2006. "Model Driven Distribution of Collaborative Business Processes". In *Proceedings of IEEE International Conference*, Chicago, 281-284.

Tiplea, F.L. and G.I. Macovei, 2005. "Timed workflow nets" *In Proceedings of IEEE International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. 25-29.

LOGISTICS APPLICATIONS

ASSESSMENT OF DELIVERY STRATEGIES BY APPLYING LEXICOGRAPHIC OPTIMIZATION AND DESIGN OF EXPERIMENTS: A CASE STUDY

Herminia I. Calvete
Dpto. de Métodos Estadísticos
Universidad de Zaragoza
Pedro Cerbuna, 12
50009 Zaragoza,
Spain
E-mail: herminia@unizar.es

Carmen Galé
Dpto. de Métodos Estadísticos
Universidad de Zaragoza
María de Luna 3
50018 Zaragoza,
Spain
E-mail: cgale@unizar.es

María-José Oliveros
Dpto. de Ingeniería de Diseño y Fabricación
Universidad de Zaragoza
María de Luna 3
50018 Zaragoza,
Spain
E-mail: mjoliver@unizar.es

Belén Sánchez-Valverde
Dpto. de Métodos Estadísticos
Universidad de Zaragoza
María de Luna 3
50018 Zaragoza,
Spain
E-mail: belensv@unizar.es

KEYWORDS
Vehicle routing. Time windows. Design of experiments. Multiobjective programming. Lexicographic optimization.

ABSTRACT

To have an effective and efficient distribution system is becoming increasingly important since this last segment of the supply chain accounts for a large proportion of the overall operational cost. In this research we aim to integrate optimization and design of experiments techniques to obtain information about the factors that influence the most the performance of the delivery system when making decisions on concentrating customers and offering more flexible periods of service in a real distribution system. For this purpose we consider the geographical distribution of customers and the tightness of the time windows and analyze their impact on relevant delivery system performance measures like the number of vehicles needed, the travel time involved and the satisfaction of customers. To deal with all these objectives and enable the decision maker to evaluate alternative solutions, a lexicographical approach is implemented in the proposed solution procedure. This work has been motivated by company's interest in designing a delivery strategy that satisfies all demand requirements and maintains high customer service levels at low costs.

INTRODUCTION

Companies have to make operational and strategic decisions in order to manage the processes involved in their supply chains efficiently. Amongst them, timely distribution of commodities to final customers at low cost, with high quality service is one of the major problems faced. The distribution network design includes such aspects as defining distribution zones, letting the customers have preferences in service, etc. All of them have a strong influence on the most central problem in distribution management, that is to say, the scheduling of delivery routes.

These kinds of problems are generically known as *Vehicle Routing Problems* (VRP). In general, their purpose is to determine a set of routes for a fleet of vehicles that is located in a central depot so that some measure of the performance of the delivery system, usually total travel time or distance, is optimized. All constraints regarding customer service level and vehicle capacity have to be held. In addition a variety of constraints can be taken into account. For instance, specific time windows, i.e., time bounds which determine the earliest and the latest time that deliveries can take place, multiple depots, simultaneous pickup and delivery, etc. Due to its importance, there is a vast literature that addresses modeling and solution aspects of VRP and its variants. It is worth mentioning that all these combinatorial optimization problems have been proven to be NP-hard. Only relatively small instances can be solved to optimality due to their huge computational requirements. For larger problems, researchers usually focus on heuristic and meta-heuristic methods to derive approximate solutions of acceptable quality in reasonable computational time. Comprehensive reviews of these problems are presented in (Ball et al. 1995; Cordeau et al. 2002; Gendreau,

Laport, and Potvin 1997; Golden and Assad 1988; Solomon and Desrosiers 1988; Toth and Vigo 2002).

Most of this research focuses on algorithms and procedures which are efficient from the point of view of the computational time involved in order to get an optimal or quasi-optimal set of delivery routes. But they pay scant attention to analyzing how decisions like offering more flexible periods of service or grouping customers could influence the delivery system performance. However, these are key decisions faced by the logistics departments of companies. In this paper we provide some insights to this question by integrating optimization and design of experiments methodologies. We propose to compare the performance of the delivery system under different configurations, aiming to determine which of these factors most influence its efficiency. To make a decision about the criterion that should be selected in order to measure the performance of the system is not easy. Therefore several criteria are used in the paper and a lexicographic optimization approach is adopted. The proposed methodology is illustrated using data from a delivery problem faced by a hardware manufacturer located in Spain. Nevertheless, it is worth pointing out that it can be applied to any medium-sized general distribution system. The remainder of the paper is organized as follows. We start by presenting a comprehensive description of the problem tackled. Then, the next Section goes on to assess the performance of the distribution system. To this end, the company goals are described. Moreover, a methodology is proposed to solve the problem of determining the best set of routes according to these objectives. The following Section provides the company with the best set of routes and allows them to evaluate the consequences of changing the importance assigned to goals. The influence of several factors in the performance of the system are then investigated by using design of experiments. Finally, we conclude the paper with a summary.

PROBLEM DESCRIPTION

Here we focus on a hardware company, which daily delivers goods to 60 geographically dispersed locations from a central depot (see Figure 1). In each location there is one hardware shop which has to be supplied. Each hardware shop has an owner, who has a share in the company. From now on, we generically identify locations with customers. The depot acts as a central buyer of goods, which are delivered to customers as they request them. In general, hardware shops have small storage spaces, so they require to be served daily. The depot is big enough to assume that there is always stock to be able to serve the customers.

Every day a set of routes must be scheduled for serving all customers, which are located at travel times ranging from 16 to 237 time units (minutes) away from the depot. A route involves information on the customers which are visited, the order in which they are visited and the departure time of the vehicle from the depot.

Each customer has a demand, which is assumed to be known

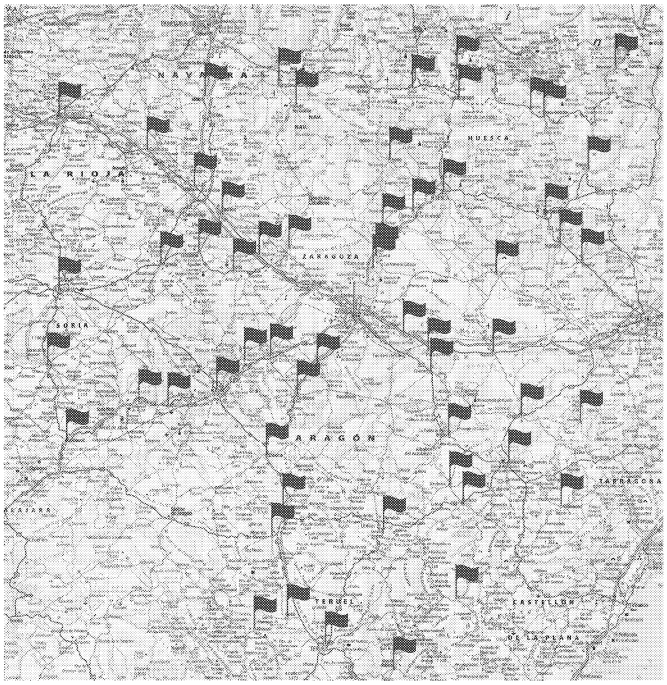


Figure 1: Geographical location of the 60 customers

when routes are scheduled, and a service time. Although there can be slight differences between demands from one day to the next, the company has provided us with a daily customer demand which ranges from 2 to 15 units. A fixed service time for vehicle unloading is associated with each customer. This service time has been estimated to be 25 time units. Travel time between the depot and each customer location, as well as between each pair of customer locations is also known.

Every customer specified a period of 90 time units along the working day when he/she would like to be served. This defined a soft time window. The lower bound of the soft time window indicates the earliest time in which the service to the customer would start, according to the preferences of the customer. Likewise, the upper bound indicates the latest time in which the service to the customer would end. These are not hard constraints, i.e. they can be violated without penalty. Nevertheless, to arrive at some form of compromise with the customers, the maximum deviation from the soft time window has been established to 60 time units. That is to say, deliveries to a specific customer cannot start earlier than an hour in advance to his/her earliest preferred service time nor finish later than an hour after his/her latest preferred service time. This defines a hard time window of 210 time units length. Notice that the lower (upper) bound of the hard time window indicates the earliest (latest) time in which the service to the customer should start (end). The hard time window associated with the depot is 960 time units.

At present, the company uses the external services of truck companies to perform all daily deliveries to customers. So, we assume that there is no limitation on the number of vehicles that can be used. In fact, the actual number of vehicles

that will be used is one of the variables of the model. For each vehicle used, there is a fixed cost of 120 units and a cost per travel time unit of 0.156. It is also assumed that vehicles are located at the depot, so they leave from and return to the depot. The vehicle capacity is 35 units, so each vehicle is able to service at least one customer. Furthermore, vehicles should not be overloaded, hence total demand of all customers on one particular route must not exceed the capacity of the vehicle assigned to this route. As well, deliveries cannot be split up amongst vehicles. That is to say, each customer is served by a single vehicle. We also assume that delivery starts as soon as the vehicle arrives at the customer location and that the vehicle leaves the customer as soon as the delivery has been completed (once the service starts, it is carried out until completion). Finally, according to legal regulations, the total daily time that a driver can work is set to 600 time units, and the total daily time that a driver can drive is set to 480 time units.

SYSTEM PERFORMANCE ASSESSMENT

As previously mentioned, in the classical VRP the single objective involved in the scheduling of routes is the minimization of travel time. However, companies usually consider additional criteria than just travel time. In particular, the company's need to provide excellent service level, which is measured in terms of customer delivery preference satisfaction, is a foremost goal. Also, since they are aware of costs, both the number of vehicles used and the total travel time are relevant. Hence, consistent with the priorities of the company, the following ranked criteria were taken into account:

Z_1: Minimize the deviations from soft time windows
Z_2: Minimize the number of vehicles involved
Z_3: Minimize total travel time

The first objective is related to service level and the two last ones are related to company costs. Although Z_2 and Z_3 could have been combined in cost by the expression $120Z_2 + 0.156Z_3$ we preferred to minimize them separately. This fact allowed us to detect more precisely the importance of each objective, as will be seen further on.

When dealing with multiple objectives, in general, there is not a single solution which optimizes each objective individually. Therefore trade-off decisions have to be made in order to get the most preferred solution. One possibility is to define an objective function as a weighted sum of these criteria. The problem is that these criteria are expressed in different measurement units. Therefore, adding them in their original form has no significance and causes a bias towards the objectives with a larger magnitude. In addition, to find normalization constants which suitably reflect their importance for the decision maker is usually complex. So, when the manager is able to rank the criteria according to their importance, a lexicographical optimization approach is best (Chankong and Haimes 1983; Steuer 1986; Zeleny 1982). We assume that a little improvement for a particular criterion is more valuable than arbitrarily large improvements for subsequent rank criteria. Then, the lexicographic method produces an optimal solution by sequentially optimizing the criteria. The main criterion is optimized first. Then, the second criterion is optimized subject to achieving the optimum with respect to the first criterion. The procedure follows in the same way until all criteria are optimized.

The resulting model to determine the best set of routes according to these criteria is a mixed-integer lexicographic multiobjective programming problem (Chankong and Haimes 1983; Steuer 1986; Zeleny 1982). General techniques to handle these kind of problems can be applied to get a solution. Nevertheless, tailored procedures are usually more efficient. Hence, we propose to use the underlying idea of the methodology EFBO (enumeration-followed-by-optimization) proposed in (Calvete et al. 2007) to deal with goal programming problems. This procedure has been proved to provide an optimal solution (the optimal number of routes and the optimal sequence of customers visited by each vehicle) in short computational times in medium-sized goal programming distribution problems.

Next, we present a brief description of the EFBO technique and the modifications needed to handle the lexicographic optimization. In short, the methodology EFBO consists of two phases. The purpose of the first phase is to enumerate all possible ways of connecting the depot and the customers in order to identify which ways provide feasible routes. After a raw computation of these routes, a filter is applied whose purpose is to remove routes which we know for sure will not be in the set of optimal routes. Upon termination of the first phase, the filtered set of feasible routes which are candidates for the optimal solution is available. The second phase of the methodology EFBO picks from this set the subset of best routes according to the objective function considered by solving a set partitioning problem.

In order to take into account the lexicographic optimization, the filtering step and the second phase of the methodology EFBO have to be modified. In case there are several feasible routes serving the same customers, after the filtering step only the most promising one is kept. This is selected by successively optimizing the objectives in the lexicographic order proposed. With regard to the second phase, if the main criterion is Z_1, the second one is Z_2 and the third one is Z_3, now the following lexicographic set partitioning problem has to be solved:

$$\begin{array}{ll} \text{lexmin} & (Z_1,\ Z_2,\ Z_3) \\ \text{s.t.} & \sum\limits_{R \in \mathcal{R}} A_{Ri} X_R = 1, \quad \forall i \in N \\ & X_R \in \{0,1\}, \quad \forall R \in \mathcal{R} \end{array} \tag{1}$$

where N is the set of customers, $\mathcal{R}$ is the set of all feasible routes which are candidates for the optimal solution,

$$X_R = \begin{cases} 1, & \text{if route } R \text{ is selected} \\ 0, & \text{otherwise} \end{cases}$$

and

$$A_{Ri} = \begin{cases} 1, & \text{if route } R \text{ visits customer } i \\ 0, & \text{otherwise} \end{cases}$$

The problem constraint ensures that all customers are just visited once.

Next we examine the analytical form of the objective functions. Let t_{ij} denote the time taken to travel from i to j through connection (i, j). Let

$$\begin{array}{ll} a_i & = \text{arrival time to customer } i \\ p_i & = \text{departure time from customer } i \\ s_i & = \text{service time of customer } i \\ [e_i^s,\, l_i^s] & = \text{the soft time window associated with customer } i \end{array}$$

Since delivery starts as soon as the vehicle arrives to the customer and once the service starts it is carried out until completion, then $p_i = a_i + s_i$. Moreover, there is a discrete set of departure times at which a vehicle can begin its assigned route. We assume that all vehicles have the same possible set of departure times, $\{d_1, \ldots, d_m\}$ where $m = 27$, $d_1 = 0$, $d_i = d_{i-1} + 30$ and $d_{27} = 780$ time units.

Let $\{1, i_2, i_3, \ldots, i_v, 1\}$ be the sequence of ordered customers of a feasible route R, where index 1 refers to the depot. As previously mentioned, the route also specifies the departure time of the vehicle from the depot. Then, the route travel time is $\tilde{t}_R = t_{1i_2} + t_{i_2 i_3} + \ldots + t_{i_v 1}$ and the route deviation from the soft time window is

$$dstw_R = \sum_{j=2}^{v} \max\{0,\, e_{i_j}^s - a_{i_j}\} + \sum_{j=2}^{v} \max\{0,\, p_{i_j} - l_{i_j}^s\}$$

Therefore

$$\begin{aligned} Z_1 &= \sum_{R \in \mathcal{R}} dstw_R X_R \\ Z_2 &= \sum_{R \in \mathcal{R}} X_R \\ Z_3 &= \sum_{R \in \mathcal{R}} \tilde{t}_R X_R \end{aligned}$$

GETTING THE BEST SET OF ROUTES

The aim of the first part of the study was twofold. First, to provide the company with the best set of routes according to the established criteria. Second, to compare this optimal solution with the optimal solution obtained when changing the order of the criteria. Table 1 provides the number of vehicles, the total travel time and the deviations from soft time windows for all these optimal solutions. They have been obtained after applying the methodology EFBO. The cost of each solution is also shown. The individual optimal value of each objective and the minimum cost are displayed in bold.

The optimal solution in accordance with the company ranked criteria is shown in the first row of Table 1 (lexicographical model M_1). Notice that it is impossible to satisfy the preferences of all customers since the deviation from soft time windows is 13 time units. A detailed analysis of the routes included in the optimal solution shows that there are two customers whose upper bound of the soft time window is exceeded by 6 and 7 time units, respectively. Moreover, the cost of this optimal solution is almost twice the cost of the solution shown in the forth row of Table 1, which is the most economical. This last solution is obtained by first minimizing the number of vehicles, then minimizing the total travel time and, finally, minimizing the deviation from soft time windows. As we could expect, the cheapest solutions are obtained when Z_1 is not the first criterion to be optimized. What is more surprising is the large variation in costs. Besides, comparing rows three and four in this Table, we notice that with a little cost increase (83.3 units) it is possible to decrease the deviation from soft time windows in 567 time units.

Table 1: Objective function optimal values and cost of the lexicographic optimal solutions

lex model	lexmin	Z_1	Z_2	Z_3	Cost
M_1	(Z_1, Z_2, Z_3)	**13**	37	11536	6239.6
M_2	(Z_1, Z_3, Z_2)	**13**	38	11527	6358.2
M_3	(Z_2, Z_1, Z_3)	909	**19**	8178	3555.8
M_4	(Z_2, Z_3, Z_1)	1476	**19**	7644	**3472.5**
M_5	(Z_3, Z_1, Z_2)	1438	21	**7519**	3693.0
M_6	(Z_3, Z_2, Z_1)	1438	21	**7519**	3693.0

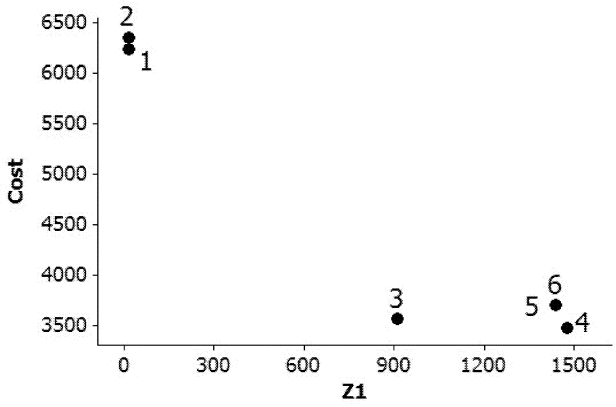


Figure 2: Scatterplot of cost versus Z_1 for models M_1 to M_6

Figure 2 displays costs versus Z_1 for the solutions in Table 1. There are two groups of solutions. The one formed by the solutions of models M_1 and M_2, which minimize Z_1 first, and the group formed by the solutions of the remaining models. Moreover, the solution of model M_3 shows a good level of achievement with regard to costs and is much better than models M_4, M_5 and M_6 regarding Z_1.

Bearing in mind that the company is part owned by the hardware shops, who finally meet the costs of delivery, it would be worth considering a change in the priorities of the crite-

ria. Although the deviation from soft time windows is large (909 units against 13 units), the optimal solution of model M_3 could be acceptable to customers, taking into account the huge decrease in costs. Furthermore, a detailed analysis of the routes of the optimal solution demonstrates that the average deviation from soft time windows is only 15.22 time units per customer, with a standard deviation of 10.93, and a maximum of 37 time units.

Let's now make some remarks on the importance of considering individually the objectives Z_2 and Z_3, i.e. the number of vehicles and the total travel time. If we had only considered deviation from soft time windows and costs we would only have had two lexicographical models: (Z_1, Cost) and (Cost, Z_1). This is, obviously, easier to handle. The optimal solutions would have been (13, 6239.6) and (1476, 3472.5), which coincide with the optimal solutions provided by models M_1 and M_4. However, it would have been impossible to provide the company managers with the solution given by model M_3, which is almost as cheap as the best solution from the point of view of costs and much better from the point of view of customer preference satisfaction.

CHANGING THE DISTRIBUTION STRATEGIES

The second part of the study gives an insight into the influence of several factors in the satisfaction of customer preferences and costs. The company pointed out the possibility of grouping customers by proximity and serving each zone in a specific period: morning, afternoon or evening. The other aspect to be negotiated with customers was the time windows strategy. The company was thinking about increasing the length of the soft time windows. Concurrent with this, the company would like to examine the potential system performance improvement when some of the customers are not allowed to specify a soft time window. Therefore, we used the design of experiments (DOE) (Montgomery 2005) which refers to an organized method for determining the relationship between factors affecting a process and the output of that process. This is an effective method for obtaining maximum information with a minimum number of experiments. For this purpose we begin by identifying the factors and the response or output. We define a number of levels for every factor, whose effect we want to find out. The method of analysis is to look for differences in the output obtained from the different factor combinations. These differences are then attributed to the factors acting alone (called single effect) or in combination with other factors (called interaction).

The company established three controllable factors:

- Factor A: The setting-up of customer clusters
- Factor B: The tightness of the soft time window
- Factor C: The percentage of customers with soft time windows

We defined a 2^3 factorial design that required choosing two levels for each factor and allowed us to analyze the impact of the three factors on each measure of the performance of the system by carrying out eight runs of the experiment for each response variable. For factor A, the two levels considered were whether grouping customers in clusters or not. We coded these levels as G or NG, respectively. When customers were grouped, there were three clusters (see Figure 3), served in the morning, in the afternoon and in the evening, respectively.

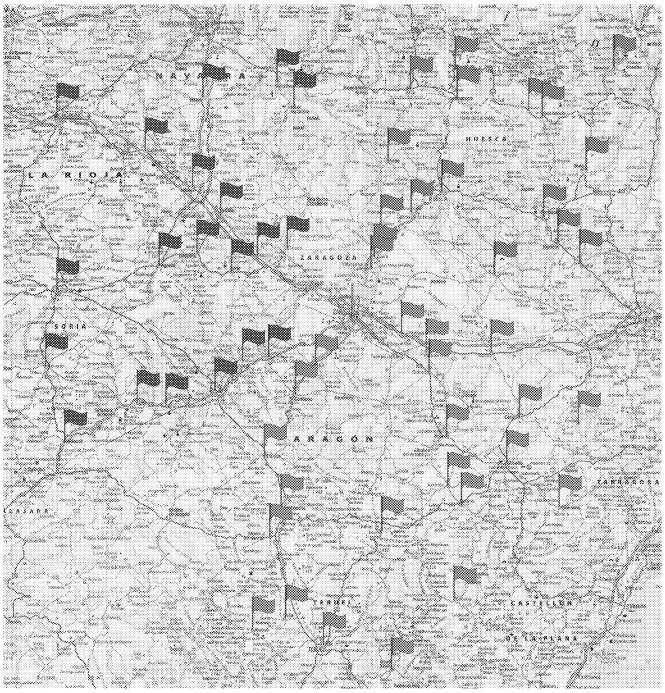


Figure 3: Setting-up of customer clusters

Table 2: Design matrix for a 2^3 factorial design

Config.	Factor A	Factor B	Factor C
C_1	NG	90	80
C_2	NG	90	100
C_3	NG	150	80
C_4	NG	150	100
C_5	G	90	80
C_6	G	90	100
C_7	G	150	80
C_8	G	150	100

For factor B the two levels selected were 90 time units (the current situation) and 150 time units. Finally, the company estimated that it would be possible to convince 20% of customers to give up the right to establishing preferences in service. So, for factor C the two levels considered were 100% and 80%. As a result, each factor combination provides a configuration of the delivery system. Its characteristics are summarized in Table 2.

We computed the optimal solutions of lexicographic models 1 to 6 under each of the eight configurations of the delivery

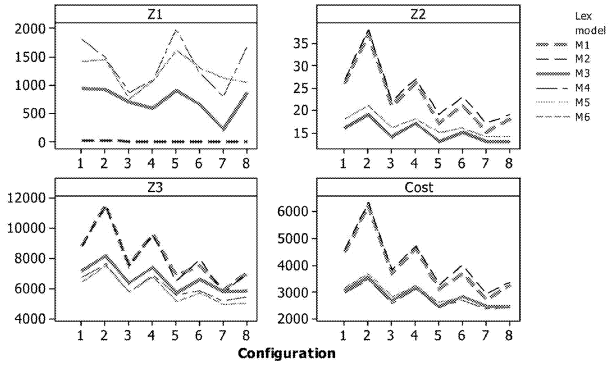


Figure 4: Optimal value of Z_1, Z_2, Z_3 and the cost for each model and every configuration

system. Table 3 displays the optimal value of the deviation from soft time windows Z_1 and the cost. The minimum value for each model is displayed in bold. At a glance, the first noteworthy result is that, whatever the lexicographic order, the cheapest solution is provided by configuration C_7 (except for model M_3, but both costs are similar). That is to say, it is better to group the customers, to extend the soft time window up to 150 time units and to limit the number of customers with soft time windows. But, from the point of view of the customer this configuration is not very appealing since it greatly limits the possibility of specifying preferences in service.

Figure 4 shows the optimal value of the three objectives as well as the cost for each model and every configuration. Looking at it, we concluded that there are two groups of models which have a similar performance with regard to cost: the group of models M_1 and M_2 and the group of the rest of models. Models M_1 and M_2 are excellent for satisfying customer preferences but they are very expensive. This is specially true for configurations C_1, C_2 and C_4. Furthermore, for all configurations, model M_3 gives good results in terms of cost and not as bad results as the other models in its group in terms of satisfaction of customer preferences. This fact strengthens the idea expressed in the previous Section of considering a change in the criterion priorities. So, it would be better to first minimize the number of vehicles and then minimize the deviation from soft time windows.

Next, we analyzed the influence of each factor in each model, taking Z_1 and cost as responses. The statistical analysis was carried out using MINITAB® Release 14.20.

Cost as the response

The results of the Analysis of Variance (ANOVA) applied indicate that factors A and C are significant in models M_1, M_3 and M_4 (the p-value being almost zero). Likewise, factors A, B and C are significant in models M_2, M_5 and M_6. Interaction AC is also significant in models M_5 and M_6. Factor A is in all cases the most significant. Factors A and B always have

Table 3: Experimental results

Configuration	$M_1:(Z_1,Z_2,Z_3)$		$M_2:(Z_1,Z_3,Z_2)$		$M_3:(Z_2,Z_1,Z_3)$		$M_4:(Z_2,Z_3,Z_1)$		$M_5:(Z_3,Z_1,Z_2)$		$M_6:(Z_3,Z_2,Z_1)$	
	Z_1	Cost	Z_1	Cost	Z_1	Cost	Z_1	Cost	Z_1	Cost	Z_1	Cost
C_1	6	4502.2	6	4617.0	929	3042.1	1788	2973.2	1412	3166.7	1412	3166.7
C_2	13	6239.6	13	6358.2	909	3555.8	1476	3472.5	1438	3693.0	1438	3693.0
C_3	**0**	3695.9	**0**	3813.7	692	2675.6	858	2584.5	**749**	2820.6	**749**	2820.6
C_4	**0**	4619.8	**0**	4731.2	585	3198.5	1081	3120.1	1064	3219.6	1064	3219.6
C_5	**0**	3118.3	**0**	3294.3	890	**2456.1**	1971	2435.6	1604	2604.6	1604	2604.6
C_6	**0**	3695.9	**0**	4003.3	645	2836.0	1205	2712.9	1294	2814.5	1294	2814.5
C_7	**0**	**2725.2**	**0**	**2932.9**	**207**	2464.3	**774**	**2369.0**	1111	**2453.1**	1111	**2453.1**
C_8	**0**	3259.3	**0**	3372.2	857	2479.2	1688	2416.8	1040	2469.0	1040	2469.0

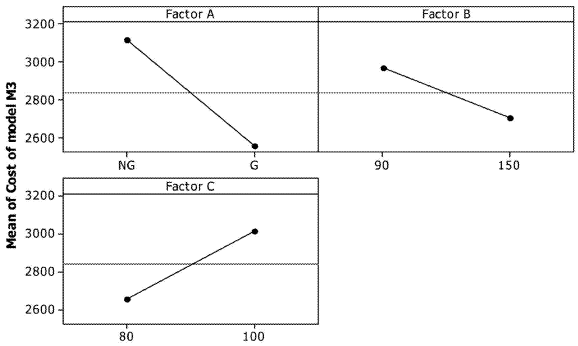


Figure 5: Main effects plot for the cost in model M_3

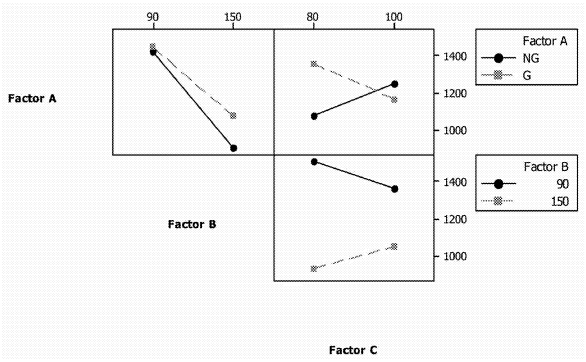


Figure 6: Interaction plot for Z_1 in model M_5

a negative effect in the cost while factor C has a positive effect. As an example, Figure 5 shows the main effects plots of factors A, B and C corresponding to model M_3. Accordingly, either grouping the customers or extending the soft time window decreases the cost, whereas increasing the percentage of customers with soft time windows increases the cost.

The deviation from soft time windows as the response

The results of the ANOVA show that there are not significant factors in models M_1 to M_4. Factor B and interaction AC are significant in models M_5 and M_6. Factor B has a negative effect, i.e. extending the soft time window decreases the deviation from soft time windows. Figure 6 shows the interaction plot of factors A, B and C corresponding to models M_5 or M_6 (both models provide the same results). Accordingly, increasing the percentage of customers with soft time windows decreases the deviation from soft time windows if customers are grouped and increases it when customers are not grouped.

Distribution strategies and lexicographic order

Finally, we assumed that only models M_3 to M_6 were acceptable in terms of cost and jointly studied the effect of factors

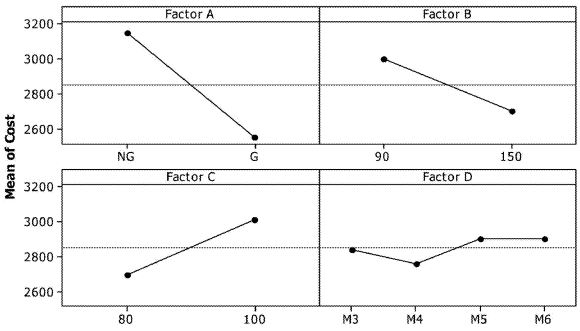


Figure 7: Main effects plot for the cost

A, B and C together with the effect of the lexicographic optimization order. For this purpose, we defined a new factor D taking the levels M_3 to M_6. First, we took the cost as the response. In this case, from the ANOVA results we deduced that factors A, B, C and D together with the interactions AB, AC and BC are significant (the p-value being almost zero). Factor A explains 58% of the variability of the cost, factor B explains 15% and factor C explains 17%. In contrast, interaction AC explains only 5%, factor D explains 2% and the remaining interactions explain less than 1% of the variability. Figure 7 shows the main effects plot of factors A to D corresponding to the cost. Hence, when restricting the attention to models with acceptable costs, in order to reduce costs, the lexicographic order is not as important as selecting properly the delivery strategy. In fact, it is specially important to be able to group customers.

Second, we took the deviation from soft time windows as the response. Now the significant factors and interactions are B, D and BC (the p-value being almost zero). Factor D explains 40% of the variability of Z_1, while factor B explains 29% and the interaction BC only explains 11% of the variability. Figure 8 shows the main effects plot of factors A to D corresponding to the deviation from soft time windows. Thus, to improve the satisfaction of customers the most important characteristic to be taken into account is the appropriate selection of the lexicographic order. Notice that model M_3 together with a soft time window of 150 time units provide the best results.

SUMMARY

In this paper, we propose to apply lexicographic optimization and design of experiments in order to analyze the performance of delivery systems. We focus on general distribution systems involving a single central depot serving a number of geographically dispersed customers who have hard and soft time windows. Moreover, multiple objectives are taken into account: minimizing the deviation from soft time windows (which is equivalent to maximize satisfaction of customers), minimizing the number of vehicles used and minimizing the

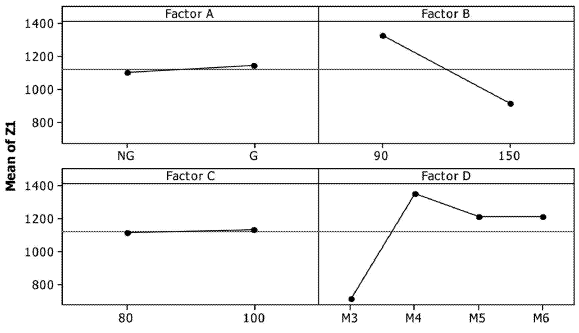


Figure 8: Main effects plot for Z_1

total travel time. The proposed methodology is illustrated using data from a delivery problem in Spain.

First, we study the influence of the importance assigned to the objectives in cost and satisfaction of customers. Models which first maximize the last one, give solutions which are prohibitive from the point of view of cost. On the contrary, we could get a solution with low cost and an acceptable deviation from soft time windows by properly selecting the order in which objectives are taken into account.

Second, we analyze the influence in costs and satisfaction of customers of factors like geographical distribution of customers, tightness of soft time windows, percentage of customers with soft time windows and importance of objectives. We conclude that grouping the customers decreases the cost irrespective of the order in which objectives are minimized. Also, allowing all customers to have soft time windows or tightening the time window increases the cost. Moreover, when restricting the attention to models with an acceptable cost, the most important factor in order to reduce costs is to group the customers. In contrast to this, if increasing customer satisfaction is the most relevant criterion then a proper assigning of importance to objectives is essential.

ACKNOWLEDGMENTS

This research work has been supported by Aragón Government under grant E58-DGA-2005-2007.

REFERENCES

Ball, M.O., T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds. 1995. *Network Routing. Handbook in Operations Research and Management Science vol. 8*. Amsterdam: North-Holland.

Calvete, H.I., C. Galé, M.J. Oliveros, and B. Sánchez-Valverde. 2007. "A goal programming approach to vehicle routing problems with soft time windows." *European Journal of Operational Research* 177 (3): 1720–1733.

Chankong, V., and Y.Y. Haimes. 1983. *Multiobjective Decision Making, Theory and Methodology*. New York: North-Holland.

Cordeau, J.F., M. Gendreau, G. Laporte, J.Y. Potvin, and F. Semet. 2002. "A guide to vehicle routing heuristics." *J. of the Operational Research Society* 53:512–522.

Gendreau, M., G. Laport, and J. Potvin. 1997. "Vehicle routing: modern heuristics." In *Local Search in Combinatorial Optimization*, edited by E. Aarts and J.K. Lenstra, 311–336. London: Wiley.

Golden, B.L., and A.A. Assad, eds. 1988. *Vehicle routing: Methods and Studies*. Amsterdam: Elsevier Science Publishers.

Montgomery, D.C. 2005. *Design and Analysis of Experiments*. 6th. New York: Wiley.

Solomon, M.M., and J. Desrosiers. 1988. "Time window constrained routing and scheduling problems." *Transportation Science* 22 (1): 1–13.

Steuer, R.E. 1986. *Multiple Criteria Optimization: Theory, Computation, and Application*. New York: John Wiley and Sons.

Toth, P., and D. Vigo, eds. 2002. *The vehicle routing problem*. l: SIAM Monographs on Discrete Mathematics and Applications, vol 9.

Zeleny, M. 1982. *Multiple Criteria Decision Making*. New York: McGraw-Hill.

A LINEAR PROGRAMMING MODEL FOR THE STORAGE LOCATION ASSIGNMENT PROBLEM – A CASE STUDY

Carla Geraldes
Industrial Management Department
Polytechnic Institute of Bragança
Campus de Santa Apolónia
5301-857 Bragança, Portugal
E-mail: carlag@ipb.pt

Sameiro Carvalho, Guilherme Pereira
Production and Systems Engineering
University of Minho
Campus de Gualtar
4710-057 Braga, Portugal
E-mail: {sameiro|gui}@dps.uminho.pt

KEYWORDS
Warehousing. Storage policy. Clusters. Linear programming. Case study.

ABSTRACT

An important warehouse operation decision concerns the definition of a storage location policy, including the definition of functional areas and the identification of each stock item location in the warehouse. Empirical approaches are often used. However, they do not guarantee that the lowest cost materials-handling solution is achieved. This paper presents a clustering approach to address the assignment of stock items to particular warehouse functional areas. A linear programming model is adapted to a case study and its performance is assessed through the use of real data. Model is solved using LINGO 9.0 mixed integer commercial solver. The output of this model may also be taken as a starting point for further detailed warehouse operation decisions, especially as far as the optimization of the order picking process is concerned.

INTRODUCTION

Warehousing deals with all material handling activities taking place in a warehouse. They include the receiving of goods, storage, order-picking, accumulation and sorting and shipping. Basically, we can distinguish two types of warehouses: distribution warehouses and production warehouses. According to Van der Berg (1999) a *distribution warehouse* is a warehouse in which products from different suppliers are collected (and sometimes assembled) for delivery to a number of customers. A *production warehouse* is used for the storage of raw materials, semi-finished products and finished products in a production facility.

Typically, distribution warehouses receive products - Stock Keeping Units (SKUs) - from suppliers, unload products from the transport carrier; store products, receive orders from costumers, assemble orders, repackage SKUs and ship them to their final destination. Frequently, products arrive packaged on large scale units and are packaged and shipped on small units. For example, SKU's may arrive in full pallets but must be shipped in cases.

Warehouse operation decisions frequently involve the definition of several functional areas and flows. Next, we briefly describe some of the most common functional areas and product flows (Figure 1):

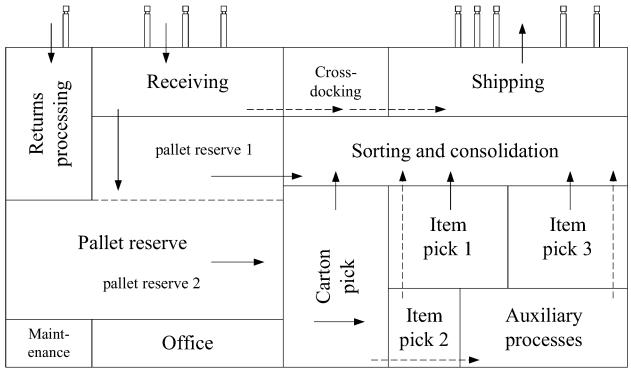


Figure 1: Example Layout of a Distribution Warehouse, (Salvendy 2001)

At the receiving area products are unloaded and inspected to verify any quantity and quality inconsistency. Afterwards items are transferred to a storage zone or are placed directly to the shipping area (this is called a cross-docking operation). We can distinguish two types of storage areas: reserve storage area and forward or picking area. The reserve area is the location where products stay until they are required by costumers' orders. The picking area is a relatively small area typically used to store fast movers products. Most of the flows between areas are the result of replenishment processes. Order picking is one of the most important functions in most warehouses. SKUs are retrieved from their storage positions based on customers' orders and moved to the accumulation and sorting area or directly to the shipping area. The picked units are then grouped by customer order, packaged and stacked on the right unit load and transferred to the shipping area.

As referred above, order picking process is concerned with the retrieval of goods from storage locations on the basis of customer's orders. This process is often the most laborious activity in a warehouse. On average the cost of order picking may represent about 55% of the operational costs in a warehouse (Tompkins *et al*. 2003). The efficiency of the order picking process depends on various factors such as the storage systems, the control mechanisms and the layout,

among others. However, in this paper, the layout is our main concern.

The time to pick an order includes travelling to storage locations, searching for the products, extracting the products, merging the products, recording the retrievals and preparing the shipment. With the ever-increasing variety of products and the increased customer service expectations, reducing order picking time is crucial. Obviously, the locations of the products in the warehouse have a major effect on the order picking time.

Traditional approaches to stock location problem ignore the characteristics of customers' orders. When considering the characteristics of customers' orders correlated information can be extracted, then it may be used for optimizing the stock location process. Based on the characteristics of the customers' orders, Liu (1999) proposed a clustering technique to group products for stock location and presented a mathematical linear programming model.

In this paper we address the storage location problem in a Portuguese distribution warehouse. In particular, we used a clustering technique to solve the storage location problem and find the structure of the clusters.

OPERATING POLICIES

Operating policies are decision rules that can be used to control processes in a warehouse, thus minimizing total handling costs. The effectiveness of the operating policies depends, apart from the efficiency of the policy itself, on factors such as layout, truck schedules that have to be met and order data (Roodbergen 2001).

According to Rouwenhorst *et al*. (2000) operating policies are short term decisions (1 year). Thus, for any warehouse it may be advantageous to re-evaluate the operating policies occasionally. For instance, if a change in the pattern of the customers' orders occurs it may be necessary to adjust the process of retrieving orders from storage. The main decisions at this level concerns assignment and control problems of people and equipment. One main decision that has to be made is the identification of the storage policy.

Storage policies assign SKUs to storage locations (Petersen and Aase 2004). Before products can be retrieved from storage, they first have to be stored. According to Goetschalckx and Ratliff (1990) a storage assignment policy is a set of rules which determines where products should be located in a warehouse. There are numerous ways to assign products to storage locations in the warehouse. Next, we describe some of the most frequently used types of storage policies: random storage, dedicated storage, turnover storage and class-based storage. This type of storage policies identification is well described in de Koster *et al*. (2007).

Products may be distributed in a warehouse randomly. For random storage policy, every incoming product is assigned to an empty location in the warehouse that is selected randomly from all eligible empty locations. If a warehouse manager employs this strategy, the first available location rule is commonly used. This policy guarantees that storage locations close to input/output (I/O) point are full. It may reduce time spent to store products and also congestion since neither the products' turnover nor the correlation between products is considered. However, if products with positive turnover are located in the back of the warehouse, then order picking times will increase. This storage policy will only work if a computer-controlled environment exists.

An alternative is to use a dedicated storage policy. According to this strategy each product is assigned to a fixed location. An advantage of this policy is that order pickers become familiar with product positions. A disadvantage is that a location is reserved even for products that are out of stock. Another possibility is to store products over the storage area according to their turnover. The products that are frequently ordered should be located close to an I/O point. Also, products with small size should be located near the I/O point so that as many products as possible are in the most accessible positions. A storage policy of this type is the cube-per-order index (COI) rule. The COI of a product is defined as the ratio of the space needed to store the product to the number of orders containing that product. Products with highest COI are located closest to the I/O point. The main disadvantage is that demand patterns vary permanently, thus each change would require a new ordering of products in the warehouse. This rule can also lead to increased congestion.

Finally, the class-based storage policy combines some of the preceding strategies. The idea is to group products into classes based on Pareto's method. Each class is then assigned to a dedicated area of the warehouse. Storage within an area is random. The advantage of this policy is that fast moving products can be located close to the I/O point and, at the same time, the flexibility and the low storage space requirements of random storage policy are applicable.

All storage policies discussed so far have not considered possible relations between products. For example, customers may tend to order a certain product together with another product. In this case, it may be interesting to store these products close to each other – complementarity. We can view this problem as a clustering problem. The aim is to group products together so that products in the same group have a high level of correlation. However, to apply clustering technique the correlation between products should be known or at least predictable. Liu (1999) proposed a binary IP formulation to clustering products in a warehouse.

A case study is presented in the next section. The use of this case example allowed us to discuss the impact of this strategy on a real problem and simultaneously redefine the storage policy of the distribution warehouse.

CASE STUDY

F&F, located at Guimarães, north of Portugal, manufactures

and distributes house appliances to about 1200 customers around the country (hypermarkets (85%), retail outlets (10%), hotels (1%), and others (1%)) and also some customers in Spain and Africa (3%). Over 25% of the products are acquired locally (including cutlery supplied from a factory owned by F&F), 20-30% is shipped from China and the remaining products are supplied from European countries. F&F processes an average of 40 customer orders per day (20 items per order, on average).

The company warehouse is a 4000 m^2 facility. Products are stored in pallets over 4000 m^2 of storage aisles (with 4 storage levels in the vertical direction), corresponding to 2000 storage spaces. The facility is divided into two main sections: picking area and assembly area.

A team of 15 workers is involved in warehouse operations which include receiving, storage, picking consolidation, sorting and shipping and some value-added activities (labelling, repackaging, etc). Receiving, storage operations and picking operations were usually carried out during the morning period. During the afternoon, while picking operations are still in progress, orders were checked for completeness (including searching for missing items) by 3 teams and then packed and prepared for shipping. F&F outsources the transportation activity and therefore it is performed by a logistics provider.

The company was facing increasing difficulties in remaining competitive due to the high operation costs and low levels of productivity to maintain existing throughput rate (output) as a result of changes in customers demand pattern over the last years: increasing need of extra labour hours, constant need of outsourcing labelling and repackaging tasks. Additionally a decrease in service level was also reported as a result of i) difficulty in meeting customers demands of shorter delivery times; ii) high level of errors in order processing (some as a result of out of stock situations).

Finally, a major problem was the increasing need of extra storage space. This problem was mainly due to the increasing variety of products available and to large inventory levels required to face long lead-times (supplier lead-times can be up to 4 months long).

An ABC Pareto´s curve showed that, in the last few years, only 43% of the existing SKUs were processed (positive turnover) and from the total set of existing customers only 8 of them were responsible for 70% of the sales.

The lack of space and the lack of operations efficiency were mainly due to the large number of obsolete items and the presence of slow moving stock items which were eating up square foot after square foot with pallets laying down the aisles preventing pickers from reaching stock positions.

The study also showed that the largest portion of the warehouse operations time was required to find the products (10%) and to walk through the entire warehouse to process an order (60%). This was the result of several factors such as: the lack of any arrangement of items on picking route lists, the absence of an efficient layout of items and the lack of an integrated information system.

Furthermore, inefficient operations planning and the incapacity to establish demand forecasts was the source of wrong inventory management policies. It was also observed that codification was not used, some of the fast movers' products were located on difficult access positions and the different functional areas were not well defined.

The described situation confirmed the need of a redesign of the warehouse and the redefinition of some operating policies. We approach this problem in two phases. In the first phase the priority was the elimination of all obsolete items, sizing the functional areas and allocating products to the areas. To do this, a mathematical model adapted from Heragu (2005) allowed us to size the functional areas and to allocate products to a functional area, minimizing the total handling costs. The results produced by the model defined a reserve area with 44% of the total space and a picking area with 15% of total space. Also 43% of the products were assigned to the reserve area, 12% were allocated to the picking area and over 45% of the products were located in both reserve and picking area (Geraldes, 2007).

As a result of this study F&F decided to redesign their warehouse: storage areas were identified and products were allocated to them. Although results revealed to be very useful in redesigning the warehouse many other problems were not yet solved. However, the output of this phase may be taken as a starting point for further detailed warehouse operations decisions, especially as far as storage location policy is concerned.

In the next step of this case study we used the mathematical model proposed by Liu (1999) to group and arrange products within the picking area.

CLUSTERING MODEL

Most material handling activities are labour-intensive and repetitive. In terms of layout for stock location, products may be grouped according to their complementarities, compatibility and popular criteria (Ballou 1999). Product grouping is the arrangement of items on stock location based on the mentioned criteria. Traditional approaches to store products in a warehouse ignore the issue of customers' orders. When considering the characteristics of customer orders, the order-item-quantity rule can be used for analysing the characteristics of items and customers (Liu 1999). In this study, a clustering technique proposed by Liu (1999) was used to group and arrange items for stock location.

Based on the characteristics of the customer orders, and recalling that 8 customers are responsible for about 70% of sales, similarity measures were defined and the mathematical programming model was used to find the structure of the clusters.

Clustering means generating groups of products. There are many ways to quantify the correlation between pairs of items. For the considered problem we took into account two similarity coefficients. The first one based on the order item quantities and a second one based on the customer order item quantities. Once the sets of similarity coefficients are defined, the mathematical programming model can be used.

Let us define a binary variable $x_{ij} = 1$ if product or customer i is assigned to cluster j and 0 otherwise. Let also y_j be a binary variable equal to 1 if product or customer is chosen as a median cluster and 0 otherwise. A median cluster is the item that represents all items in that cluster.

The general formulation of the model can be stated as:

$$\max \sum_{i=1}^{P}\sum_{j=1}^{P} s_{ij}x_{ij}$$

$$\sum_{j=1}^{P} x_{ij} = 1 \quad i = 1,...,P$$

$$x_{ij} \le y_j \quad i,j = 1,...,P$$

$$\sum_{j=1}^{P} y_j = K$$

$$x_{ij} = 0 \text{ or } x_{ij} = 1 \quad i,j = 1,...,P$$

$$y_j = 0 \text{ or } y_i = 1 \quad j = 1,...,P$$

S_{ij} represent the similarity measure adopted and K is the number of clusters under consideration. The objective function is used to select K products as medians such that the sum of similarity measures from all products to their respective median is maximized. The constraints are used to restrict the number of clusters to K and to assign each product to a cluster, but only to a cluster that has actually been designated.

Application to the F&F case study

In order to improve the materials handling efficiency, similarity measures for products and customers were considered. In this case study we used two different correlation measures. Therefore, grouping is based on the probability that any pair of customers order the same products – s_{ij}. Further, grouping was based on the probability that any pair of products appear in the same order – ss_{ij}.

In this study and for the time period under consideration, 8 out of 542 customers were considered. Based on the first proposed measure a similarity matrix between customers was constructed (Table 1).

Each element in the similarity matrix (s_{ij}) represents the probability that the respective pair of customers ordering the same products. For instance, we found that the probability of customers C1 and C8 order the same items is 48%.

Table 1: Correlated Matrix for Customers

	C1	C2	C3	C4	C5	C6	C7	C8
C1	1.0							
C2	0.02	1.0						
C3	0.04	0.46	1.0					
C4	0.23	0.33	0.45	1.0				
C5	0.01	0.25	0.19	0.43	1.0			
C6	0.09	0.11	0.31	0.75	0.17	1.0		
C7	0.08	0.09	0.51	0.05	0.27	0.04	1.0	
C8	0.48	0.05	0.29	0.32	0.13	0.15	0.05	1.0

The same approach was then used to define clustering structure for products. Data were collected from the 184 products allocated to the picking area and similarity matrix based on the probability of any pair of products appearing in the same customer order was constructed (Table 2).

Table 2: Correlated Matrix for Products

	P1	P2	P3	P4	P5	…	P184
P1	1.0						
P2	0.49	1.0					
P3	0.55	0.89	1.0				
P4	0.49	0.69	0.46	1.0			
P5	0.16	0.85	0.28	0.10	1.0		
…	…	…	…	…	…	…	
P184	0.26	0.09	0.43	0.38	0.81	…	1.0

Now, each element in the similarity matrix (ss_{ij}) represents the probability of the respective pair of items appearing in the same order. For instance, products P184 and P5 appeared in the same order in 81% of the analysed orders.

Results

The model was solved considering the two correlated measures and the following structures of clusters were obtained. The clustering results for customers are presented in Table 3.

Table 3: Clustering Results for Customers

Number of clusters	Grouping of Customers
1	{C1, C2, C3, C4, C5, C6, C7, C8}
2	{C2, C3, C7};{C1, C4, C5, C6, C8}
3	{C1, C8};{C2, C3, C7};{C4, C5, C6}
4	{C1, C8};{C2, C3, C7};{C4, C6};{C5}
5	{C1, C8};{C2};{C3, C7};{C4, C6};{C5}
6	{C1};{C2};{C3, C7};{C4, C5};{C6};{C8}
7	{C1};{C2};{C3};{C4, C6};{C5};{C7};{C8}
8	{C1};{C2};{C3};{C4};{C5};{C6};{C7};{C8}

For example, if we assume three clusters, products ordered by customers C1 and C8 must be stored close to each other. The same happens to products ordered by customers C2, C3 and C7 and also for products ordered by customers C4, C5 and C6.

When considering the correlated measure for products, a

different clustering structure was identified. Some of these clusters are displayed in Table 4, when 30 clusters were considered.

Table 4: Cluster Structure for Products

Grouping of Products
{P8, P37, P70, P90, P101}
{P103, P107}
{P11, P6, P42, P47, P87}
{P14, P51, P58, P81, P92, P112}
{P44, P60, P88}
…
{P52, P113, P120, P129}

Results from cluster analysis for products tell us that, for example, products P103 and P107 must be located next to each other within the storage area. So, using this similarity measure the information on products that are located next to each other is well identified.

As a result of this study F&F decided to define their storage policy and results for cluster analysis based on the similarity between products were used. Although the model results were very useful for the stock location problem a sequencing rule was also necessary. Therefore, we proposed that the product type with the largest demand should be assigned to the location closest to the I/O point. If any unselected product belonging to the same cluster as the already selected product exists, it should be located to the adjacent empty slot of the rack. This rule was applied iteratively until all products were stored.

CONCLUSIONS

For real warehousing situations, the question of what type of storage policy to use is always a demanding task. In particular, this work addressed a correlated storage policy. Two different similarity measures were considered and clustering structures were obtained. Although the use of this clustering technique to solve the storage location problem is promising it does not explicitly consider an objective related to operating costs. Another disadvantage is the fact that sequencing rules are necessary for implementing the arrangement of the stock location.

A storage strategy is considered optimal if it minimizes the average time required to store and retrieve a product while satisfying the constraints of the warehouse system. Since routing decision is strongly interrelated with the storage policy, future research is needed on the order picking process in order to evaluate the performance of this storage strategy.

REFERENCES

Ashayeri, J. and Gelders, L. 1985. “Warehouse design optimization”. *European Journal of Operation Research* 21, 285-294.

Ballou, R. 1999. “Business Logistics Management”. 4th ed. Prentice Hall. Upper Saddle river, New Jersey.

Bartholdi, J. and Hackman, S. 2006. “Warehouse & Distribution Science”. Release 0.76. http://www.warehouse-science.com.

Berg, J. Van der et al. 1998. “Forward-reserve allocation in a warehouse with unit-load replenishments”. *European Journal of Operation Research* 111, 98-113.

Berg, J. van der et al. 1999. “Models for warehouse management: Classification and examples”. *International Journal of Production Economics* 59, 519-528.

Caron, F. et al. 1998. “Routing Policies and COI-Based Storage Policies in Picker-to-Part Systems”. *Internationall Journal of Production Research* 36, 3, 713-732.

Carvalho, J. et al. 2001. “A Auditoria Logística - Medir para Gerir”. 1st ed. Edições Sílabo, Lisboa.

Frazelle, E. 2002. “World-Class Warehousing and Material Handling. McGraw-Hill.

Geraldes, C. 2007. “Modelos de programação linear aplicados à gestão das operações de um armazém – caso prático”. M.S. Thesis. Dep. Production and Systems. University of Minho. Braga. Portugal

Goetschalckx, M. and Ratlift, H. 1990. “Shared Storage Policies Based on the Duration Stay of Unit Loads. *Management Science* 36, 9, 1120-1132.

Gray, et al. 1992. “Design and operation of an order-consolidation warehouse: models and application”. *European Journal of Operation Research* 58, 14-36.

Hackman, S. and Platzman, L. 1990. “Near-optimal solution for generalized resource allocation problems with large capacities”. *Operations Research* 38, 5, 902-910.

Hassan, M. 2002. “A framework for the design of warehouse layout”. *Facilities* 20, 13/14, 432-440.

Heragu, S. et al. 2005. “Mathematical model for warehouse design and product allocation”. *International Journal of Production Research* 43, 2, 432-440.

Koster, R.de and van der Poort, E. 1998. “Routing orderpickers in a warehouse: a comparison between optimal and heuristic solutions. *IIE Transactions* 30, 469-480.

Koster, R. et al. 2007. “Design and control of warehouse order picking: A literature review. *European Journal of Operational Research* 182, 481-501.

Liu, C. 1999. “Clustering techniques for stock location and order-picking in a distribution center”. *Computers & Operations Research* 26, 989-1002.

Mulvey, J. and Crowder, H. 1979. “Cluster analysis: an application of Lagrangian relaxation”. *Management Science* 25, 4, 329-340.

Petersen, C. 1997. “An Evaluation of order Routeing Policies”. *Journal of Operations and Production Management*, 17, 1, 1096-1111.

Petersen, C. and Aase, G. 2004. “Improving order-picking performance through the implementation of class-based storage”. *International Journal of Physical Distribution & Logistics Management*, 34, 7, 534-544.

Rao, A. and Rao, M. 1998. “Solution procedures for sizing of warehouses”. *European Journal of Operation Research* 108, 16-25.

Ratliff, H. and Rosenthal, A. 1983. “Order-Picking in a rectangular Warehouse: A Solvable Case of the Travelling Salesman Problem. *Operations Research* 31, 3, 409-421.

Roodbergen, K. and Koster, R. de. 2001. “Routing methods for warehouses with multiple cross aisles”. *International Journal of Production Research* 39, 9, 1865-1883.

Rouwenhorst, B., et al. 2000. “Warehouse design and control: Framework and literature review”. *European Journal of Operation Research* 122, 515-533.

Salvendy, G. ed. 2001. “Handbook of Industrial Engineering: Technology and Operations Management”. 3rd ed. John Wiley & Sons, Inc. Chapter 81, 2083-2109.

Tompkins, J. et al. 2003. “Facilities Planning”. John Wiley & Sons.

USING 3D SIMULATION IN AN INDUSTRY INTERNAL LOGISTIC PROCESS

Vik Pavel

Dept. of Manufacturing systems, University of Liberec
Studentská 2
461 17 Liberec 1,Czech Republic
E-mail: vikpavel@seznam.cz, frantisek.manlig@tul.cz

Luís S. Dias
José A. Oliveira
Guilherme Pereira
Dept. Produção e Sistemas, Universidade do Minho
Campus de Gualtar
4710-057 Braga – P, Portugal
E-mail: {lsd|zan|gui@dps.uminho.pt}

KEYWORDS
Logistic Trains, Milk Runs, Simulation, Arena, Lean Production, 3D Animation.

ABSTRACT

The paper describes the usage of logistic trains (LT) in an internal supply process. Modern industries frequently use LT to improve performance by reducing manual work, work in process, product switch times, etc. The LT design and process analysis are performed using computer simulation. The results of simulation are visualised by using 3D graphics and animation. This paper focuses on the use of computer simulation in Arena connected by 3D Arena Player which is used as the visualisation tool.

The project outcomes, based on the simulation optimisation, lead to the reduction of logistic costs and improvement in the availability of material in production lines. These in turn, reduce work in process material (by reducing stock levels and buffers in the production lines) and setting the minimum number of LT. The benefits of 3D animation are easy to state as they validate spatial layout, allow better designing of LT routes and could limit shock risk with different layout options.

DESCRIPTION OF THE MR

Production lines can be supplied by items through several methods. One of them is the Milk Run (MR) system (http://www.ipaslovakia.sk) which is a special type of direct transport. The main principle of MR is supplying goods from deposits to delivery points in fixed defined routes and times and also picking empty transport units (Oliveira et al. 2003; Oliveira 2007; Rodrigues and Oliveira 2004). It is very useful when connected with the kanban system. Figure 1 shows an example of the MR system in the automotive industry. It can be compared with the subway-train that is directed by a defined harmonogram where it stops at each station and takes in a few passengers while a few get out. The subway-train is almost never empty. This principle can be used in internal and external logistics. For example, delivering by forklifts is based on another principle – a taxi. The most used facilities for MR are the logistic trains consisting of one engine (locomotive) and few wagons (http://www. fantozzi.com). The idea comes from the past when the cars delivered milk at exactly defined times from the milk farms.

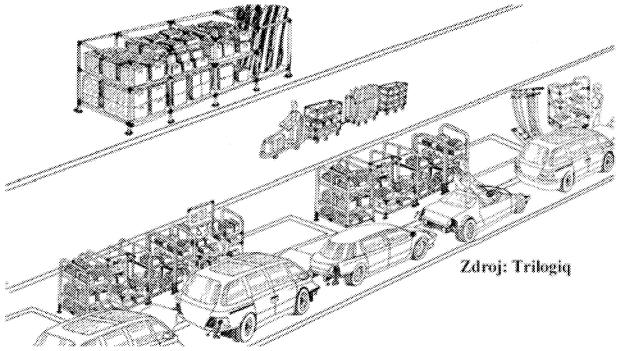


Figure 1: Example of MR usage

SIMULATION USAGE

The LT design and process analysis are performed using computer simulation. The computer dynamic simulation of production systems belongs to modern tools for solving logistic processes, which are illustrated by the created simulation model. By using models for research and experimentation, we improve our knowledge about the system, prior to implementation or changing the real system. The marketplace is awash with programs for discrete event simulation. One of the most popular systems is the Arena Simulation Tool by Rockwell Automation (Dias el al. 2007) (Kelton el al. 2004).

The project aim and the first model

This project describes a case of the usage of MRs in the electrotechnical industry (production of car radios).
The primary aims of the project were solved by simulation by determining the number of MRs for all production links, finding the best routes, structuring of delivered items and cycle times of each MR. The MR was solved by a simple simulation model using the transporter system. Figure 2 shows a screenshot of the Arena simulation model.

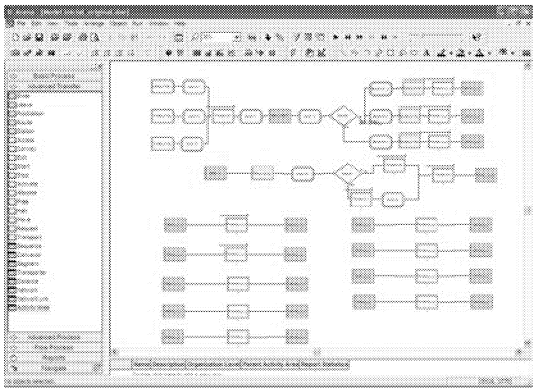


Figure 2: The screenshot of simple simulation model to solve the milk-runs problems

The inputs for the model are:

- Layout and distance matrix of each delivery point
- Suggested route sequences for MRs
- MRs – speed and number
- MRs scheduling

The model validated the estimated 20 minutes cycle time for complete supermarket load and production line supply and also the number of MRs.

Figure 3 shows a plant layout and suggested routes and delivery points (stations). The supermarket is a depot that contains prepared boxes with delivering items.

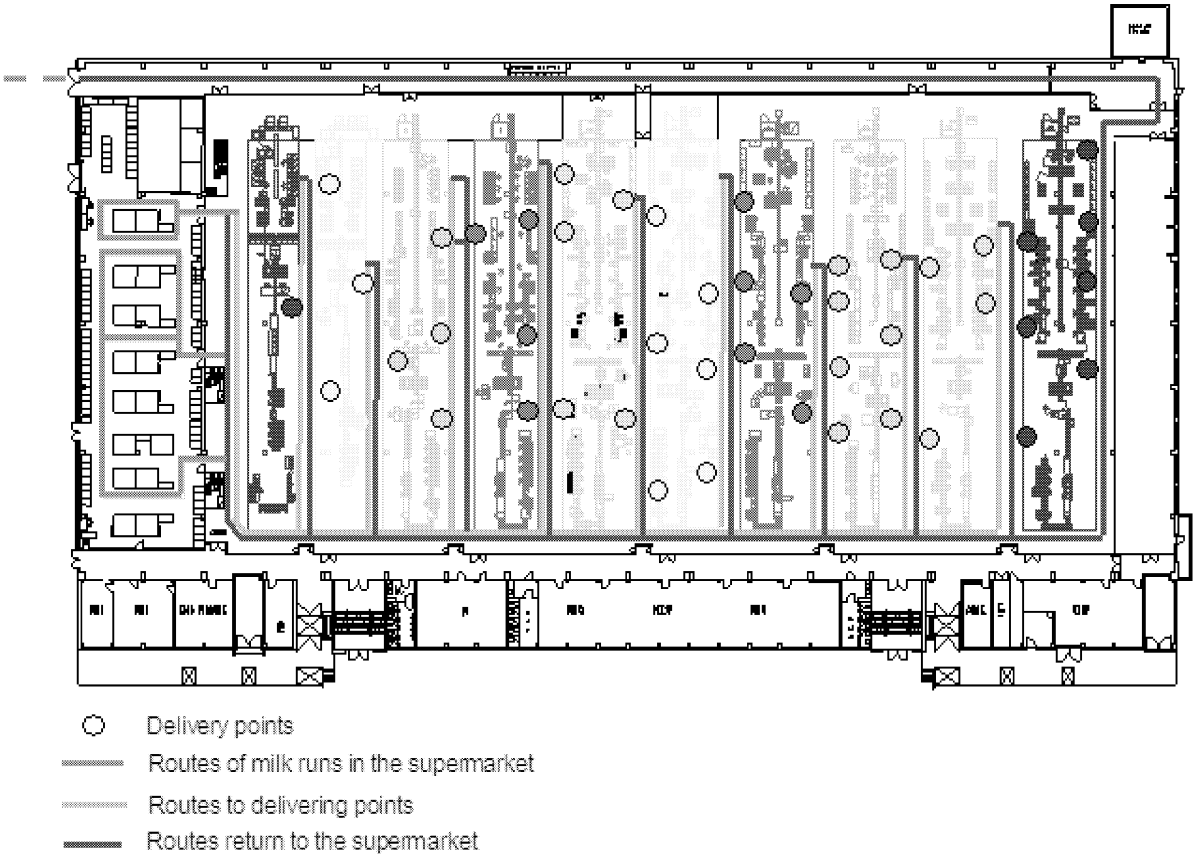


Figure 3: Plant Layout with routes and delivery points

USAGE OF 3D ANIMATION AND 3D SIMULATION

Animation (3D animation) creates an illusion of movement of objects with the help of fast gradual imaging of small different pictures. The most often used method is the frame-by-frame method, where the program forms individual pictures and saves them to a data medium (it is applied while rendering animation to the data-format AVI, MPEG, WMV or others). The second type of animation is an algorithmic one. Each frame is counted and displayed in real time by graphic algorithms (Vik 2005; Vik et al. 2005).

To complete the first approach and suggest the logistic system, a 2D simple simulation model has been used. Subsequently, a more detailed simulation model was prepared for creating 3D animation of MR processes. The Arena 3D Player was used to create 3D animation. It is a powerful post-process tool for creating and viewing 3D animations of Arena simulation models. The 3D Player enables both method of animation – algorithmically (with movements of the user captured in the 3D world and also frame-by-frame animation by just capturing the screen).

Creating a simulation model

The conveyor facility (Arena modules) has been used to track the movements of MRs in the model (instead of transporter modules). The main reason for choosing the conveyor facility instead of transporters is the conveyor's

ability to easily define movements and also capture the realistic movements of MRs. The model uses complete logistic trains with one pusher and a number of wagons. Simpler models use only one transporter representing the entire train.

Generating playback file

The two most popular approaches to connect simulation data with graphics are the online and post-process approaches.
One example of the on-line principle use is Witness, a simulation tool using the Witness Virtual reality tool (creation 3D world and animation). It means that simulation and 3D animation are calculated and displayed at the same time. It is very convenient and useful. For example, changes in simulation are immediately managed in animation without the time delay caused by generating the file. The disadvantages are that there is more dependency on computer power and the animation could also be less realistic. http://www.lanner.com/en/simulation_professionals /witness_vr.php).

The second approach, post-process, is based on connection with another external tool for creating animation using a file, e.g., connection between the SimPro simulation tool with 3D graphics and visualisation software 3ds max. This connection results in very realistic graphics (shadows, lights, textures, etc.). The disadvantages are that it takes a longer time for rendering animation depending on the level of the model's details.

The connection between Arena simulation and 3Dplayer is managed by a pbf (*PlayBack File*) file format. It contains information about the movements and changes in objects (dynamic changes).
A tree view pane (see Figure 6, on the left side of picture) provides an organisation map of the 3D animation objects and model. It provides the list of objects that have been defined in the simulation, control over visibility of each category and select specific objects for editing. The 3D Player contains a shape´s library which can be filled with other objects (VRML models). It is also possible to create or change the models in simple editor that has been provided.

Creating 3D world

Figure 4 shows the images in the whole procedure of 3D animation creation. If a plant layout exists, it is better to import it first for easier building of the 3D world. The main reason is that Arena 3Dplayer is unit- less with respect to distance (distance units can be inches, meters, furlongs, etc.) so the imported layout (with a fixed proportion) is useful to provide a good proportion between other parts of the 3D world. It also proves useful as a reference when inserting other parts.

The list of parts that it is possible to insert is shown in the tree view pane. These could be divided into parts contained in the simulation model (red colour – resources, stations, transporters etc.) and static parts contained only in the 3D world (blue colour – models of parts of production links, machines, walls, columns etc).

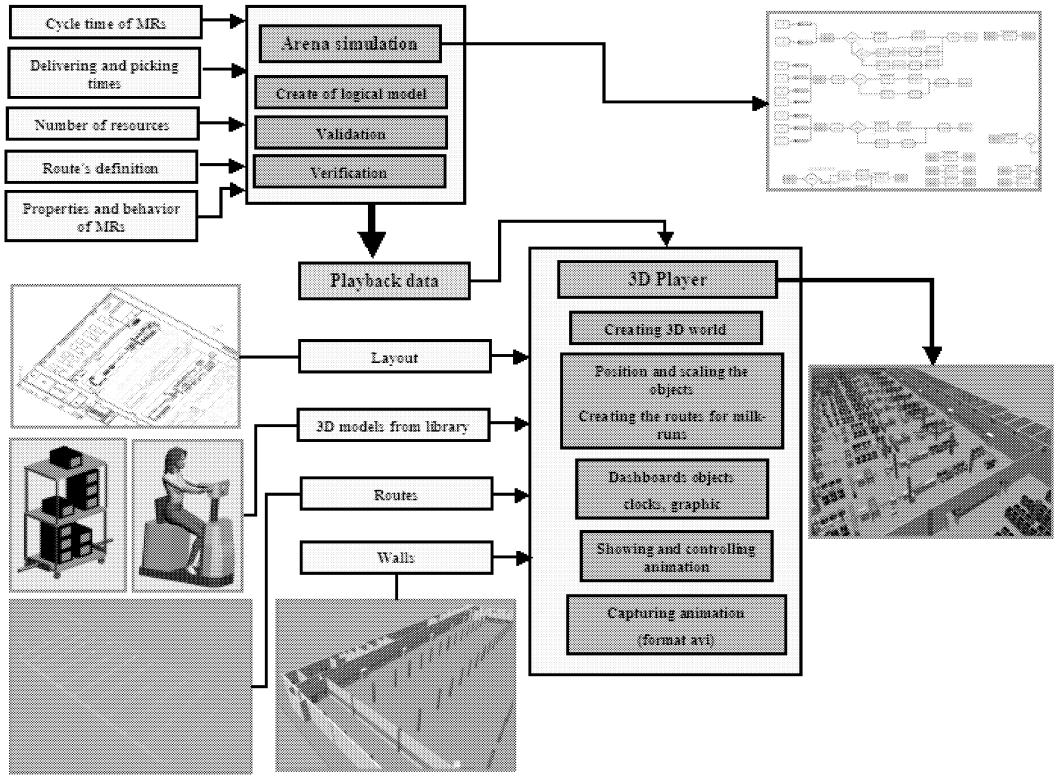


Figure 4: A process chart of creation Arena 3D animation (post process)

Purple coloured items denote inclusion in both simulation model and 3D world. A library of default elements exists which can be extended or changed in a simple but effective editor. When extending the default elements, only 3D models in VRML 97 or VRML 2.0 format can be used. There are also some restrictions in importing models – e.g., 3D Player does not support extrusions, built animations and textures from VRML files.

The routes and tracks of conveyors, transports, queues or distances are created in the 3D world like a sequence of points. These sequences can create more realistic shapes of routes in the real world (see Figure 5).

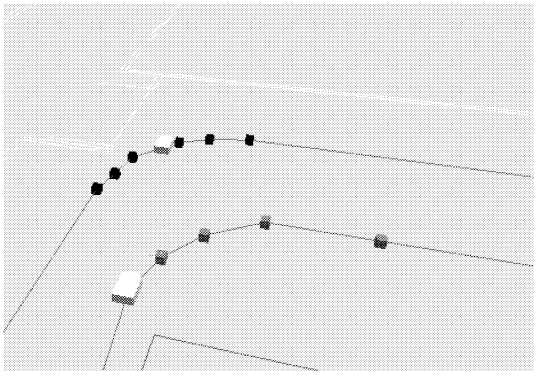


Figure 5: Creation of routes

Viewing the animation

It is possible to use standard functions (start, stop, time shifts or change of animation speed) for controlling the animation. Figure 6 shows the 3D world with standard functions.

CONCLUSION

The main advantages of using Arena simulation tools and Arena 3D Player (compared with other 3D simulation (SimPro, Witness, ShowFlow, Enterprise Dynamics) are the following:

- Speedy creation of the 3D world and animation
- Realistic and fluent movements of dynamic parts
- Ability to use the model's library (adding and editing)
- Ability to use the structure tree of 3D elements
- Ability to use the dashboards objects to view values of variables during animation (time, graphs, texts, etc.)

The main disadvantages are the following:

- Post-process connection between simulation and graphics data
- The model supports only VRML97 and DXF data format and with some restrictions imposed during importing.
- Export of graphics data (3D world, layout, created models from library) is not allowed.
- Animation can be exported only to AVI format by capturing the screen.
- The shape editor supports only basic shapes (box, cone, cylinder, sphere) and their transformations (translation, rotation, scaling).
- The 3D model cannot be automatically created (generated) by using the library.

The main disadvantage could be data export. 3D graphic and animation can be used for validation of spatial layout. In this model, it is possible to create optimal placement of resources. However, no feedback to CAD is possible.

Using computer simulation is an ideal tool for design, optimization or testing of logistic processes. It is often very hard to describe some results gained by simulation– e.g., suggestion of new controlling processes, function of machines or principle of new layout design. For these occurrences it is very useful to apply 3D animation and virtual reality.

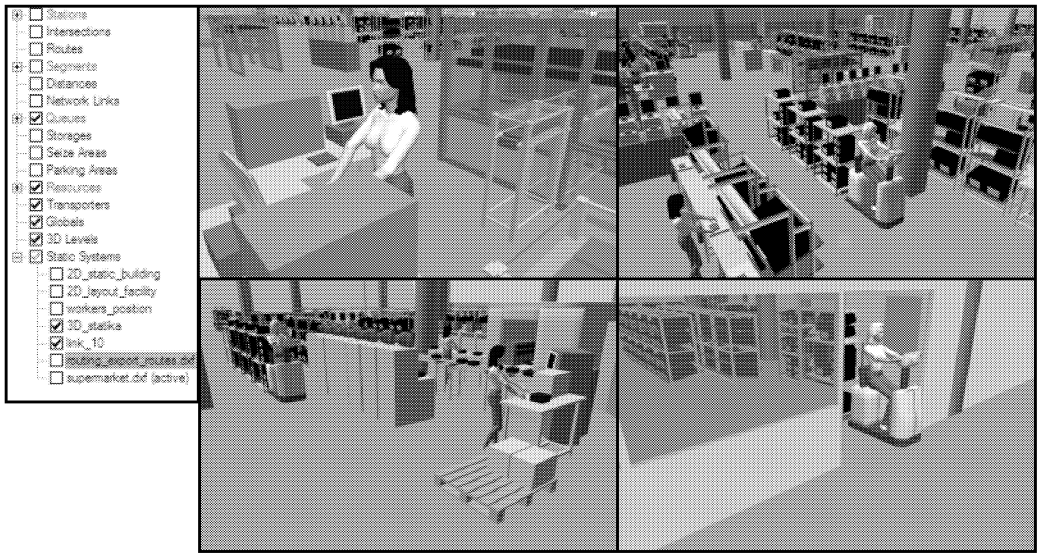


Figure 6: Shows of 3D simulation (the tree view pane and frames from the animation)

The main advantages of animation are larger plasticity of visualization which is the opposite of the simulation model that contains insubstantial information for outsiders.
Using 3D animation and VR is familiar to people of all professions – managers, operators, specialists and also absolute outsiders. So the whole production team understands process´ problems easily.

Summary of using 3D animation in the manufacturing processes:

Advantages of 3D simulation:

- Better quality of designing layouts and better detection and stripping of pertinent errors in layouts.
- Increasing communication in the planning team.
- Increasing skills of new workers.
- Increasing qualities of technical setting.
- Easier consultation with supplier companies.
- Easier presentation of results obtained by simulation
- Better comparison between planned situations and actual realization
- Easy archiving and retrieval.

Disadvantages:

- Elaborate process (creation of 3D models and libraries, rendering of animation)
- Complex hardware requirements
- Expensive software requirements
- Required skills in these programs

This part of project is managed for the first-time usage. The subsequent parts of the project will focus on the creation of more realistic animation (e.g., blinking of alert lights, moving the workers and items, changing numbers of carried items). As 3D Arena does not support these actions directly, they will be managed through VBA which is the programming language integrated in Arena Simulation Tool. The other possibility is generating own data for animation (instead of pbf) and creating 3D animation in another 3D tool – e.g. professional 3D studio max (using Maxscripting) or free Blender (Python scripting). That spreads possibilities of using more realistic 3D animation. Both mentioned graphic programs also support 3D engine for faster rendering of 3D scenes.

REFERENCES

Costa, B.; L. Dias, J.A. Oliveira, G. Pereira. 2007. "Modelos de Simulação do Sistema de Abastecimento de Linhas de Produção". *Actas do VIII Congreso Galego de Estatística e Investigación de Operacións*, Eds. B. Mendez, P. Miguel, I. Jurado, B.López, C.Sallero,153-158. ISBN 978-84-690-9136-4.

Dias, L.S. 2005. Modelação Automática Interactiva de Simulação. Tese de Doutoramento. Universidade do Minho.

Dias, L.S.; A.J.M.G. Rodrigues; and G.A.B. Pereira. 2005. "An Activity Oriented Visual Modelling Language with Automatic Translation to Different Paradigms". Proceedings of the 19th European Conference On Modelling And Simulation (ECMS 2005), Riga, Letónia, 452-461.

Dias, L.S.; G.B. Pereira; and A.G. Rodrigues. 2007. "A Shortlist of the Most Popular Discrete Simulation Tools". *Simulation News Europe*, vol 17, nº 1, 33-36. ISSN 0929-2268.

Kelton, W.D.; Randall, P.S.; and D.T Strurrock. 2004. Simulation With Arena. (3rd edition) McGraw-Hill.

Oliveira, M.; M.S.F.B.S. Carvalho; and G.A.B. Pereira. 2003. "Warehouse Operations: a case study", BEX´03, Business Excellence I, Performance Measures, Benchmarking and Best Practices in New Economy, (Guimarães Jun. 10-13) 577-579.

Oliveira, J.A. 2007. "Scheduling the truckload operations in automatic warehouses". European Journal of Operational Research, 179, 3, 2007, 723-735.

Rodrigues, P. and J.A. Oliveira. 2004. "Scheduling Operations to Minimize the Total Weighted Tardiness in an Automated Warehouse System". Proceedings of the 21st International Manufacturing Conference (Limerick Sep. 1-3), 251-258.

Vik, P. 2005. "Využití 3D animace a virtuální reality v počítačové simulaci". TU v Liberci - KVS 2005. Diploma thesis.

Vik, P.; J. Štoček; F. Manlig; and R. Havlík. 2005. "Využití 3D animace při návrhu výrobního procesu". *In Modelování a optimalizace podnikových procesů 2005*. Sborník referátů, Plzeň: ZČU v Plzni.

http://www.fantozzi.com - pages of producer MRs

http://www.ipaslovakia.sk - pages focused to industrial engeneering and logistics

http://www.lanner.com/en/simulation_professionals/witness_vr.php - pages about Witness Virtual reality toolbox.

BIOGRAPHY

PAVEL VIK was born in 1982 in Mladá Boleslav, Czech Republic. He studied at the Technical University of Liberec where he obtained his MSc degree in Manufacturing Systems, in 2005. He is making doctoral studies in Manufacturing Systems and Processes and he is attending a ERASMUS-SOCRATES program at the University of Minho. His main research interests are 3D animation and virtual reality in design of production systems and vizualization of computer simulation.

LUÍS S. DIAS was born in 1970 in Vila Nova de Foz Coa, Portugal He graduated in Computer Science and Systems Engineering at the University of Minho, Portugal. He holds a PhD degree in Production and Systems Engineering from the University of Minho, Portugal. His main research interests are Operational Research and Simulation.

JOSÉ A. OLIVEIRA was born in 1966 in Matosinhos, Portugal. He studied Mechanical Engineering at the University of Porto, Portugal. He graduated with a Ph.D. in Production and Systems Engineering at the University of Minho, Portugal. His main research interests are Optimization with Heuristic Methods in Systems Engineering.

GUILHERME PEREIRA was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management in the University of Minho, Portugal. He holds an MSc degree in Operational Research and a PhD degree in Manufacturing and Mechanical Engineering from the University of Birmingham, UK. His main research interests are Operational Research and Simulation.

CASE STUDIES

SIMULATION BUILDING BLOCKS FOR BUSINESS ENGINEERING OF MOBILE FIELD SERVICES

Tennov Simanjuntak
Del Polytechnic of Informatics
Jl. Sisingamangaraja, Laguboti
Tobasa 22381 – Sumatera Utara, Indonesia
tennov@del.ac.id

Yan Wang
System Integration & Technology, Accenture
Gustav Mahlerplein 90
1082 MA, Amsterdam, The Netherlands
Yan-nl.wang @accenture.com

Marijn Janssen
ICT Section
Faculty of Technology, Policy, and Management
Delft University of Technology
Jaffalan 5, 2628BX Delft, The Netherlands
m.f.w.h.a.janssen@tudelft.nl

Alexander Verbraeck
System Engineering Section
Faculty of Technology, Policy, and Management
Delft University of Technology
Jaffalan 5, 2628BX Delft, The Netherlands
a.verbraeck@tudelft.nl

KEYWORDS
Discrete Event Simulation, Building Blocks, Business Engineering, Mobile Services

ABSTRACT

The use of mobile solutions aimed at improving field service processes have come to the attention of managers in many industry sectors. In the Internet age, the process starting with having a vague idea and ending after implementation needs to be run through within an ever decreasing time frame. To get a better grip on changing business processes and their added value, managers are looking for ways to gain quickly insight into various options to support their decision making. In this paper, simulation building blocks are proposed as a solution to improve process analysis and modeling in the domain of mobile field services. They are aimed at providing insight into the dynamics of different process models and supporting quantitative analysis. The findings show that the presented simulation building blocks satisfy the functionalities desired by the problem owners and fulfill the requirements for building blocks.

INTRODUCTION

The need for business mobility has been the ground for managers to turn the attention to mobile solutions adoption. Mobile solutions provide "mobile services" to support business mobility by enabling reachability and access to computing power from remote areas, where conventional information technology cannot deliver (Gebauer and Shaw 2002). These services allow business transactions and information exchanges to be conducted from anywhere at anytime (Davis 2002; Lyytinen & Yoo 2002). As a result, the flexibility of organizations can be increased and in this way the productivity of resources, especially of field-service workforces, can be improved.

Engineering mobile field services is not a simple task. Of the factors hampering mobile solutions adoption is managers' lack of "know-how" in maximizing business values offered by mobile services (Lehmann et al. 2004). Managers face difficulties in realizing business transformation and process redesign (Wang et al. 2005a). Therefore, a simulation based support tool is needed that is usable and useful for managers to engineer mobile filed services (Wang et al. 2005b).

In this research, we develop a prototype of a support tool based on the concept of building blocks taken from domain specific simulation. The research objective is to develop domain specific simulation building blocks for business engineering of mobile field services. The basic premise underlying this vision is that the provision of building blocks will accelerate the construction of simulation models for mobility enhanced field service processes.

ARCHITECTURE OF BUILDING BLOCKS FOR SIMULATION BASED INQUIRY

Simulation plays an important role in business process engineering endeavors. The use of simulation creates added values by supporting the integration of dynamic aspects of business processes modeling and facilitating the anticipation of future changes in process design (Giaglis et al. 2005). In this way, simulation is able to reduce the risk of direct implementation failure.

Valentin and Verbraeck (2002) propose an architecture consisting of building blocks for constructing domain-specific discrete event simulation-based environments. They argue that developing simulation-based environments within the same domain involves similar modeling constructs, which are used repetitively. Thus, it is better to structure and package the repetitive process and make them available to the modeler for continual use. Verbraeck et al. (2002) identified set of requirements that must be fulfilled by a building block. The most important requirements are as follows.

- *Quality:* provide necessary information for the users, e.g. the explanation of conceptual design and the procedure for simulation model construction.
- *Speed*: support rapid development of simulation models.

- *Maintainable and extendable:* easy to maintain and to extend.
- *Generality:* applicable for several cases within the same domain.
- *Reusability:* multi-time usability of building blocks and the designs in several implementations.

These requirements will be used to evaluate the developed building blocks.

RESEARCH METHOD

To pursue our research objective, we used the research approach depicted in
Figure 1. The approach is organized into five main steps: 1) literature study, 2) case study, 3) design, 4) implementation, and 5) evaluation. Firstly, a literature study was conducted to investigate business-engineering, simulation, and the use of mobile services in businesses. Next, two case studies were investigated to provide input for building blocks design. The design process is comprised of conceptualizing the problem domain, generalizing this and then specifying the building blocks. Conceptualization of the problem domain is aimed at a reduction in the complexity of the specific domain, and results in conceptual models for each case. Based on the conceptual models of the cases, we generalized the necessary building blocks of the domain, and derived our first conceptual model of building blocks. The design step ends with the specification of the model, which include selecting the simulation environment into which the building blocks should be implemented.

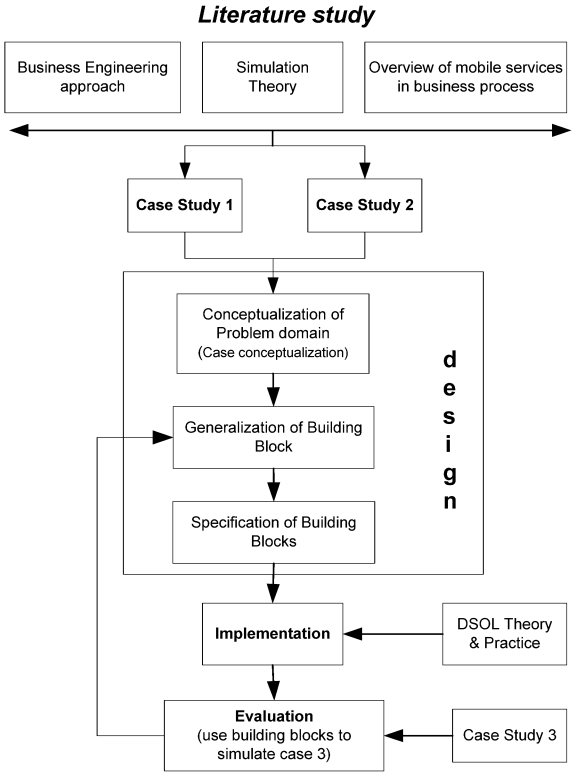


Figure 1: Research approach

In implementation, the specified building blocks conceptual models were implemented using a programming language. In turn, the building blocks were used to construct a simulation model for the evaluation case. Next, we evaluated the fulfillment of the requirements for the building blocks. We included a feedback loop to the generalization of the building blocks to enrich the generality, given the insights gained from the evaluation case.

CASE STUDY DESCRIPTION

The case study was aimed at gaining knowledge of the problem domain of mobile field services. The problem perception of managers regarding the adoption of mobile solutions was investigated. Furthermore, we identified the business processes in which significant improvements could be made by making use of mobile services. To ensure the generality of the building blocks, we performed two case studies. The first case study concerns the security and patrolling sector (G4 Falck case) and the second was in the logistic sector (KCII case). G4 Falck is the leading security and patrol company in Belgium and it provides security patrol and response service for its customers. KCII is a global medical technology company which operates in the rental field, renting medical appliances to its customers, which are mainly hospitals.

From the study, we learned that mobile services can be used to streamline and automate the process of service planning for order delivery. Order delivery can be assumed to be an alarm intervention in the G4 Falck case, or the delivery of product in the KCII. We found that the critical change caused by the use of mobile solutions were to be found in the dispatching processes. The existing dispatching policy (without mobile services support) is highly dependent on the *availability* of mobile workers. This means it relies on information about the "busy" or "not busy" status of mobile workers which is acquired through traditional mobile communication terminals, such as a mobile phone. A PDA (Personal Digital Assistance), with an online function, can be used as a terminal for field workers to access enterprise information system remotely. The embedded GPS receiver of PDA can enable a location-based service (LBS), which allows the dispatcher to track the real location of a worker, and even to trace the remaining time to finish their current assignments. Hence, a mobile solution is proposed to support the dispatching decision with more real time information on *availability*, *distance to order site*, and *remaining time of current assignments* for each field worker. Based on this information, the dispatching decision will be made on the *ETA (Estimated Time of Arrival)* which is the summation of the remaining time of a field worker on a current assignment and travel time.

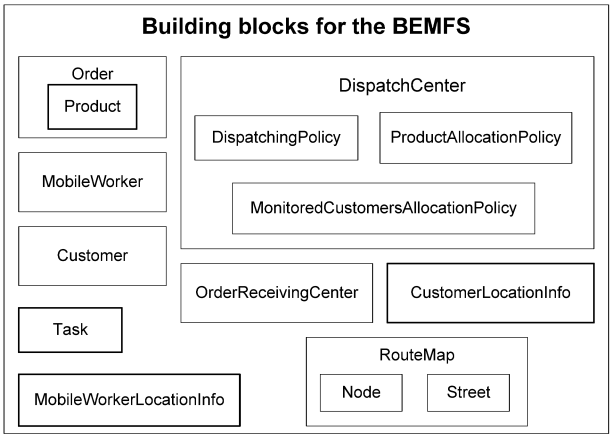
Mobile solutions offer benefits by enabling dynamic dispatching mechanism and increasing immediate responses to order. Yet, the managers from the cases were reluctant to make any investment decision for a mobile solution. They wanted to be provided with insight into the proposed dispatching process. Furthermore, the managers questioned if the investment cost could be recovered from the

performance improvements, in term of response time and resource utilization.

DESIGN AND IMPLEMENTATION OF BUILDING BLOCKS

We opted for Object Oriented Analysis and Design technique for conceptualization (Meyer 1997). Firstly, a conceptual model for each of the cases was constructed and followed by generalization of building blocks conceptual model (Simanjuntak 2006). The generalization is done by finding functional similar classes, means similar operations and attributes, to form a generic representation of those classes. Building blocks are direct translations of the generic classes. The next step is to specify the building blocks within the framework of DSOL (Distributed Simulation Object Library), a Java-based simulation environment created at Delft University of Technology (Jacobs 2005). We opted for DSOL as it is an open source simulation environment that gives freedom for model developers to extend libraries for domain specific purposes and supports multi-formalisms modeling. The specification is made upon the simulation engine DEVSSimulator provided in DSOL, which is a DEVS formalism-based simulator for event-scheduling. In the implementation step, the translation of building blocks conceptual model into computer language was done in the Java programming language. This resulted in the building blocks for modeling the mobile field services in terms of library of java objects. The overview is given in Figure 2.

There are fifteen building blocks produced. Each building block provides specific functionality for user to construct a simulation model. The *Order, RouteMap*, and *DispatchCenter* are the building blocks with different level of granularity. These are composed of other building blocks and form bigger building blocks with more functionality. The granular structure enables better maintainability and extendibility because the changes in part of the building blocks will not affect the others.



.

Figure 2: Overview of the building blocks

EVALUATION

The usability and usefulness of the building blocks was evaluated in practice. In addition, it was investigated whether they met the requirements for building blocks. For the evaluation purpose, we used a fictitious but realistic case of an: Enterprise Solution Vendor.

Case study background

The enterprise solution vendor, targets Small and Medium Sized Enterprise (SMEs). The company provides technical support at customers' sites if a problem is reported. The customer contacts the helpdesk by phone, fax, or email. Having received the call, the helpdesk contacts the technical support department to dispatch a technical support engineer. Subsequently, the engineer travels to the customer site and fixes the problem.

As the number of customers increases, the number of calls for technical support is increasing. A mobile solution is sought to increase the utilization. The proposed mobile solution can provide a service for tracking the locations of technical support engineers. It also enables real time data communication for the engineers to access the company database remotely. These services should enable dynamic dispatching and allocation of technical support engineer. However, there is a lack of knowledge about how to dispatch the technical support engineers using a dynamic dispatching scheme and the managers want to know what efficiency gains can be accomplished in terms of minimization of response time and maximization of utilization of their technical support engineers.

Simulation Model Construction

The test on usability will be the answer to the question: "Can we build simulation model for the aforementioned problem situation using the created building blocks?" The usefulness will be evaluated on whether the simulation model built on the architecture of building blocks provides insight into new dispatching process enabled by mobile services and whether the model produced the necessary information.

The simulation model construction started with drawing a conceptual model by mapping the building blocks abstract concept (conceptual model) to the Enterprise Solution Vendor case. Using classification and exemplification, we mapped the structure and process view of both systems. In structural mapping, the elements of the system related to the dispatching process together with their relationships were mapped. During the process mapping, the process interaction and message sequences between the elements were mapped. We found that the abstract concept could be directly translated to produce a conceptual model for the case with only minor changes. The process continued with specification within DSOL framework. We created artificial values for input variables and defined simulation properties. A map to locate all stationary objects in simulation model was also made. The map is depicted in Figure 3. There were fourteen customers distributed over different locations and served by three technical support engineers. The technical

support engineers are not shown on the map. At the beginning of simulation, all the technical support engineers are located at the office. The accessibility of the customers is marked by dashed line.

In the next step the conceptual model was implemented in DSOL. The simulation model was created by instantiating the building blocks. After the programming completed, the model was run. The runtime view of the simulation model is presented in Figure 4 and statistical output in Figure 5. Next, the verification was performed. The simulation model was verified by checking if the model represents the conceptual model, and if the input variables and simulation properties were coded correctly.

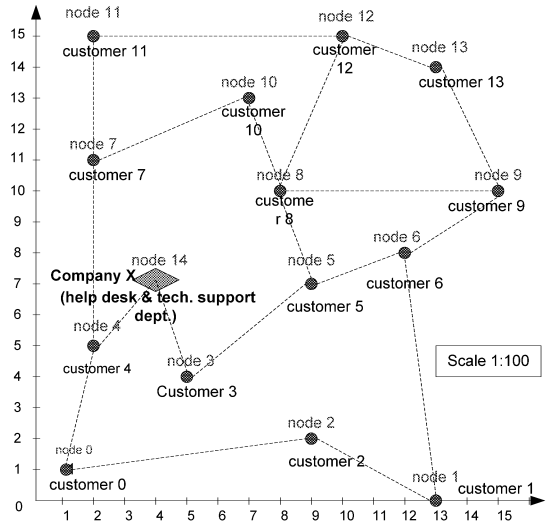


Figure 3: Location of stationary objects

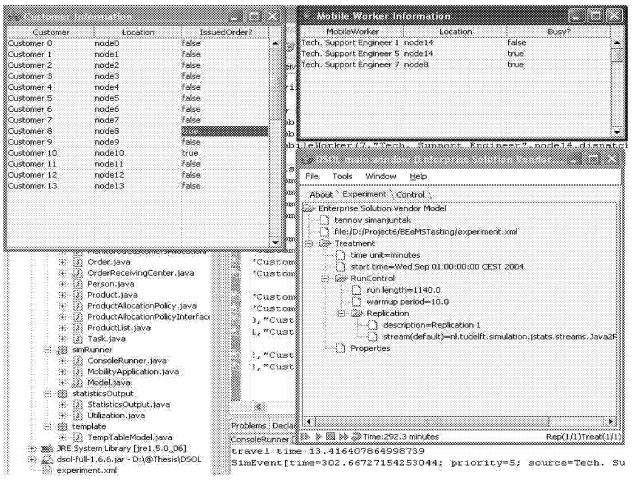


Figure 4: Runtime view of the model

Evaluation of Building Blocks

The case study illustrated the usability of building blocks. Our experience showed that the building blocks could be used to create a simulation model for modeling a dynamic dispatching process enabled by mobile services support. Furthermore, the evaluation also showed the usefulness of the building blocks. The simulation model created using building blocks provided system performance indicators in terms of response time and the utilization of mobile workers. Managers can use the outputs to justify their decision regarding mobile solution investment.

Requirement for Quality: This requirement is related to the provision of necessary information for the user who needs to construct simulation models by using the building blocks library. In this research, we provided documentation of the design and implementation of building blocks. We expect that the documentation can be used as guidance for a user to construct a simulation model for mobile field services related cases. However, due to time constraints, we were not able to test this.

Requirement for Speed: This requirement means that the use of building blocks must support the rapid development of a simulation model. In our case study, the use of domain specific building blocks enables the automatic generation of simulation models by mapping between the building blocks and conceptual model. Only a few changes in design and coding needed to be made. Furthermore, the model was constructed by merely creating instances of building blocks. As a result, the time needed for constructing the simulation model was significantly reduced.

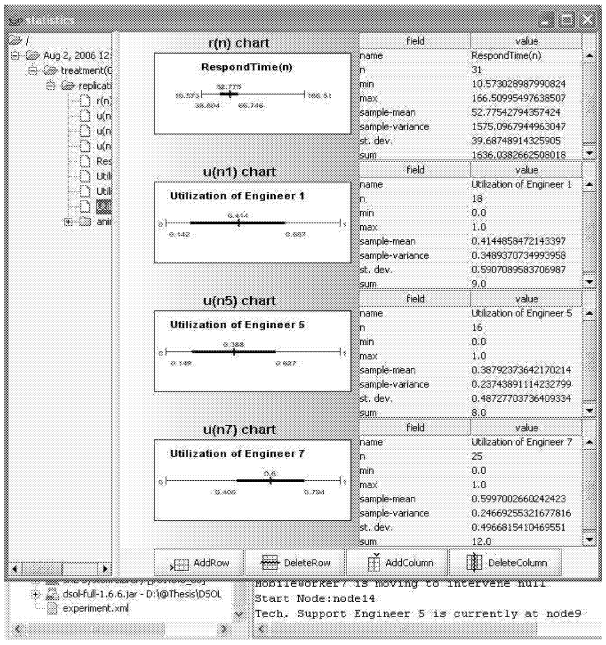


Figure 5: Statistics output of simulation model

Requirement for maintainability and extendibility: We designed the building blocks for certain functionality which was specific and different from other building blocks. As a consequence, the information contents (i.e. the attributes and operations) inserted into building blocks were kept to a minimum. This ensures that the changes in certain building block will only affect the internal logic and not the interfaces with the environment and in this way do not pose any changes to other building blocks. If we needed a building block which performs several functionalities, we composed several building blocks to form a building block at higher

level of granularity. For example, DispatchCenter and RouteMap (see Figure 2). This architecture enables better maintainability and extendibility of building blocks.

Requirement for Generality: the generality requirement was considered to be the applicability of the simulation building blocks for several cases within the same domain. This requirement was met as we were able to model a third case study. In this research, building blocks were created based on the inputs from two different cases from different business sectors. Furthermore, we focused on service planning and emphasized modeling dispatching process mechanism enabled by mobile services. We analyzed the differences between the two cases in terms of entities which built the system and process interactions between those entities. Afterwards we constructed generic building blocks which included the information and functionalities from the two cases.

Requirement for Reusability: reusability means the ability to reuse building blocks as software elements and as a design in several implementations or simulation studies. This can be operationalised as the capability of a building blocks library to be plugged into other solutions or the ability of a conceptual model to work as a framework to develop support tool for various cases. Reusable-building blocks means saving time during the development of simulation model. Reusability is related to the amount of information content put into a single building block. The more information content put into building block the less self-contained and reusable it is. In our design, we limited the information and functionality in the building blocks and created them at different levels of granularity. In this way, this architecture provides the building blocks with the capability to be reusable in several implementations. From the perspective of design, UML was used which is a general standards for the modeling and documentation of systems (Richter and Marz, 2000). This standardization enables the modelers and the users to communicate the design of simulation in terms of the process, requirements, and performance features of the simulation model regarding its structure and dynamics. UML enables researchers or modelers to understand the design of building blocks and to reuse the knowledge of the design as for their purposes.

CONCLUSION

In this research, we developed building blocks to facilitate the fast modeling of business processes in the domain of mobile field services. We demonstrated that our building blocks can be used tackle the complexity found in the domain and prevent the need to waste time by developing a simulation model from scratch.

Particularly, we implemented the building blocks, and tested them by simulating the mobility enhanced dispatching process in a given case. The building blocks were used in our test case study to model the situation quickly. The evaluation shows that the building blocks capture the requirements for quality, speed, maintainability and extendibility, generality and reusability. Our suggestion for future research is to enrich the values added of the building blocks by incorporating real and complex business process and involving real actors in the design process.

REFERENCES

Davis, G.B., "Anytime/AnyPlace Computing and The Future of Knowledge Work", in: Communications of the ACM, Vol. 45 No. 12, pp: 67-93, 2002.

Gebauer, J. and Shaw, M. J., "A Theory of Task/ Technology Fit for Mobile Applications to Support Organizational Processes", in: the 1st Workshop on e-Business (WeB), Barcelona, 2002.

Giaglis, G. M., Hlupic, V., de Vreede, G.J., & Verbraeck, A., "Synchronous Design of Business Process and Information System Using Dynamic Process Modeling", in: *Business Process Management Journal, Vol. 11, No. 5, 2005*.

Jacob, P.H.M., *The DSOL Simulation Suite; Enabling Multi-Formalisms in a Distributed Context*, PhD Dissertation, Delft University of Technology, 2005.

Lehmann, H., Kuhn, J., and Lehner, F., "The Future of Mobile Technology: Findings from a European Delphi Study" in*: Proceedings of the 37th Hawaii International Conference on System Science, Hawaii, 2004*.

Lyytinen, K. and Yoo, Y., "Issues and Challenges In Ubiquitous Computing", in: *Communications of the ACM, Vol. 45 No. 12, pp: 63-65 2002*.

Meyer, B., *Object-Oriented Software Construction*, Prentice Hall, 2nd Ed., 1997.

Richter, H. and Marz, L., "Toward a Standard Process: The Use of UML for Designing Simulation Model", in: Proceedings of the 2000 Winter Simulation Conference, J.A. Joines, R.R. Barton, K. Kang, P.A. Fishwick (Eds.).

Simanjuntak, T., *Simulation Building Blocks for Business Engineering enabled by Mobile Services*, Master Thesis, Delft University of Technology, 2006.

Valentin, E. and Verbraeck, A., "Simulation Using Building Blocks", in: *F.J.Barros and N.Giambiasi – Proceedings of AI, Simulation and Planning in Highly Automated Systems, 2002*.

Verbraeck,A.,Saanen,Y.,Stojanovic,Z.,Valentin, E., Meer, K. V. D., Meijer, A., and Shiskov, B., "What are building blocks?", in: *A. Verbraeck and A.N.W. Dahanayake (eds.), BETADE Research Program, TU Delft, 2002.*

Wang, Y.,Van De Kar, E.,Meijer, G., Hunteler, M., "Improving Business Process with mobile workforce solution", in: *Proceedings of International Conference on Mobile Business (ICMB), Sydney, 2005a.*

Wang, Y., Van De Kar, E., Meijer G., Hunteler M., "Designing Mobile Solutions for Mobile Workers-Lesson Learned from a Case Study", in: *Proceedings of International Conference on E-business 2005, Xian, 2005b.*

QUALITY MANAGEMENT IN BUILDINGS DESIGN PHASE

José Carvalho
Instituto Superior de Engenharia do Porto
Rua Dr. António Bernardino Almeida
4200 -Porto-Portugal …
Email: jmc@isep.ipp.pt

Jean-Claude Mangin
Gérard Sauce
Université de Savoie – Polytech'Savoie – LOCIE
.Chambery. France
jean-claude.mangin@univ.savoie.fr
gerard.sauce@univ.savoie.fr

KEYWORDS: Buildings Design, Quality Plan

SUMMARY: The design phase is unanimous considered as a source for many non-quality situations in buildings construction. In this paper, it is described a quality approach (Quality Plan) in the design phase, to ensure his technical quality, based on the principles of prevention, technical check and continuous improvement process. The Quality Plan, is developed on the basis of ISO 9000 (2000) Standars and includes a network of indicators, suitable for building design, to ensure the monitoring of the quality system. The data and records are handled by a computerized software tool, designed specifically for the purpose. The originality of the work, it is based, on a full-scale experiment on design, for tree project buildings in Portugal. Analysis and comparison of the results obtained, demonstrate the value of this approach, for the overall quality of building design.

1. INTRODUCTION

Numerous studies, have shown that the design phase have a most important impact on the final level of quality to be achieved. A French study developed by the Bureau Securitas, 1979, analyzed 10.000 non-quality cases in various buildings, between the years of 1965 and 1978 (cp. Table 1).

Design	Construction	Materials	Occupation
43%	43%	6%	8%

Table 1: Origin of non-quality situations in France

Another author, Costa, 1995, refers the results of a similar investigations carried out in several European countries (cp. Table 2).

Origin	B	UK	G	D	RO
Design	46%	49%	37%	36%	37%
Materials	15%	11%	14%	25%	22%
Occupation	8%	10%	11%	9%	11%
Construction	22%	29%	30%	22%	19%
Others	9%	1%	8%	8%	11%

Table 2: Origin of non-quality situations in various European countries

These results, clearly point the design phase in the origin for more than a third of non-conformities occurred in buildings and in all cases his responsibility is higher than the construction phase. It is the emphasis on the need to invest in this phase of the construction process, in action for the prevention and control.
Ironically, we find that the papers, studies and references available on the design activities, associated with the quality control process, are relatively rare. This paper, describes a methodology supported by the establishment of an operative Quality Plan, which is based a Quality Manual tailored to the activities of design and having as a reference the ISO 9000 (2000) Standards. The principles of prevention, technical control and continuous process improvement, are still present. A plan of testing, applied to tree real designs has been developed and supported a doctoral dissertation. It is important to note that this study is based on the Portuguese construction management reality.

2. METHODOLOGY FOR ENSURING THE TECHNICAL QUALITY OF DESIGN PHASE

2.1. The process approach

The structure of the Quality Manual consists in three parts:

- Introduction
- Identification and description of the processes
- Identification and description of the procedures.

Processes are defined by the ISO 9001:2001 Standard, as a set of interactive activities, which transform inputs into outputs. Normally, the output from a process leads to the input to the following process, knowing that its planning and implementation are place in a controlled manner to achieve value.
It was possible to identified management, realization and mixed (management and support) processes (cp. Table 3).

Realization Processes	Management Processes	Mixed Processes
P1-Analyze the contract P2 - Realize the design P3 - Provide technical assistance	P4 - Planning the Quality System P5 - Evaluate and improve the Quality System	P6 - Human resources management P7 - Material resources management

Table 3: The processes

The realization processes are critical to the activities of design phase. The Process P2 - Realize the design, consists in all activities required to develop the design, in order to provide the customer with the technical and legal requirements to execute the building construction. Our model includes five sub-processes:

- SP2.1 - Implementing the Quality Plan
- SP2.2 - Defining the outline project brief
- SP2.3 - Conducting project brief
- SP2.4 - Realizing the outline design
- SP2.5 - Realizing the detailed project brief

2.1. The Quality Plan

The Quality Plan was identified as the Sub-Process SP 2.1. During the development of a specific design, the Quality Plan must be done, thereby requiring the company to reflect on aspects of organizing, planning, prevention and check. The proposed plan, provides a software tool for monitoring the system, in order to obtain a set of indicators, true dashboard design's routing for quality.
Figures 1, illustrate a quality management route type. This pattern, presents the input information and the produced information in circles, and the handling actions in rectangles.

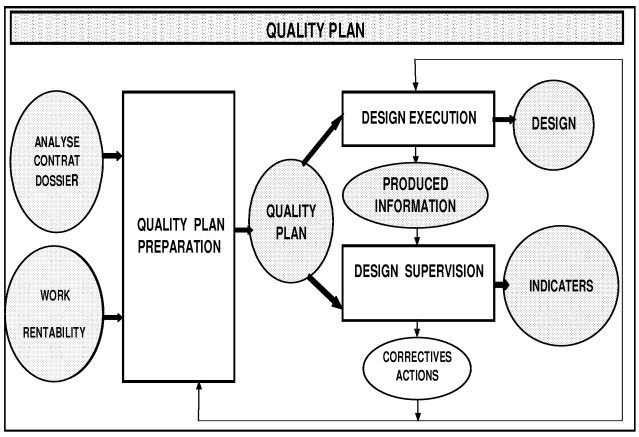


Figure 1: Quality Plan Model

Prepare a quality plan includes the following actions:

The design planning - The design planning is a key document of the Quality Plan. It defines the different stages of the design and the task duration. This first phase is planning the development of the Quality Plan and updating its own, and of course the other tasks such as collecting basic data, the sub-processes of the design, the tasks of production, the internal and external meetings, the check actions, the design activities and rectifications, the internal and external approvals, the monitoring and synthesis. As a result, the design planning provides expected hours of work, the costs and delays, by technical specialty, by intervening and by phase. These parameters will be compared during the development of the design, with the real data recorded. Table 4 summarizes information needed and produced by the design planning.

Input		Output
Dossier contract analysis	▶	Planning of activities and critical paths
	▶	Evaluation of human and material resources
Data labour productivity	▶	Evaluation of considered working hours
	▶	Costs estimation
Cost of labour hours	▶	Delays estimation

Table 4: Information needed and produced

Organizing design -Design teams must begin by setting the rules for its operation, with accurate characterization of the subject of his work, the detail list of all persons, the pattern of internal and external organization, the definition of responsibility and the circuit of information flows.

Develop preventive actions -These actions are a key to Quality Plan. These preventive actions are to anticipate any technical problems that frequently arise in a design. For example, establishing the interfaces between the various components and techniques followed during the design, will avoid many appearance of non-quality. The designer should validate the interfaces identified, otherwise control will be negative. For example, if it validates the interface between the structure and air conditioning, that means it has studied the issue.
The definition for the check actions, allows identifying the important stage of the design, which will detect non-conformities. The idea is to define and schedule the activities of self technical check and the internal technical check of the design.

Define the documents type -This helps to define and organize the support documents required, which will contribute to track the activity on a daily basis. One example is the calculation notes, meeting report, documents, lists for non-compliance, corrective actions, and so on.

Documents type	
Memorandums	Closure monitoring and non-conformities
Reports type	Closure monitoring and corrective action
Meeting reports	keywords

Table 5: Registration forms for supporting documents

2.2. Production activities. Design development

Design development is a production in the large sense. It is the realization of studies, architectural and technical, from the sketch to the plans. This includes all stages of validation by the client.

2.3. Supervision actions

Report time spent - Along with its production activities, all members of the team will report their work hours, spent during the various production activities, meetings or checks, with a suitable computer tool.

Monitoring process -The control is based on a set of indicators to monitoring the progress of the design, both in terms of costs and deadlines, which are usual, but especially in tracking non-conformities and their consequences. The records of hours from production activities, meetings or checks, are carried out on commercial software and are processed and compared with what was prevue in planning design.

2.4. The indicators network

The Quality Plan has six quality indicators; five are economic and one is an indicator of the non-conformities detected. We get indicators by phase, by technical specialties, and by the general process.

IP2.3 - Delays variation
IP2.4 – Design hours variation
IP2.5 – Design non-conformities / hours variation - Show influence of the design non-conformities, on the production hours variation
IP2.6 - Design costs variation
IP2.7 - Design non-conformities / costs variation - Show influence of design non-conformities, on the work cost variation
IP2.8 - Number of design non-conformities - Accounting for design non-conformities observed.

Each of these indicators takes a maximum score of **5** and a minimum score of **1**, obtained by a linear function according to the figure 2, score prepared on the basis of a value calculated as a percentage.

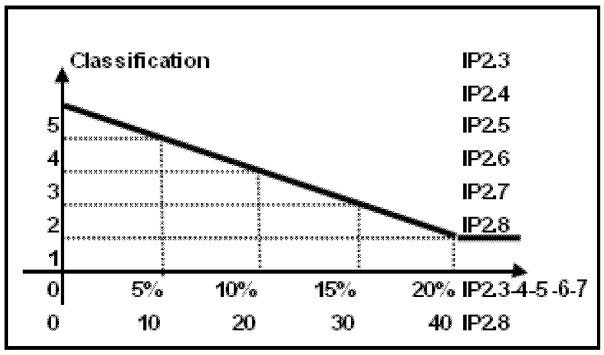


Figure 2: Chart for calculating indicators

Depending on the classification obtained for the indicators, profiles can be realized.
These are progress indicators for the design. The evaluation of the production activities, of technical specialty and persons, have fundamental role in setting up the process of continuous improvement.
These indicators are final indicators for the process P2. They informs in a logic of post-evaluation and contributes to the improvement of the design management practices.
The numerical analysis and overlay comparative profiles, makes it possible to assess quality during the progress of the design, by detecting the less efficient of those involved and technical specialties, which are problematic. It is possible to have a progressive analysis of the design and another final analyse. Figure 3 represents de design supervision structure

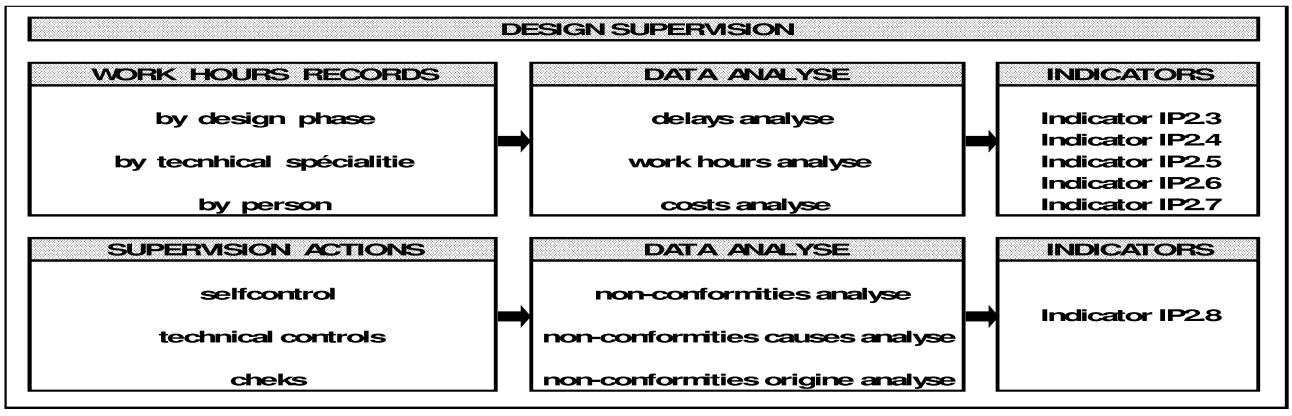


Figure 3: Design supervision

3. THE EXPERIMENTATION PLAN

The experimental approach was based on tree private buildings designs, in Portugal. Two of them, designated by PRO 1and PRO 2, are inserted in a urbanization, located in Espinho, a small town in the Porto region. The third one, designated by PRO 3, is the design of a building on an isolated lot, located in the town of Matosinhos, also in the Porto region,

The phases planned for the designs, were the outline design and detailed project brief, which is usual in Portugal for such private operations.
The experimental plan includes the studies for civil engineering -structures, water supply, drainage of wastewater and pluvial water, thermal comfort and acoustic comfort.
The design team is integrated on a Portuguese commercial company, which has 25 full time workers.

The team used a software tool called "Gesto," available on the Portuguese market that is adapted for design activities. It enables the recording and processing data - work hours, day by day, for everyone involved.
Quality Plans have been implemented in several stages.

a) 1 ª stage - PRO 1 Design – Espinho

The PRO 1, took place in an identical manner to that normally used by the team on other similar designs. It means that nothing will be changed in the established routines.
The design coordinator has calculated the number of hours of work and costs. During the design, he collected for each phase, the number of hours expended.
At the end of design implementation, technical checks were carried out with the analysis of non-conformities, their causes and their origins. The findings were communicated to the team, which developed then the final draft with the necessary corrections.
The comparison between "planned" and "realized" were made and then results a first set of data and indicators.

b) 2 ª stage - PRO 2 Design – Espinho

The PRO 2 design is similar to PRO1, from functionally and constructively points of view. The technical team was the same as the PRO 1.
The first activity of this stage, was to put in action some adjustments to the Quality Plan type used in PRO 1.The implementation of the plan was made step by step, in parallel to the development of the production actions, led by the design coordinator.
The tables and records were made gradually and at the end of the outline design and of the detailed project brief, it was then possible to implement two rounds of technical checks.

c) 3 ª stage - PRO 3 Design – Matosinhos

The activities planned for the PRO 3, were the same as those referenced to the PRO 2. But, in this case the site is different, the functional program has changed and some elements of the design team were different.
Adaptations on the Quality Plan were introduced, in logic of continuous improvement comparing to the previous stage (PRO2).
Thus, a third batch of data and indicators has been obtained, which led to further analysis and retrieval final conclusions on the proposed methodology.

d) Results Analysis

One final, the Process P2 indicators were analysed, witch results are shown on the Table 6 and Figure 4 below.
It is possible to observe an overall improvement in performance indicators, which is reflected in growing averages based on the good results. For example, for PRO 1, the average indicators has the value of 1.13 and for PRO 2 and PRO 3, the averages are respectively 2.84 and 2.09.
Generally, all the indicators show a positive development, but the results of the design PRO 3 are actually more consistent in comparison with those of PRO 2.

Indicators Table- process P2						
	IP 2.4	IP 2.5	IP 2.6	IP 2.7	IP 2.8	Average
PRO 1	1,00	1,00	1,00	1,66	1,00	1,13
PRO 2	1,66	3,56	1,97	3,74	3,27	2,84
PRÓ 3	2,87	2,93	3,41	3,21	2,50	2,98

Table 6: Table indicators

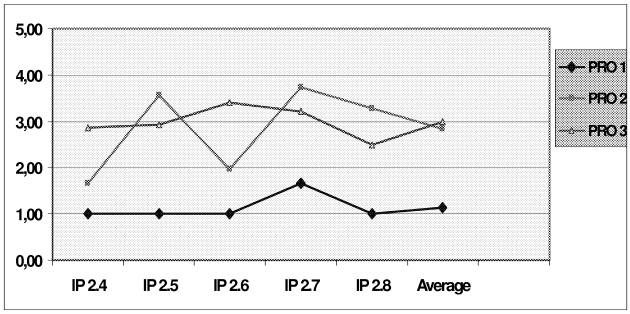


Figure 4: Indicators profile

In general, the non-conformities most frequently for the three designs are errors and omissions, including drawings or omitted details incomplete, incompatibilities between designs and incomplete specifications

e) Costs variation

Variation: (Real costs - Expected costs) / Expected costs- The evolution of this variation has been positive, from 45.19% (PRO1), to 15.22% (PRO 2) and 12.28% (PRO 3).

Relation between the design costs and design sale value- The results show that for the PRO 1, the design has been in deficit (-10.46%) and that for the other two designs, PRO 2 and PRO 3, it was possible to reverse the situation with obtaining positive results, +12.83% and +15.79% respectively. We can reasonably state to the beneficial impact of the introduction of quality plan

The cost of nonconformities -In the PRO1 design, the costs of non-conformities are quite high (17.23%) compared to budgeted costs, but we obtained better performance with the implementation of the quality plan, especially for the PRO 2 (6.55%), which can be explained by its similarity to the PRO1.

4. CONCLUSIONS

The design phase is shown in unanimous way, as a primarily responsible for the non-quality in buildings. The proposed Quality Plan is the establishment of rigorous design management, united with a follow-up not only in costs and

time, but also with non-conformities. The numerical analysis and overlay comparative profiles, makes it possible to assess the credit quality during the design progress, by detecting from those involved the less efficient and from technical specialties which are problematic. It is possible to have progressive and final analysis of the design.
In addition to this, are put in place preventive actions, especially monitoring of design implementation through indicators, which can detect and correct very early the potential errors.

The experimental plan based on three real designs, was difficult to set up, and show the real prospect of the problems and the complexity of the design of buildings.
It was also possible to establish progressively the Quality Plan for two purposes:

- Analyse the impact of the plan, on the design phase quality, with the monitoring of the indicators.
- Initiating the process of improving the Quality Plan on successive projects.

Quality Plans in place, have proved to be efficient, especially for all the prevention, planning, organization and check actions that were generate.
Note that the mere, existence of a Quality Plan creates a more attentive of the persons, who know they are under supervision and produced work of higher quality. It can be said, that there is a direct influence and another indirect.
The results obtained in terms of experimentation are not exceptional, but are sufficiently positive to partially validate the quality methodology proposed. The implementation of Quality Plans requires a great deal of work, to train and create knowledge among various persons, to adopt a more attentive attitude during the design phase, and similarly, the adoption of procedures is slow, but progressive.
Anyway, it is possible to state that the main objectives of the quality plan have been achieved:

- The detection of non-conformities was a reality and their number have decreased;
- Variation between real costs and expected costs, have followed a very positive development, ;
- The experimentation plan provides important feedback for future designs;
- During testing, awareness raising and training with persons have proved very effective.

5. REFERENCIES

Borges, Ferry.1990. "2º Encontro Nacional Qualidade na Construção" .LNEC Lisboa

Bureau Securitas.1979. «Annales de L' Institut Technique du Bâtiment et des Travaux Publics, n.º 378 .Paris.

Chartered Institute of Building. Project Management for Construction & Development. Ed. Longman 1996.

Crosby, P.B., Quality Free, McGraw – Hill, 1979

Carvalho, J. 2002. "A gestão da qualidade nos serviços dos empreendimentos ". Dissertação. ISEP. Porto

Carvalho, J. 2007. "Gestion de la qualité en phase de conception des opérations de construction de bâtiments. Thèse. Ph.D.Université de Savoie. France.

Direito Prtuguês .1972. "Portaria de 7 de Fevereiro de 1972"

Estingoy, Philippe et Rabatel, Michel – Montage et suivi d´une opération de construction. Le Moniteur. 2002

Juran, J.M - Quality Control Handbook - McGraw – Hill, 1979

Marrington, James – The Improvement Process – Mcgraw Hill, 1987

NF EN ISO 9000 (2000). Principes essentiels et vocabulaire ». AFNOR

NF EN ISO 9001 (2000). . Exigences. » AFNOR

Pires, A. Ramos – Qualidade - Sistemas de Gestão da Qualidade (colecção Silabo – 2ª edição)

Taguchi, G., Introduction to off line Quality Control – Angoya, 1979

Woodward, John F. Construction Project Management. Ed. Thomas Telford – 1997

Walker, Anthony. Project Management In Construction. Ed. Blackwell Science - 1996.

DEFINITION OF A LOGISTICS MODEL IN THE CONSTRUCTION INDUSTRY – CASE STUDY

Tiago Pinho
José Telhada
Maria Sameiro Carvalho
Systems and Production Department, School of Engineering
Gualtar, University of Minho, Braga, Portugal
E-mail: {tiago.pinho, telhada, sameiro}@dps.uminho.pt

KEYWORDS
Logistics, Supply Chain Management, Construction Industry, Information Systems, Information Technologies.

ABSTRACT

Throughout times, we have witnessed an improvement in the construction tools, materials and techniques along with the exponential growth of this kind of industry. On the other hand it is very difficult to find improvements in the information management and coordination tools for the construction projects. An integrated Information System that supports construction projects seems like an idealistic (impossible even!) idea. Nevertheless, it is the future path for the construction companies to become more competitive and efficient. An Information System that can provide real-time information on the current state of a construction project is the main focus of this paper. Many of the project managers receive the information, regarding their project, with a long delay, so that they can no longer give a quickly response to the necessary changes.

INTRODUCTION

In the past, logistics was seen as a function to simply deliver the goods to the costumer, at the right time and in the right place, at the optimal cost. That definition is no longer sufficient. Nowadays, issues like customer satisfaction, quality and real time information are now the main focus. The construction industry is not an exception, since it has to deal with this reality. In order to correspond to the market expectations, the companies have to invest in their internal supply chain optimisation.
In the construction sector, a large degree of uncertainty exists, normally at different levels. This is, inevitably, reflected in a much bigger complexity of management of the resources involved, in comparison with what happens in other logistic chains more centralized (especially those more oriented for the final products output and commercialization). To illustrate such a complexity one may imagine that a given company has to extract some of the raw materials and produce some of the components in somewhat centralised infrastructures; however, it has to distribute all these materials and components to a large number of construction sites that are widely dispersed across the country. The company also needs to continuously establish new contracts with construction-site nearby suppliers, and to coordinate their deliveries with its own projects. And, furthermore, it is well-known that construction projects are very vulnerable to schedule variations (delays) due to an enormous variety of factors and most of them are out of company´s control (unexpected works/difficulties on operations, bad weather conditions, etc).
This article is a result of project carried out to a Portuguese construction company, which consisted in the redefinition and implementation of its logistic model. As the company business was growing up, it became clear to managers that their traditional planning and management strategies could no longer be effective in the increasingly competitive market. To deal with the increasing complexity of its logistics system, the company decided to redesign sustainable organization solutions supported by large investments in Information Systems and Information Technologies (IS/IT). This article reports the main developments carried on in this area, namely the development of a web portal for supporting the supply chain. Preliminary results show that the global performance of the company supply chain has improved significantly.
The main objectives of the project were defined as:
- To elaborate a detailed analysis of the company´s logistic system, identifying the main gaps and constraints, aiming to define a conceptual model of its logistic network;
- To plan processes of re-engineering for the logistic activities, to develop appropriate methodologies and tools, aiming basically the compression of time, eliminating waste and improving the global performance of the company;
- To develop the integration of all the relevant information into an Information and Communication System. This will allow the development and implementation of an advanced decision-making support system that can promote significantly the coordination of the global logistic network of the company, including external suppliers, transporters and construction outsourcing companies.
A more strategic objective of the project, in a socio-economic view, is to contribute to the improvement of the efficiency and competitiveness of the construction industry.

SUPPLY CHAIN MANAGEMENT

The Supply Chain Management (SCM) in the construction can be seen as the network of installations/resources and activities that provides added value to the final customer, in the functions of project design, contact management, acquisition/provision of materials and services, production

and delivery of raw material and management of the installations/resources (Love et al. 2004).
The logistics movement of materials regarding the construction projects can be divided in 3 categories: *Supply Logistics*, *On-site Logistics* and *Reverse Logistics* (Figure 1).

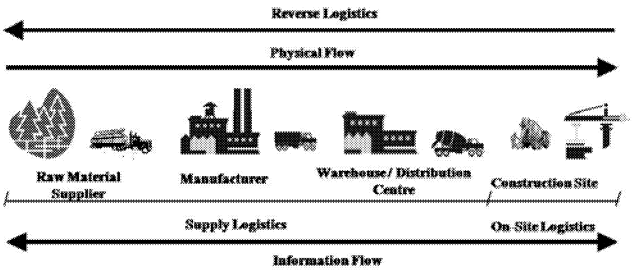


Figure 1: Logistics Flow

In the first case, its main focus is the supply of materials, services and workforce to the construction projects. The second refers to the internal logistics in the construction site, namely, stock management, movement and storage of materials. The third logistics category is orientated to the construction projects devolutions to the main warehouse, where it is necessary to transport and verify the integrity of the materials for future use.
To support the construction projects it is necessary to elaborate a bill of materials that needs to be delivered in the construction site according to a defined planning. The concept of "order/request" means the quantity of materials/equipments/ workforce needed to execute an activity. In order to accomplish the activities, the construction site managers have to request materials/equipments/workforce to the main logistics centre.
The order processing system set many logistics activities in motion, such as (Saad et al. 2002):
- Determining the transportation mode, carrier and loading sequence;
- Inventory assignment and preparation of picking and packing lists;
- Warehouse picking and packing;
- Updating the inventory file, subtracting actual products picked;
- Automatically printing replenishing lists;
- Preparing shipping documents (a bill of lading, if using a common carrier);
- Shipping the product to the costumer.

INFORMATION IN CONSTRUCTION

The implementation of Information Systems and Information Technologies permit to obtain higher levels of integration that can result in improved productivity, higher levels of co-operation and co-ordination and competitive advantages for business. However, construction enterprises are still facing great difficulties in implementing this process both internally and among partners as the result of the use of disperse software packages with no integration, no real time information available for decisions makers and partners, non utilization of bar coding for tracing products and little or none existing formal training, either in logistics and ITs. Additionally, the lack of appropriate communication and Information Systems prevents a deep change in the construction industry culture, a prerequisite to achieve successful supply chain integration.
Many reports describe the success of other industries in adopting extensive and effective supply chain integration, cooperation and organizational culture, introducing new business processes and achieving significant cost reductions, generating gains in both productivity and quality. Despite the difficulties, there is an emergence to the adoption of such managerial approaches in construction industry. Most can be done, despite that on-off projects, wide geographical and functional dispersion and cyclical demand for products, among many other obstacles, that may inhibit the achievement of the "ideal" level of integration in construction supply chains.
In the last decades, an increasing number of studies report some few advances in the coordination of information flows. For example, (Lambert et al. 1998) provided a review of the progress towards the adoption of SCM relationships in construction, concluding that the most common forms of collaboration have been always at the contractual level and at punctual times during the project life, although since mid 80's there has been a tendency to engage on high levels of collaboration involving design, building activities, and even partnering relationships. Also, (Briscoe and Dainty 2005) reported that integration in construction has been mainly observed as partnering arrangements and limited coordinated working among parties: the emphasis has been done on transactions rather than on relationship exchanges. More recently, another study (Tarantilis et al. 2008) identifies the importance of a Web system to help several business areas, namely, construction industry. This system was tested in more than 40 companies in Greece, in areas such as construction, education, production, insurance and finance.
Many Internet and traditionally-based electronic commerce (e-commerce) companies, whether focused on business-to-business (B2B) or business-to-consumer (B2C) markets, have come to realise that easy access to information and communication and the delivery of their products or services are important drivers in developing market competitiveness. To support electronic logistics (e-logistics) and reverse e-logistics it is necessary to maintain this competitiveness (Sarkis et al. 2004). The use of an Internet-based portal gives users the ability to access the information in the construction sites, not only inside the company intranet. Web-based systems for information management are intended to extract and filter information from a large amount of information in a database (Kopanaki et al. 2001). The main attributes of an Internet-based portal are: real-time information, the ability to build and unleash the power of relationships, and the effect it has on the economics of software (O´Keeffee 2001). Although the web portal is an effective outside and inside tool, it doesn´t work on its own, it needs a backend system for support. This can be a single database or a complex system to support the communications.

WEB PORTAL

After the analysis made of the supply chain, in the company and in the construction site, it was decided to implement several features in the web portal (Figure 2):

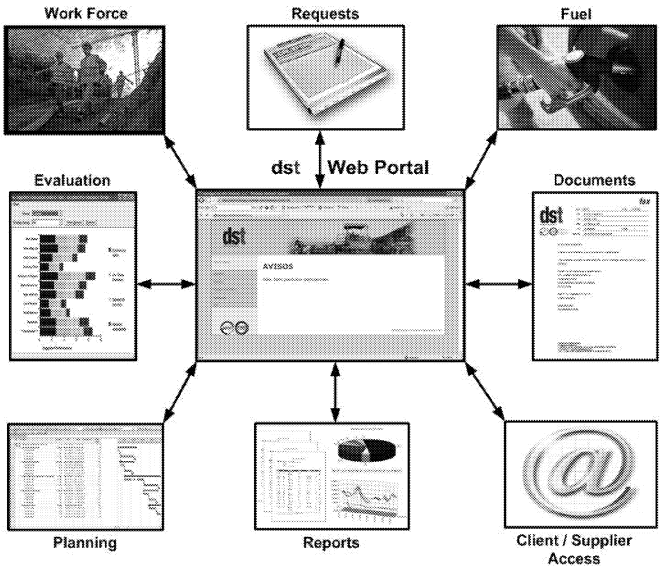


Figure 2: *dst* Web Portal Features

- Ability to launch requests of materials: the portal connects to the order processing system of the company, which immediately returns all relevant information about stocks and expedition details related to requests;
- Ability to launch requests of equipments: the user selects the equipment, based on the characteristics the user needs;
- Ability to launch requests of transportation equipments: the user selects the vehicle, based on the characteristics user needs;
- Ability to display the status of fuel consumption: the user can then manage the stock of the fuel stored on the tanks in the construction sites;
- Allows performing project management, with the association of the resources (materials, equipment and work force) to the activities executed on-site;
- Ability to display the status of requests;
- Ability to generate reports;
- Ability to validate documents and generate official documents for materials transportation;
- Ability to evaluate suppliers (and clients to evaluate the company);
- Possibility to consult the stock available on-site;
- Possibility to consult information about external suppliers, e.g., information about their products.

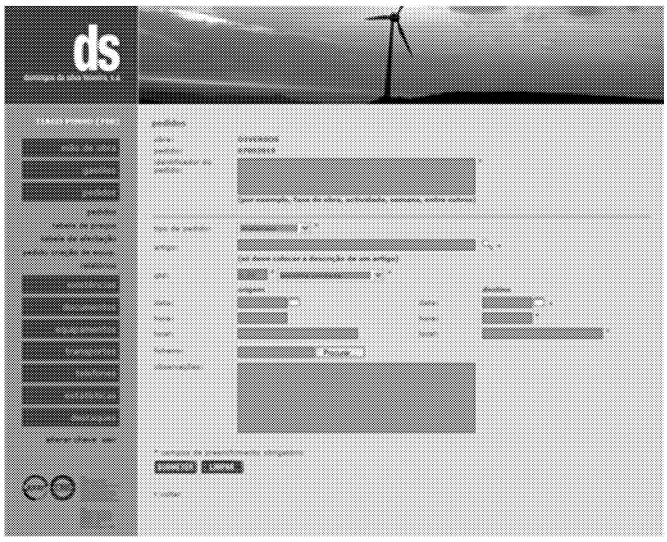
All of these features were implemented with only one purpose: the optimization of the communications and therefore the internal supply chain. For the web portal users, it is now very simple to analyse data, since they had alerts to the changes made in their information. Another advantage of the company's web portal is their portability, e.g. any user with an internet access can easy get all the necessary information in real time.

MAIN DEVELOPMENTS

Some of the features implemented were of the highest importance for the adequate control of the construction projects. Some have "only" economical importance, but others are more operational, like project management and planning. Here are descriptions of the more important features:

Workforce Management
- Engineers work times introduction (the company has an access and time control system using PDA devices (Pinho et al. 2007), but in case of system failure, the web portal can substitute the system for all the employees);
- Distribution of the engineers' total time cost per all the construction sites in which they worked during a period;
- Monitoring of the work days and absence.

Requests Management
- Materials, equipments and transportation requests /orders (Figure 3);


Figure 3: Requests Submission

- Possibility to consult the status of the request (pending, cancelled, ready to dispatch, finished, etc.) and see who is processing and when (for time response purpose);
- Possibility to cancel and claim the requests and their processing;
- Along with the complete record of the request, for future use, it is possible to analyze the material consumption of the construction project and see if they are according to the project;
- The transportation requests are associated with the truck fleet (it is possible to see the trucks last known position).

Fuel Management
- Equipments fuel supplying (owned, rented and sub-contractors equipments): this feature is also integrated with the accounting system in order to debit the sub-contractor fuel consumption;
- Fuel tanks stock management, including the report of robberies, reading errors and the input of more fuel distribution, which helps to maintain the control on the fuel tank in the construction site. All the movements made in the fuel tanks are reported in this feature in order to debit or credit the construction sites.

Documents Management
- Suppliers documents introduction (Figure 4): when a material from the supplier arrives, the users can introduce all the data relevant for that document, since there is no direct connection to the suppliers' database;
- Documents validation (for suppliers' payment effects): all the documents need to be validated before payment, so the user can validate the documents in this feature;
- Information on current account of the suppliers;

- Results reports: this is where all the accounting information is collected and organized.



Figure 4: Documents Introduction

Stocks Management

- Construction sites stock management, introducing the material consumptions (associated with the activities) and the material waste and material inventory correction;
- Main warehouse headquarters stocks available online: in real time you can see the consumption in the main warehouse and the remaining stock.

Fleet Management

- Fleet management information: all the trucks report periodically their location, speed, driver and state of driving;
- Associated with the truck, there is the information of the driver, with his contact number, so that the user can contact him if needed;
- Using online maps (Figure 5) provided by Google Maps (Google 2007), it is possible to see the location of the entire fleet all in one map or separated by types of vehicles.

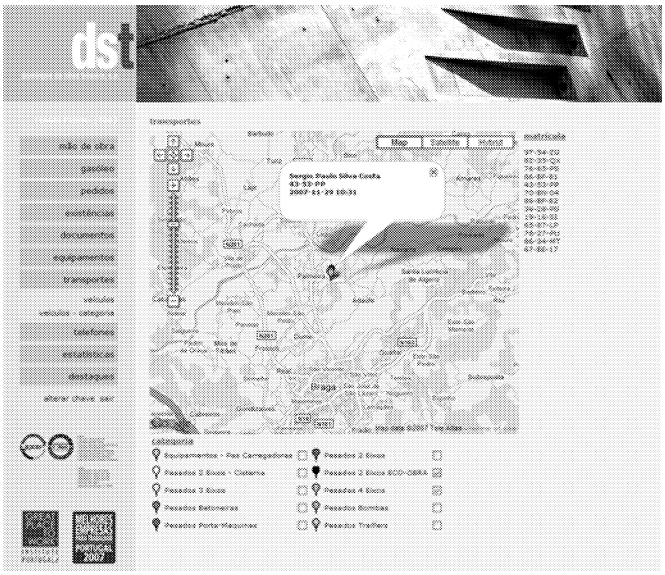


Figure 5: Fleet Management

- Associated with the requests from the construction site users, it is possible to see the location of a truck that is carrying a specific shipment. This integration with the requests is an important feature, since users can track closely the location of their cargo and coordinate the arrival of the truck at the construction site.

CONCLUSIONS AND FUTURE WORK

The use of the Internet in the business world is not a new idea, but the development of an e-logistics portal in construction is something that can be seen as an important innovation.

The main work in this project is devoted to carry out a reengineering process and creating new procedures/ proceedings for the internal logistic chain of the company in order to implement an efficient Information System capable of supporting the construction projects.

After the implementation of the new proceedings, a new form of communication can be developed. Therefore, the use of an Information System and, in particular, a web portal has proven to be essential to the improvement of the company´s global system. The web portal is already in use by the company´s engineers through which they can access real-time data, regarding their construction projects, essentially, the requests, the workforce control, the fuel control and the economic position of the construction project. On the other hand, through the web portal is also possible to access the information gathered by a fleet management system, using GPS technology, and a time clock/access control of the employees. As a result, it becomes possible to allocate, more efficiently, costs to construction sites and to monitor their activities.

Effectively, the information flows and the efficiency of communication across departments and construction sites have been greatly improved, allowing that most of the warehouse and distribution problems have been vanished or drastically reduced.

Other features can be developed in order to extend the current Information System, such as the implementation of an advanced work planning feature, which is the possibility for users to develop work plans, to assign resources to activities, track online progress, manage budgets and analyze workloads. All of these features will be dynamically updated once the users input more information.

Another project perspective that can also be developed is the integration with suppliers. For example, this Information System would have an interface through which suppliers can access to prices request for a specific product and leave their best offer, not only about prices but also about material availability and time delivery.

ACKNOWLEDGEMENTS

This work would not be accomplished without the help of the *dst* Information Systems Department, namely Rui Caridade, Paulo Rebelo, Eva Lima and Sónia Freitas. A special note to the *dst* administration, in the person of José Teixeira, that provided all the tools needed for the development of this project.

This project is also financed by the *FCT – Fundação para a Ciência e Tecnologia.*

Last, an acknowledgement to the staff of the *Systems and Production Department* of the *Minho University*, in Braga.

REFERENCES

Briscoe G. and A. Dainty. 2005. "Construction Supply Chain Integration: an Elusive Goal?" *Communications of the* Supply Chain Management: An International Journal, vol. 10(4), 319-326.

Google. 2007. "Google Maps API – Version 2," Access in August 2007 (http://code.google.com/apis/maps/)

Kopanaki E., Karkaletsis V. *et al*. 2001. "MITOS: An Integrated Web-based System for Information Management," In *8th Panhellenic Informatics Conference*, Greek Computer Society.

Lambert, D. M., Stock, J. R. and Ellram, L. M. 1998. *Fundamentals of Logistics Management*. McGraw Hill, Boston.

Love, P., Irani, Z., and Edwards, D. 2004. "A Seamless Supply Chain Management Model for Construction," *Communications of the* Supply Chain Management: An International Journal, vol. 9, 43-56.

Saad, M., Jones, M. and James P. 2002, "A Review of the Progress Towards the Adoption of Supply Chain Management (SCM) Relationships in Construction," *Communications of the* European Journal of Purchasing & Supply Management, vol. 8, 173-183.

Sarkis J., Meade L. M. and Talluri S. 2004. "E-Logistics and the Natural Environment," *Communications of the* Supply Chain Management: An International Journal, vol. 9, 303-312.

Tarantilis C., Kiranoudis C., Theodorakopoulos N. 2008. "A Web-Based ERP System for Business Services and Supply Chain Management- Application to Real-World Process Scheduling", *Communications of the* European Journal of Operational Research, vol. 187, 1310–1326.

O´Keeffee M. 2001. "Building Intellectual Capital in the Supply Chain - the Role of e-commerce," *Communications of the* Supply Chain Management: An International Journal, vol. 6, 148-151.

Pinho T., Telhada J. and Carvalho M. S. 2007. "Web Developments in the Construction Industry: A Web Portal (PDA Web Site)," In Proceedings IADIS International Applied Computing, Salamanca, 797- 800.

BIOGRAPHIES

TIAGO PINHO studied industrial engineering at the University of Minho, Portugal. Since 2005 he is PhD student in the same institution. He is mainly interested in logistics applied to construction, GPS and Information Systems.

JOSÉ TELHADA is Auxiliary Professor at the Industrial Engineering Department of the University of Minho, Portugal. He is mainly interested in Operational Research and Logistics issues applied to manufacturing industry and services, in general, and Public Transportation systems modeling and planning, in particular.

MARIA SAMEIRO CARVALHO is Associate Professor at the Industrial Engineering Department of the University of Minho, Portugal. She is mainly interested in Operational Research and Logistics.

AUTHOR LISTING

AUTHOR LISTING

Aguirre J. **79**

Barkaoui K. **95**

Calvete H.I. **103**
Campos A. **79**
Cannella S. **13/33**
Carvalho Brito A.E.S. ... **41**
Carvalho J. **128**
Carvalho S. **111**
Chibani N. **19**
Ciancimino E. **13/33**
Correia A. **79**
Cortez P. **5**

Diamantas V. **58**
Dias K.S. **116**
Dias L.S. **47**

Feliz-Teixeira J.M. **41**

Gale C. **103**
Geraldes C. **111**
Grama C. **79**

Hicheur A. **95**

Janssen M. **123**
Jemai A. **19**

Kirytopoulos K. **58**
Koblasa F. **47**
Kostanyan A. **89**
Kožusznik J. **55/71**

Leischnig S. **27**
Leopoulos V. **58**

Mangin J.-C. **128**
Martinovič J. **55**
Moalla M. **19**
Moreau R. **63**
Müller E. **27**

Neves Silva R. **79**

Oliveira J.A. **47/116**
Oliveros M.-J. **103**

Pavel V. **116**
Pereira G. **47/111/116**
Pinho T. **133**

Rutsch A. **74**

Sameiro Carvalho M. **133**
Sanchez-Valverde B. **103**
Sauce G. **128**
Schreurs J. **63**
Schumann C.-A. **74**
Silva A. **5**
Simanjuntak T. **123**
Smiri K. **19**
Stokic D. **79**
Štolfa S. **55**

Telhada J. **133**

Unucka J. **55**

Varosyan A. **89**
Verbraeck A. **123**
Vondrák I. **55/71**

Wang Y. **123**