13^{TH} EUROPEAN CONCURRENT ENGINEERING CONFERENCE 2006

ECEC'2006

EDITED BY

Uwe Baake

and

Enver Yucesan

APRIL 17-19, 2006

ATHENS, GREECE

A Publication of EUROSIS

Printed in Ghent, Belgium

13TH EUROPEAN CONCURRENT ENGINEERING CONFERENCE 2006

3RD FUTURE BUSINESS TECHNOLOGY CONFERENCE

FUBUTEC 2006

ATHENS, GREECE

APRIL 17-19, 2006

Hosted by

Athens Imperial Hotel

Organised and Sponsored by:

EUROSIS
GFI Consulting
DaimlerChrysler do Brasil
DaimlerChrysler AG
Smart GmbH
Avnet 3D
Ghent University

International Programme Committee

Günther Schuh
RWTH Aachen
Aachen, Germany

Renate Sitte
Griffith University
Gold Coast Mail Centre
Queensland, Australia

Vaclav Snasel
VSB-Technical Univ.of Ostrava
Ostrava-Poruba
Czech Republic

Paul Stratil
Smart GmbH
Böblingen, Germany

Rik Van Landeghem
Ghent University
Zwijnaarde, Belgium

Uwe v. Lukas
ZGDV e.V.
Rostock, Germany

Danny Van Welden
KBC WRB
Brussels, Belgium

Lode Vermeersch
KBC Asset Management
Brussels, Belgium

Valentin Vitanov
Cranfield University
Cranfield, United Kingdom

Vesna Bosilj Vuksic
University of Zagreb
Zagreb, Croatia

Ursula Wellen
Deutsche Post AG
Bonn, Germany

All the articles published in this Proceedings have been peer reviewed.

EUROSIS-ETI publications are ISI-Thomson and INSPEC referenced

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

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

EUROSIS Publication

ISBN: 90-77381-24-4

FUBUTEC
C
E
C
2006

PREFACE

This year in its 13th edition, the ECEC is held at the Athens Imperial Hotel, Athens Greece, together with the 3rd edition of the FUBUTEC conference. As convergence grows between concurrent engineering, future business technology and the needs of a cleaner environment, we are happy to see that this year's event covers all three areas, as you can ascertain from the different sessions:

1. Simulation in OR and Knowledge Management
2. Simulation in Business and Risk Analysis
3. Business Gaming
4. Discrete Event Simulation
5. Sustainable Development
6. Supporting Technologies
7. Formal Methods and Techniques
8. Engineering Process Management
9. Collaborative CE-Environments for Virtual Teams
10. CE-Enhanced Lean Manufacturing
11. Practical Applications and Experiences

We are also happy this year, that we can welcome Fanuel Dewever of IBM Business Consulting Services as our keynote for 2006. His talk will cover the Open and Collaborative Innovation in a European Network of Living Labs. We are sure this talk will fire your professional imagination.

We also would like to take this opportunity to thank the organizers of the conference, and especially Philippe Geril from EUROSIS-ETI in Ghent for his efforts and work, and to all Session Chairs, Members of the Technical Program Committee and Reviewers for their efforts to make this Conference a success.

We hope that this event will lead to further CE activities, which will also bring people from industry, research and education closely together. As chairmen of the 13th European Concurrent Engineering Conference and the 3rd Future Business Technology Conference, we thank you all for supporting this event and in wishing you a pleasant and rewarding stay in Athens.

Dr. Uwe Baake

Prof. Enver Yucesan

CONTENTS

CONTENTS

CE-ENHANCED LEAN MANUFACTURING

PRACTICAL APPLICATIONS AND EXPERIENCES

LATE PAPER

SCIENTIFIC PROGRAMME

SIMULATION IN OR AND KNOWLEDGE MANAGEMENT

A BUSINESS PROCESS MODELLING APPROACH FOR KNOWLEDGE MANAGEMENT

Vesna Bosilj-Vuksic
University of Zagreb, Faculty of Economics
Department of Business Computing
Trg J.F. Kennedya 6, 10000 Zagreb, Croatia
E-mail: vbosilj@efzg.hr

KEYWORDS
business process modelling (BPM), knowledge management (KM), simulation modelling, Croatian financial institutions, ARIS.

ABSTRACT
This paper presents an overview of business process modelling applications and analyses the relationship between business process modelling (BPM) and knowledge management (KM) projects, focusing on the suitability of the simulation-based approach in BPM projects. Based on results from a survey of Croatian leading financial companies, hypotheses were tested regarding business process modelling as a mechanism for knowledge capturing, externalisation, formalisation and re-use. Empirical findings suggest that organizational knowledge embedded in the business process repository could be used as the foundation for knowledge management system development. This approach is recognized, though still not implemented in Croatian business practice.

INTRODUCTION

Knowledge management (KM) enables the creation, communication, and application of knowledge of all kinds to achieve business goals (Tiwana, 2000). It is increasingly recognized as an integral part of an organization's strategy to improve business performance (Carillo *et al.*, 2003; Zack, 1999). Organizational knowledge integrates a company's experiences, company-specific knowledge, culture, communications and decision-making procedures, as well as the detail of business processes (Stewart, 1997). Business process modelling (BPM) is a methodology that enables enterprises to specify their business processes as a series of activities and transactions that together achieve a given business objective. As an approach, business process modelling focuses on understanding the underlying business processes and developing business process repositories where business rules are one of the most important elements for the detailed and formalised description of all facts (knowledge) to be implemented during information system (IS) development (Kovacic, 2004). Many different methods and techniques can be used to model and manage business processes. Since business change involves changes in people, processes and technology, and these changes happen over time, simulation appears to be a suitable process modelling method. Therefore, the focus of this paper is on process orientation and simulation modelling for the purpose of knowledge management. It compares the theoretical overview of BPM application in different areas and the results of empirical research, investigating how organizational knowledge embedded in the business process repository is utilised within leading Croatian financial institutions. The paper is structured in the following manner: a business process-based orientation for knowledge management is introduced and the overview of BPM applications is discussed, stressing the role of discrete event simulation in BPM projects. The results of BPM projects conducted in two Croatian leading banks and the insurance company are presented in the context of knowledge management. Finally, conclusions and future research orientations are provided.

BUSINESS PROCESS-BASED ORIENTATION FOR KNOWLEDGE MANAGEMENT

Organisational knowledge, as an important element of overall business knowledge, could be systemized, documented and retrieved in a business process repository developed using business process modelling tools within business process change projects. BPM could be considered an important contribution and approach to the process of KM. The role of BPM in KM is threefold: (1) business processes, if modelled and captured in the business process repository, are part of the codified intellectual capital of the organisation; (2) knowledge processes in an organisation should be a part of the business process repository; (3) the business process repository could be used for knowledge creation, sharing and distribution (Apshavalka and Grundspenkis, 2003; Woitsch and Karagiannis, 2002; Persson and Stirna, 2002).

A process is a knowledge asset that can create, store and disseminate best practices, procedures and standards. The business process repository contains existing process knowledge documented in the form of business rules: policies and procedures, job descriptions, business forms and application code, relational data-base management system rules (tables, constraints, and triggers). It could enable employees to reuse and adopt knowledge and best practices from previous business process restructuring efforts. Business rules support business policies that are formulated in response to an organization's mission, vision, objectives and goals. They are usually embedded in technology or documents, providing guidance to business processes (Burlton, 2003). By developing a business process and business rules repository, it becomes possible to classify existing knowledge and make it transparent, to identify knowledge carriers and knowledge users and to define them in specific roles. This structuring of knowledge or knowledge processes creates the conditions for configuring, administering and, where necessary, modifying a knowledge management system efficiently and effectively.

However, the approaches that focus on knowledge management within the business process level are limited (Papavassiliou and Mentazas, 2003; Papavassiliou *et al.*, 2002). Moreover, although business process modelling tools and/or workflow management systems (WFMS) adequately support the modelling and enactment of business processes, they still do not provide the required support for knowledge-related activities. As such, it becomes clear that an approach that explicitly integrates knowledge management activities into the business process environment is missing. The continued development of business process modelling tools, which will support the transformation of the integral model of business processes into the knowledge repository, will permit for stronger ties between these two otherwise separate areas.

OVERVIEW OF BUSINESS PROCESS MODELLING APPLICATIONS

Business process models can be used to serve a wide number of applications, for example to drive a strategic organizational analysis, to improve the existing processes, to derive requirements and specifications for information systems design, or to support the (semi)automated execution of processes or so called workflows (Paul *et al.*, 1999). The focus of this section is to discuss the application of BPM in different areas, depending on different project objectives and goals.

BPM in Business Process Change

Business Process Change (BPC) could be recognized as a form of continuous organizational change in which companies change and improve their business models, strategies and goals. To realize business process change, most companies use different methods and tools which integrate components for static and dynamic modelling and measuring the performance of the processes. Regardless of the methodology used, business process models play an important role in the different phases of BPC (Desel and Ervin, 2000). According to trends recognized from current business practice and the literature (Bal, 1998; Hommes and Reijswoud, 2000; Harmon, 2003; Aguilar-Saven, 2004), the typical features of an integrated BPM tool could be summarized as: data and organization modelling function, static process modelling, dynamic process modelling, data and process models interfaces, repository and publisher. Among these, the data and static process modelling features were most widely used in practice, whereas the dynamic process modelling feature was less frequently used. The aim of using an integrated BPM tool is to develop a framework that: (1) is easy for modellers and users to design and understand; (2) interrelates several business process modelling methods and techniques; (3) encourages standardization; (4) provides a single business process repository and the use of a common process vocabulary; (5) provides model analysis, validation and testing; (6) is able to analyse, simulate, tune and optimize the processes of a company, and (7) is formal enough to serve for software development purposes.

Application of BPM in Business Process Management and IS Development Projects

Business Process Management enables the design, analysis, simulation, optimization, automation and diagnosis of business processes by separating process logic from the applications that run them; managing relationships among process participants; integrating internal and external process resources and monitoring process performance (Burlton, 2003; Greasley and Barlow, 1998). This term is occasionally used to refer to various automation efforts such as Workflow Management (WFM) systems, XML business process languages and ERP solutions. However, this approach is too narrow and does not comprise the entire Business Process Management concept. The focus of traditional workflow management systems is on the automation of business processes. The main purpose is to avoid the programming, enabling the transformation of business process diagrams into tailor-made applications. As a result, there is little support for the analysis, design and diagnosis phase (van der Aalst *et al.*, 2003). The IS/WF modelling environment should have a formal foundation, providing a structured way of identifying and capturing all information, relationships and business rules that make up a business process (Kovacic *et al.*, 2003).

Many WFM systems do not support the simulation, verification and validation processes, or the collection and interpretation of real-time data. Therefore, Business Process Management extends the traditional WFM approach. However, a very serious problem for WFM and Business Process Management systems is the inability to translate business models into information (workflow) models precisely and without ambiguity as business process models still lack a formal foundation (Kovacic, 2004; van der Aalst *et al.*, 2003). Although software interfaces between process modelling and IS modelling are developed, these interfaces might provide some syntactical translation but cannot bridge the semantic gap between business processes and IS models. Here the manual revision of IS models is often more efficient and useful than the use of interfaces, but this problem is expected to be resolved by the developers of Business Process Management tools using the appropriate rule-transformation approach and introducing the rule repository.

INTEGRATING DISCRETE EVENT SIMULATION IN BPM PROJECTS

Simulation of business processes creates added value in understanding, analyzing and designing processes by introducing dynamic aspects. Computer based simulation models of business processes can help overcome the inherent complexities of studying and analyzing organizations and therefore contribute to a higher level of understanding and designing organizational structures (Giaglis *et al*. 1999). Kettinger *et al.* (1997) mention simulation as one of the modelling methods in their survey on business process modelling methods. However, it is only recently that dynamic modelling, in particular Discrete Event Simulation (DES), has been considered an essential component of process modelling and the term business-process simulation has been coined (Scholz-Reiter *et al.*, 1999). This section focuses on typical business motivations for the use of discrete event simulation (DES) in an attempt to analyze and define the suitability of DES to the range of BPM projects.

There are certain modelling requirements specific to simulation-assisted business process change modelling (Giaglis and Paul, 1996): (1) processes need to be formally modelled and documented; (2) modelling should take the stochastic nature of business processes into account, especially the way in which they are triggered by external factors; (3) there is a need to quantitatively evaluate the value of proposed alternatives; (4) the evaluation is highly dependent on the objectives of the particular study and (5) modelling tools should be easy to use to allow users of the processes to be involved in the modelling process. Greasley and Barlow (1998) also identified several areas in BPM projects where simulation modelling can be useful: identification of processes for change, identification of change possibilities, definition of process vision, understanding current processes and design and prototyping of new processes. Davies (1994), for instance, describes how some of these issues were addressed in a study of the use of simulation in a back-office process management system in the financial services sector. Giaglis and Paul (1996) suggest how simulation modelling could be used to assist in each stage of the five-step framework for the implementation of BPR proposed by Davenport (1993). Giaglis *et al*. (1998) reported a real-life case study where discrete-event models of business processes were developed to assist two companies in realizing the expected impact of EDI on key business performance indicators. A case study of modelling and automating business processes of a medium-sized bank introducing Internet technologies (intranet, workflow management system, Lotus Domino) was described in the paper by Nikolaidou *et al*. (2001). Business process modelling was conducted using the discrete-event simulation (Petri-Nets).

The modelling of business or "human-based" systems does, however, present a number of additional issues which may limit usage in this area. These include the difficulty of modelling situations where a large amount of discretion is available in how tasks are undertaken and the difficulty of using simulation for prediction when the method of carrying out tasks evolves over time. Popovic and Jaklic (2004) identified other issues concerning simulation modelling, including problem definition issues, issues regarding data collection, socio-political issues, hierarchical and modular modelling issues, granularity issues and integration issues and multi-perspective issues. Many process improvement methodologies also consider only single processes and allow improvements to that specific process without considering the effect of changes on other processes within the business. It may be that improvements to the studied process may have a detrimental impact on the business as a whole. To be absolutely sure that process improvements benefit the business, the entire business should be modelled and evaluated. However, it is very difficult (if not impossible), expensive and time consuming to simulate a model compounded of many interacting complex processes. Barber *et al*. (2003) therefore suggest developing a limited simulation model for chosen segments of business processes.

One of the major goals of Business Process Management is to realize continuous process improvement and business activity monitoring (BAM). Thus, BPM vendors are offering greater capabilities in this area. Almost all vendors offer at least some sort of administrative console with metrics and reporting capabilities. Other vendors specialize in process monitoring and offer enhanced analysis functionality. Through reports and analysis, companies can take steps towards process optimization. In terms of BAM, simulation can be useful at the process design and monitoring stages, though not in real time. Integrating simulation models with Business Process Management tools still represents a major problem. There are some solutions available for transformation and linking (interfaces) though problems still exist.

EMPIRICAL RESEARCH ON THE PROCESS-BASED APPROACH FOR KNOWLEDGE MANAGEMENT: CASE STUDY OF CROATIAN FINANCIAL INSTITUTIONS

In light of the above discussion and theoretical findings, the following hypotheses are proposed:
H1: *Business process modelling is implemented in different areas, depending on different project objectives and goals*.
H2: *Simulation modelling is recognized as a mean by which business processes may be analysed and evaluated, prior to implementing business process change*.
H3: *The business process model builds up a company-wide (organizational) knowledge base and could be the starting point for knowledge management system development*.

For the purpose of this study, a new project entitled "IT Aspects of Knowledge Management and Business Process Management Implementation in Banking" was launched in April 2004. The key objective of this project was to offer information on the status of KM and BPM practices in Croatian financial institutions. The project was realized by

IT researches of the Department of Business Computing (Faculty of Economics, University of Zagreb, Croatia).

Research methodology and findings

The study was conducted on a sample of 41 banks operating in Croatia in December 2003. The survey was based on a questionnaire resulting in 23 responses, representing a strong response rate of 56%. One of the questions posed in the survey was "Is the project of business process modelling/reengineering (BPM/BPR) ongoing or already completed in your bank?" A positive response was given by 10 banks (25% of the total number of surveyed banks). This percentage indicates that there is awareness of the need for business processes restructuring, as well as the existence of a relatively high percentage of banks which are still not paying sufficient attention to the improvement of business processes. However, the results showed that most BPM/BPR efforts have not focused much on knowledge (if at all) which is critical, considering that knowledge should be treated as the principal competitive factor. In order to obtain better insight into the actual standing, objectives, methods, tools and success of the business process modelling projects in Croatian financial institutions, the two leading Croatian banks (Privredna Banka Zagreb and Zagrebacka Banka) and the leading Croatian insurance company (Croatia Osiguranje d.d.) were selected as a representative sample. In the period June-September 2004, in-depth interviews were conducted with management staff dealing with this area. The questions posed to participants were grouped into three logical entities:

- general questions about the company (structure, size, financial strength of the company);
- questions on BPM projects in the company (start of project implementation, who initiated the project, project objectives, who implements the project, which methods and tools are used, project results, employee reactions to the project, does the project help in improving knowledge management, how results of BPM project are used or could be used in KM projects); and
- a short description of the project considered to be representative and/or successful and the quantitative and qualitative project results (which is outside of the scope of this paper).

Interview results were analysed and verified by the interview participants.

Research results: analysis and discussion

Comparison of the results showed the existence of common features of BPM projects in the selected banks and the insurance company:

- The primary objective for the companies was the development of an integral business process repository (the business process model covered about 80% of all processes in the companies). The business process repository is continually maintained, changed and supplemented. Business performance measurement methods are increasingly used: process models are enhanced with attributes necessary for analysis, measuring process performance, simulation modelling and creating proposals for improving processes. Companies use the developed repository as a foundation for the development of a new IS, as well as for changes to the existing IS, to the extent possible due to existing limitations (problems in the transformation of the process model into a data model). The business process repository is not available on the company intranet; however, this possibility is defined as a future objective.
- Companies do not approach changes as a one-time project, but instead treat them as a continuous process, one in which company employees participate as users who possess knowledge of the company operations, as well as in-house experts – employees specialized for organizational and business improvement tasks. The company position on the need for continuous implementation of BPM projects is evident in the decision to form special organizational units (directives, offices, divisions), with the task of improving operations.
- The possibilities offered by the business process repository as a basis for knowledge management system development are still insufficiently used. A knowledge perspective should be added and KM tasks should be assigned to roles during BP modelling. However, management has accepted the idea of the need to launch KM projects, in which one of the basic strategies will be the use of organizational knowledge built and structured into the process repository.
- The leading financial institutions in Croatia use the same tool (ARIS) for business process modelling. This fact is not surprising, considering that the consulting company Gartner ranked this tool as the best in its category. This tool is also the most represented on the Croatian market.

The use of tools for business process modelling and many years of implementing reengineering projects suggest advantages and positive impacts, as well as highlighting certain problems and shortcomings. The most significant advantages for the companies included in the study were:

- The development of the process repository was presented to managers and employees for the first time as a comprehensive, clear and detailed overview of all the key processes and their participants, thereby allowing for better understanding of the existing way of doing business, as well as insight into shortcomings, as well as the possibility for improvement.
- The process repository permits for the documentation and standardization of the process (procedure, routine, business rules) and with it the implementation of ISO standards. In this way, the quality of operations is improved, resulting in a positive impact on employee satisfaction and efficiency, as well as that of users and business partners.

- Detailed analysis and measurement of process performance, as well as simulation modelling, allows for the creation of alternative scenarios and proposals for improvement. With the implementation of the selected solutions and measurement of results of the conducted changes, significant positive impacts are visible. Qualitative impact are shown (greater efficiency, savings in time and human resources, shorter life cycle for products/services) while those of a qualitative nature are described (better quality products/services, greater satisfaction and trust on the part of the user).
- Information technology experts are extremely satisfied with the results of the business process modelling projects, providing them with an excellent basis for information system development and enabling faster and easier communication with the users, thus making the entire process considerably more efficient. They also expressed the need for development of more detailed process models including descriptions of business rules, procedures and steps, since such low-level models could be used to develop program-codes.

The following problems were noted in the implementation of the project and the use of BPM tools:

- BPM projects are relatively long-lived (1-3 years). The most time-demanding were those conducted first, and their objective was to develop a model of all the key processes in the company. After spending much time, human resources and financial resources, these projects resulted in a large number of models, the true value of which was noted and used only in later projects of a narrower scope, directed at analyses, performance measurements and improvements to individual processes, or their segments (sub-processes). This is a problem noticed in practice as the risk of over-analyzing existing business processes which led to the long period of modelling, producing huge documentation on "as-is" business processes and getting stuck in the business process analysis phase of the project. Therefore, the volume of business process models must be defined and strictly limited to the scope of the project.
- Simulation modelling is considered a very useful feature of BPM tools, but can only be used by experts. Most business managers and analysts could not use simulation due to a lack of knowledge and the tools' complexity. This problem could be avoided by improving awareness of simulation within the business community, by teaching simulation in business and management seminars and training courses in a systematic manner.
- Rule repository is the core of a development environment providing appropriate tools for process, workflow, data and organisation modelling, process refinement, as well as KM system development. Though the selected processes, or their segments, were analysed and measured in detail, the possibility of describing the process at the level of business rules was not fully taken advantage of, although this would create an appropriate foundation for the development of a WFM system and KM system. This opportunity was not implemented for two main reasons: (1) this approach requires additional time and resources, and (2) the existing software interfaces do not provide an automatic, reliable and secure transformation of generated business rules in the appropriate IS model and KM model.
- The results of these projects proved that the greatest drawback of ARIS, like other BPM tools, is the inability to connect and transform business process models into information system models. Although formally, the integration of business process models and information system models could be resolved through the use of software interfaces, experience from practice shows that there are still great limitations and serious problems. At the moment, the solutions provided by the software developers cannot be implemented in a rational and reliable way.
- The implemented projects were met with resistance from the employees, most frequently due to fear of change and a lack of understanding of the true project objectives. Despite this, the projects were successfully completed, thanks primarily to continued strong support from top management.

Following analysis and discussion on the study results, hypothesis **H1** could be accepted since the results of the research proved that BPM is applied in different areas, confirming that the best results are still achieved in business process change projects. It is also used as a basis for IS development, despite the current disadvantages. Hypothesis **H2** could be accepted since the research confirmed the assumption that simulation modelling is recognized and used as a very useful feature of BPM projects. The implementation of BPM for KM systems development is very limited, but usually declared as a project to be realized in the very near future, confirming the assumption stated in **H3**.

CONCLUSIONS

According to the resources from the literature and the results of the empirical research, it can be assumed that an approach that explicitly integrates knowledge management activities into the business process environment should be developed. This approach will be based on integration of business process rules with the organizational knowledge documented and stored in the business process repository. The continued development of BPM and KM software tools should enable the transformation of integral business processes model into the knowledge repository.

Based on the conducted empirical research, it can be concluded that Croatian financial institutions have taken a positive approach towards initiating BPM projects. The implementation of modern methods and tools for business process modelling and simulation creates a high quality foundation for improving operations. Since the management of these companies has accepted the need for a knowledge management strategy, this research could serve them to implement an integrated "business process and knowledge management centric" framework. Organizational knowledge, structured in the form of business rules, is already stored in business process repositories. The development and implementation of the knowledge management system should enable employees to search, retrieve, distribute and transfer organizational knowledge throughout the company. Simulation modelling as an integral part of BPM tools can be used to investigate knowledge management processes, to simulate missing data needed for knowledge management and to evaluate alternative models of knowledge management strategies.

REFERENCES

Aguilar-Saven, R.S. 2004. 'Business process modelling: Review and framework'. *International Journal of Production Economics*, No. 90, 129-149.

Apshavalka, D. and Grundspenkis, J. 2003. 'Making organisations to act more inteligently in the framework of organisational knowledge management system'. in *Scientific Proceedings of Riga Technical University on Computer Science*, Vol. 17, RTU, Riga, 72-82.

Bal, J. 1998. 'Process analysis tools for process improvement'. *The TQM Magazine*, Vol. 10, No. 5, 342-354.

Barber, K.D., Dewhurst, F.W., Burns, R.L.D.H. and Rogers, J.B.B. 2003. 'Business-process modelling and simulation for manufacturing management: A practical way forward'. *Business Process Management Journal*, Vol. 9, No. 4, 527-542.

Burlton, R.T. 2003. *'Business Process Management: Profiting form Processes'.* Sams Publishing.

Carrillo, P.M., Robinson, H.S., Anumba, C.J. and Al-Ghassani, A.M. 2003. '„'MPaKT: A Framework for Linking Knowledge Management to Business Performance'. *Electronic Journal of Knowledge Management*, Vol. 1, No. 1, 1-12.

Davenport, T.H. 1993. '*Process Innovation: Reengineering Work through Information Technology'*. Harvard Business School Press, Boston.

Davies, M.N. 1994. 'Back-Office Process Management in the Financial Service: A Simulation Approach Using a Model Generator'. *Journal of the Operational Research Society*, 45, 1363-1373.

Desel J. and Ervin T. 2000. 'Modeling, Simulation and Analysis of Business Processes'. in: *Business Process Management* (edsitors.: van der Aalst, W. et al.), Springer Verlag, Berlin, 129-141.

Giaglis, G.M. and Paul, R.J. 1996. 'It's Time to Engineer Reengineering: Investing the Potential of Simulation Modelling in Business Process Redesign'. in: *Business Process Modelling* (Eds. B. Sholz-Reiter and E. Stickel, Eds.), Springer-Verlag, Berlin, 313-332.

Giaglis, G.M., Paul, R.J. and Doukidis, G.I. 1998. 'Dynamic modelling to assess the business value of electronic-commerce'. In: *Proceedings of the 11th International Electronic Commerce Conference*, Vol. 1: Research, June, 57-73.

Giaglis, G.M., Paul, R.J. and Hlupic, V. 1999. 'Integrating simulation in organizational design studies'. *International Journal of Information Management*, No. 19, 219-236.

Greasley A. and S. Barlow. 1998. 'Using simulation modelling for BPR: resource allocation in a police custody process'. *International Journal of Operations & Production Management*, No. 18, 978-988.

Harmon, P. 2003. '*Business Process Change: A Manager's Guide to Improving, Redesigning and Automating Processes*', Morgan Kaufmann Publishers.

Hommes, B. and van Reijswoud, V. 2000. 'Assessing the Quality of Business Process Modelling Techniques'. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, Vol. 1, 1-10.

Kettinger W.J., Teng J.T.C. and Guha S. 1997. 'Business process change: a study of methodologies, techniques, and tools'. *MISQ Quarterly ,* March 55-80.

Kovacic, A. 2004. "Business renovation: business rules (still) the missing link". *Business Process Management Journal*, Vol. 10, No. 2, 2004, 158-170.

Kovacic, A., Groznik, A. and Krisper M. 2003. 'Business renovation: from business process modelling to information system modelling'. *International Journal of Simulation*, Vol. 2., No. 2, 41-50.

Nikolaidou, M., Anagnostopoulos, D., Tsalgatidou, A. 2001. 'Business process modelling and automation in the banking sector: A case study', *International Journal of Simulation Systems, Science & Technology*, Special Issue on: Business Process Modelling, Vol. 2, No. 2, 65-76.

Papavassiliou, G. and Mentazas, G. 2003. 'Knowledge modelling in weakly-structured business processes', *Journal of Knowledge Management*, Vol. 7, No. 2 2003, 18-33.

Papavassiliou, G., Mentazas, G. and Abecker, A. 2002. 'Integrating Knowledge Modelling in Business Process Management'. in *Proceedings of ECIS'2002*, June 6-8, Gdansk, Poland, 851-861.

Persson A. and Stirna J. 2002. 'Creating an organisational memory through integration of enterprise modelling, patterns and hypermedia: The hyperknowledge approach'. in *Information Systems Ddevelopment: Advances in Methodologies, Components and Management* (edited by M. Kirkova, J. Grundspenkis, W. Wojtkowski, W.G. Wojtkowsky, S. Wrycza and J. Zupancic), Kluwer Academic/Plenum Publishers, pp. 181-192.

Popovic, A. and Jaklic, J. 2004. 'Problematika simuliranja poslovnih procesov', In: *Management in informatika: Zbornik posvetovanja DSI 2004*, 166-172.

Paul, R.J., Giaglis, G.M. and Hlupic, V. 1999. 'Simulation of Business Processes'. *The American Behavioral Scientist*, 42 (1999), pp.1551-1576.

Scholz-Reiter, B., Stahlmann, H.D. and Nethe, A. 1999. ‚*Process Modelling*'. Springer-Verlag, Berlin.

Stewart, T.A. 1997. '*Intellectual Capital: The New Wealth of Organizations'*. Nicholas Brealey Publishing Limited.

Tiwana, A. 2000. '*Knowledge Management Toolkit, the: Practical Techniques for Building a Knowledge Management System*'. Pearson Education.

van der Aalst, W.M.P., ter Hofstede, H.M. and Weske, M. 2003. *'BPM 2003'*. Springer-Verlag Berlin Heidelberg, 1-12.

Woitsch R. and Karagiannis, D. 2002. 'Process-oriented knowledge management systems based on KM-services: The PROMOTE approach, in *The Practical Aspects of Knowledge Management, Proceedings of the 4th International Conference*. Vienna, Austria (edited by D. Karagiannis and U. Reimer), Springer, 398-412.

Zack, M. 1999. 'Developing a Knowledge Strategy'. *California Management Review*, Vol. 41, No. 3, 125-145.

An Ontology Oriented Approach for Knowledge Criticality Analysis

Charles-Emmanuel Foveau
Christophe Roche
Christophe Tricot
Equipe Condillac - LISTIC
Université de Savoie
Campus Scientifique 73376 Le Bourget du Lac cedex,
France
E-mail: {charles-emmanuel.foveau, christophe.roche, christophe.tricot}@univ-savoie.fr

KEYWORDS
Criticality, critical knowledge, ontology, knowledge management, knowledge cartography, human resources management.

ABSTRACT

Knowledge criticality analysis is an important economic stake for enterprise. As a matter of fact, managers and decision-makers must be able to identify as quickly as possible what knowledge are critical or could be critical in a close future in order to plan the enterprise's strategy in particular in human resources management.

This article presents a methodology provided with tools for knowledge criticality analysis. This methodology relies first on enterprise ontology building and then on knowledge cartography. The enterprise ontology represents knowledge about competencies, jobs and projects when knowledge cartography allows to visualize the critical information mapped on the ontology.

To qualify a knowledge as critical, it is necessary to define criticality criteria. We shall present some examples mainly based on the central notion of competency.

The result is a set of maps of critical knowledge which provides useful simulation tools. These tools have been integrated in the Os-Skill software environment, an industrial application dedicated to knowledge and human resources management.

INTRODUCTION

In an economical environment which is more and more competitive, the wealth of enterprise relies more on its know-how and knowledge than on its production means. To be able to have a global view and understanding of knowledge as well as to be able to identify critical or noticeable knowledge are crucial and strategic issues.

Knowledge criticality analysis needs first to represent and structure knowledge, then to analyse it according to criticality criteria which have to be defined.

At last but not least, it is necessary to provide decision-makers with indicators like graphical tools (e.g. interactive maps) in order to have a global view and to visualize the critical knowledge.

KNOWLEDGE

Knowledge Management

The rapid changes resulting from the globalization of markets, the economy deregulation and the impact of new information and communication technologies (Grundstein and al. 2002) have obliged enterprises to change their way of managing their knowledge (Zacklad and Grundstein 2001).

Knowledge management raises several issues: project management, return on experience management or competitive innovation. It is the reason why some large companies have invested in knowledge management systems like MKSM (Method for Knowledge System Management) (Ermine and al. 1996), KADS (Knowledge and Analysis Design Support) and CommonKADS (Breuker and al. 1994) (Schreiber and al. 1999) or CYGMA (Dieng and al. 2001). In spite of powerful functionalities, such systems are quite difficult to deploy and not very easy to use. Furthermore they do not provide functionalities for knowledge criticality analysis or knowledge cartography.

Knowledge Criticality

There are different kinds of knowledge as there are various sources of knowledge in enterprise. As a matter of fact, collaborators make use of and create different kinds of knowledge and know-how for their activities. In a same way, projects and jobs are defined in terms of required knowledge and competencies. This knowledge can be classified according to three main categories: explicit knowledge – the specific know-how which characterizes the company's capability to design, to produce, to sell and to support its products and services – and the individual and collective skills which characterize the enterprise capacity to act and to evolve (Grundstein and Rosenthal-Sabroux 2003).

Each part of this knowledge has not the same degree of significance or criticality for the enterprise in the sense that the lost of some part could be harmful for the company (Aubertin and al. 2003). It the reason why identifying critical knowledge is a central issue for enterprise.

According to (Ermine 2003), knowledge criticality analysis relies on knowledge cartography which can be drawn up according to three points of view: processes, domains and organisation.

Concepts

Whatever the point of view, the main concepts on which knowledge cartography is based are project, competency, job and collaborator.

The competency, which can be simply defined as a knowledge in action in a given context (Roche and al. 2005), is the central concept of our approach. As a matter of fact, everything is defined in terms of competencies: a job or a project mobilizes competencies which are held by collaborators and trainings are set up in order to acquire or improve competencies.

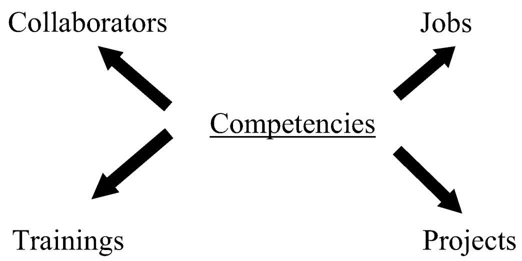


Figure 1. The competency-oriented approach of enterprise

In a such competency-oriented model of enterprise, and in the framework of this article, the knowledge criticality will be mainly defined in terms of critical competencies, and then in terms of critical projects and jobs.

ENTREPRISE ONTOLOGY

The knowledge criticality analysis requires to start by representing the knowledge of the enterprise, and more precisely by representing the main concepts of competency, jobs, project (or processes). It means by representing the enterprise ontology.

Ontology, defined as a common vocabulary of terms and meaning shared by a community of practice, is a central issue in knowledge management. It finds applicability in many application domains including knowledge based systems, natural-language processing, database design, datawarehousing, computer supported collaborative working, intelligent information retrieval, web semantic and from a more general point of view, in information systems and enterprise integration. The reason for why ontology is so popular is mainly due to what they promise: "a shared and common understanding of a domain that can be communicated between people and computers" (Duineveld and al. 2000).

Although one of the main objectives of ontology is to normalize the meaning of terms, there is neither really a consensus about its definition itself: "*although* ontology *is currently a fashionable term, no agreement exists on the exact meaning of the term"* (Gruber 1995), nor about building methodology: "*Ontologies are becoming increasingly popular in practice, but a principled methodology for building them is still lacking.*" (Guarino and al. 2000).

Following the "famous" definition of Gruber « *An ontology is an explicit specification of a conceptualization.* » (Gruber 1992), we shall consider that: "*An ontology is a shared conceptualisation of a domain, i.e. the description of the concepts, relationships and properties of the domain. Defined for a given objective, an ontology expresses a point of view shared by a community of practice. An ontology is represented in a language (explicit ontology) whose theory (semantics) guarantees the ontology properties of consensus, coherence, sharing and reuse"*.

In the framework of our application – human resources and knowledge management application – the knowledge criticality analysis relies on the representation of enterprise competencies, jobs and projects. So, we have defined a different ontology for each of them.

Today there are several software environments for ontology building (Duineveld and al. 2000), (Roche 2003). We have chosen two of them: Protégé 3.1 because it is one of the most famous environments and OCW (Ontology Craft Workbench) for its methodology.

Protégé 3.1 is a free open source ontology editor (Protégé 2005). The Protégé-OWL editor enables users to build ontologies in the W3C's Web Ontology Language (OWL) and supports the creation, visualization, and manipulation of ontologies in different representation formats. An OWL ontology includes descriptions of classes, relationships, properties and instances. As OWL is a logical-oriented language providing inference mechanisms, it is possible to derive new knowledge from an OWL ontology.

The figure below illustrates a part of the financial job ontology.

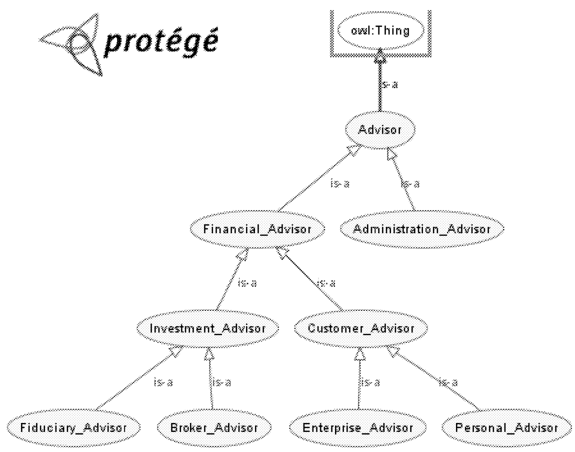


Figure 2 An ontology defined with Protégé

Although the Protégé definition of an ontology fits very well to our objectives ("*An ontology describes the concepts and relationships that are important in a particular domain, providing a vocabulary for that domain as well as a* *computerized specification of the meaning of terms used in the vocabulary.*" (Protégé 2005)), there is no methodology for building ontology and there is no means for managing the terminology aspect of the domain.

As a matter of fact, it is very important to help the experts of the domain in finding the suitable technical terms and then to define them. It is also important to have ontology-oriented tools for knowledge management (classification, information retrieval, etc.).

It is the reason why we have chosen the Ousia© methodology of the Ontologos corp. company (Ontologos corp 2005). The Ousia approach guides the experts in the terminology and ontology definition from texts to concepts.

The Ousia© methodology relies on four linked tools. The first one, LCW (for Linguistics Craft Workbench), allows to extract terms from texts (it means the specialized terms used by people to express enterprise knowledge). LCW is based on automatic analysis linguistic tools.

The second one, SNCW (for Semantic Network Craft Workbench), is a frame-oriented environment for building semantic network. It is the first step for defining in a semi-formal formalism the concepts and their relationships. SNCW takes the extracted terms from LCW as so many concept names.

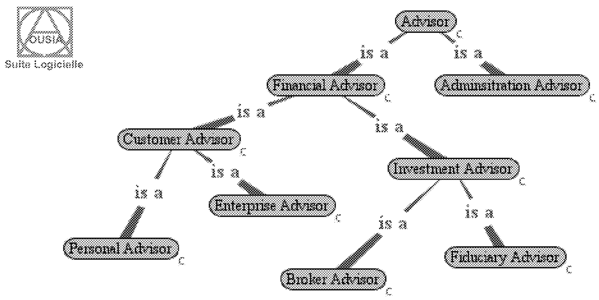


Figure 3 A semantic network defined with SNCW

Although the result looks like very similar to the previously defined Protégé ontology, a SNCW representation is only a semantic network without logical formalization – a logical approach is necessary if we want a sound and coherent representation -. But in the both cases there is no really definition of what is a concept and of what are the differences between concepts. We need a more epistemology-oriented environment for ontology acquisition.

The next tool, OCW (for Ontology Craft Workbench), is a software environment dedicated to building formal ontologies: creation, manipulation, visualization, etc. OCW relies on a methodology for building ontologies based on the definition of concept by specific differentiation: the most important is not the "is a" relationship, but what is the difference between concepts (the "is a" relationship is a consequence of the definition by differentiation) (Roche 01). If the result is a very formal definition of concepts, the most important feature of OCW is its methodology for ontology building based on linguistic and epistemological principles.

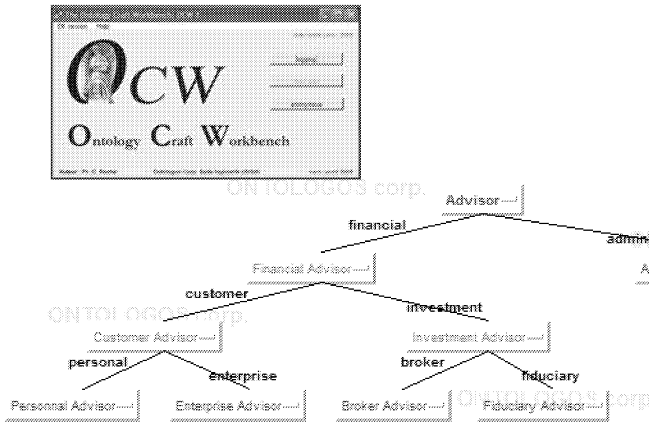


Figure 4 An ontology defined with OCW

The last environment is TCW, for Terminology Craft Workbench, which enables experts to link the enterprise lexicons to the formal ontology. TCW handles semantic relationships like synonymous and multilinguism as well as lexicons in different languages.

We have defined the enterprise vocabulary in TCW including the lexicons of the different practice communities the words of which are linked by linguistic relationships to the normalized vocabulary and whose semantics relies on the OCW ontologies of competencies, jobs and projects.

Let us notice that the Ousia software provides several additional tools for ontology-oriented semantic annotation and information retrieval based on OCW ontology and TCW terminology. Theses tools will be used for knowledge cartography.

KNOWLEDGE CARTOGRAPHY

Our approach of knowledge criticality analysis relies on knowledge cartography. As a matter of fact, knowledge cartography is a means for getting a global view and understanding of a knowledge base and quickly identifying critical knowledge.

The last section of this article will present the criticality criteria we have selected for our industrial applications in the field of human resources and knowledge management.

The enterprise knowledge analysis according to the three axles of competencies, jobs and projects, allows to build three different maps. The building of a map on a given axle corresponds to the projection (mapping) of all information and knowledge to the associated ontology.

So, the problems to be solved are on the first hand to annotate information with the relevant concepts and on the second one to map information on the ontology and to visualize the result.

Semantic annotation

In our application, knowledge cartography is first a classification of the enterprise knowledge on the concepts of the three ontologies (viewed as virtual directories). Our approach is an ontology-oriented semantic annotation which associates to every document the relevant concepts, i.e. the concepts the document refers to (Handschuh and al. 2003), (Kiryakov and al. 2005).

The following figure presents the ontology indexing process.

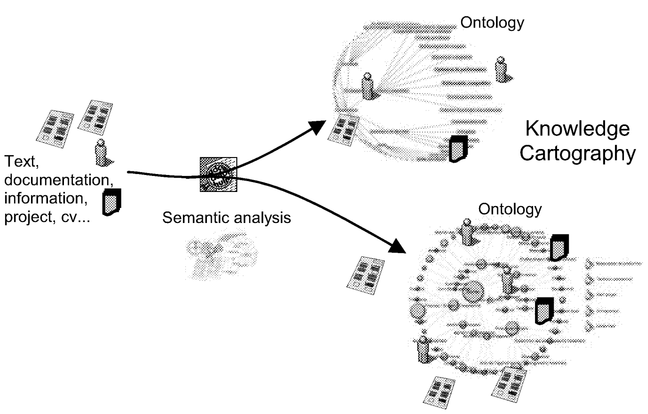


Figure 5 The ontology indexing process

Every document, whatever it is (CV, project, individual professional description, return on experience, job description, etc.) is analysed using linguistic tools based on the TCW terminology. Based on the occurrences of the words of usage which both appears in the document and in the terminology, a semantic marker is associated to the document (a semantic marker contains the concepts referenced by the document and their frequency). Such a semantic marker allows to map the document to the different ontologies in order to obtain the enterprise knowledge's cartography. Let us notice that a same document can be classified on several concepts and several ontologies.

Let us notice that the ontologies are extra linguistic. It means that they are independent of the language used for writing documents. The separation, inside the terminology, between words of practice used for writing texts and the concepts of the ontology on the other hand, allows to index on the same ontology documents whatever their writing language. Then it is possible to query in a given language and to find documents written in another language.

Cartography

Cartography plays a central role in knowledge management (Card and al. 1999). In our case, it allows to visualize the knowledge distribution on the ontologies of competencies, jobs and projects. The maps we have built rely on the ontology structure and on the hierarchical relationship between concepts.

Using graphical forms (the representation of nodes in terms of graphic variables, e.g. form, size, colours, etc.) allows to visualize important information such as the number of projects using a same competency (size), if a competency is critical because it is held only by few people (color), and so on.

At last, maps must be interactive. They must allow users to navigate inside the maps and the annotated knowledge base. Users must be able to find their way inside their information space without getting lost as well as to identify at all the time where they are and where they need to go.

Yellow pages

As we are going to see it in the next section, the competency, job and project maps allow the knowledge criticality analysis. For a more general point of view, they provide an interactive access and management of the enterprise knowledge.

The figure below illustrates one of the possible applications, in this case, the job map used as a job directory (yellow pages). The job ontology is represented as a tree projected onto a sphere (the "eye tree" paradigm (Tricot and al. 2005)) with the collaborator cards associated to the relevant nodes.

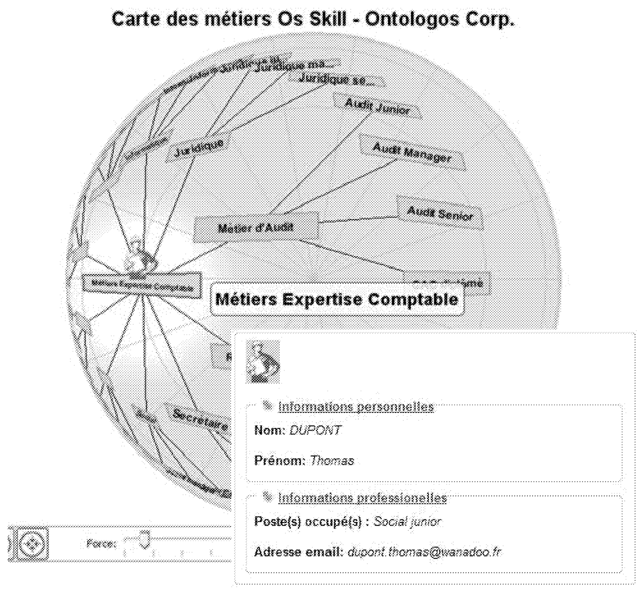


Figure 6 Os-Skill: the jobs' map

Semantic and interactive maps are very useful decision making tools. As a matter of fact, it is one thing to classify information, but quite another thing to find them or navigate in an ontology-annotated knowledge base (Tricot and Roche 2005).

CRITICALITY ANALYSIS

Our approach of knowledge criticality analysis relies on knowledge cartography regarded as a means to evaluate and locate critical knowledge. With semantic maps and the definition of criticality criteria, users can easily detect critical knowledge (highlighted through the different graphical forms in the maps).

Criticality criteria

There are different ways to analyze the enterprise knowledge, according to different points of view: organization, human resources, projects, competencies, etc.

As we previously saw it, the enterprise knowledge is represented through three different ontologies: the ontologies of competencies, of jobs and of projects.

Our approach is competency-oriented. It means that everything is defined around competency: collaborators are defined through their competencies and are evaluated according to them, jobs mobilize competencies as well as the realization of projects, and the enterprise know-how is assets of competencies. Competency is then a central issue in enterprise modeling, it also makes it possible to conciliate the different enterprise, collaborator and market points of view.

Therefore it is important to know when a competency is critical, becoming critical or could be critical. From this information, we can deduce what jobs and projects are also critical.

The criticality of a competency can be defined in different ways.

It can be explicitly defined according to a scale defined by the enterprise. For example:

- when the enterprise detects new needs in competency within a context of innovation;
- according to its rate of use;
- or how difficult to acquire the competency is;
- etc.

It can be also automatically evaluated according to different criteria:

- if a competency is used in many projects but held by few collaborators, it could be critical;
- if the competency is held by "sensitive" collaborators (e.g. collaborators who are close to retire from business or at the end of a fixed duration contract).

At last, it is important to notice that a competency is not a static resource. It lives and evolves within the organisation's life, for example in the field of law competencies or enterprise security competencies. Norms and technologies evolve regularly and organizations must respect them. Collaborators have competencies which have a life time. This life time depends on how fast they evolve. Therefore, a competency at the end of its life time is critical.

Map of critical competencies

The map of critical competencies (figure 7) visualizes the criticality of competencies. The use of different graphical forms (size, colour) allows to focus on the most important points.

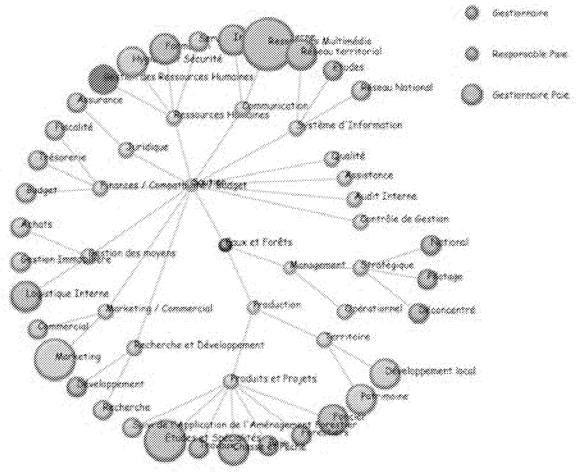


Figure 7 the critical technical competency map (XML&SVG)

In this way, the user can know which competencies could be problematic for the enterprise. By analysing this map, he will be able to know, for example, if some critical competencies are close to disappear or not enough held by collaborators (the size of circles is directly linked to the number of collaborators who hold the competencies or to the number of projects which mobilize these competences). Let us notice that knowing if competencies are going to disappear or to be less essential is important to set up in-house or continuing trainings).

Map of critical jobs

Since jobs are defined in terms of competencies, it is possible to create, in a same manner, a map of critical jobs if we consider that a job is critical if it mobilizes critical competencies.

But a job can be also qualified as critical if it is held by "sensitive" collaborators (e.g. close to retire from business or under a temporary contract).

The map of critical jobs gives another vision of enterprise knowledge criticality. Then it is possible to plan recruitments without using the competency cartography.

Map of critical of projects

Like jobs and competencies, projects can be mapped too. That can bring up some information about them. The user can determine if a project is critical, if some competencies it mobilizes are held by a small number of collaborators or not, and so on. The projects criticality will depend on their importance in the organization and on the resources needed for their accomplishment. It can help the user to control recruitment.

APPLICATION: Os-Skill

This ontology-oriented knowledge criticality analysis has been implemented in an industrial application called Os-Skill. Os-Skill is an intranet solution for knowledge and human resources management which provides different functionalities (figure 8). For example, each collaborator has an access to the portal, can see his competencies and is able to make a career plan. The maps of critical knowledge will be exploited by decision-makers to have an overview and a global understanding of the enterprise knowledge and know-how, whereas human resources managers will exploit them to plan recruitments and trainings.

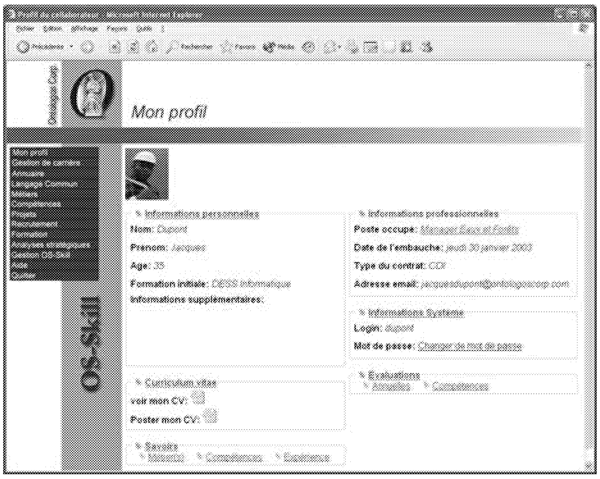


Figure 8 Os-Skill: collaborator's profile page

Simulation

Knowledge criticality analysis is not a static view of the enterprise. As a matter of fact, it is crucial to offer to decision-makers and human resources managers a long term vision of the company according to the three axles of analysis of competencies, jobs and projects. This is the simulation aspect of the knowledge cartography. Such a functionality is a useful means to help managers to identify the knowledge which could be critical not at the present time but in few years. Knowledge cartography simulation exploits criteria like competency life time, collaborator ages, etc. and the number of years for the forecasts. In this way, managers can anticipate, if nothing changes in the company, which knowledge should be a problem. Therefore, they will be able to plan what they should do in the next years to prevent those problems (figure 9).

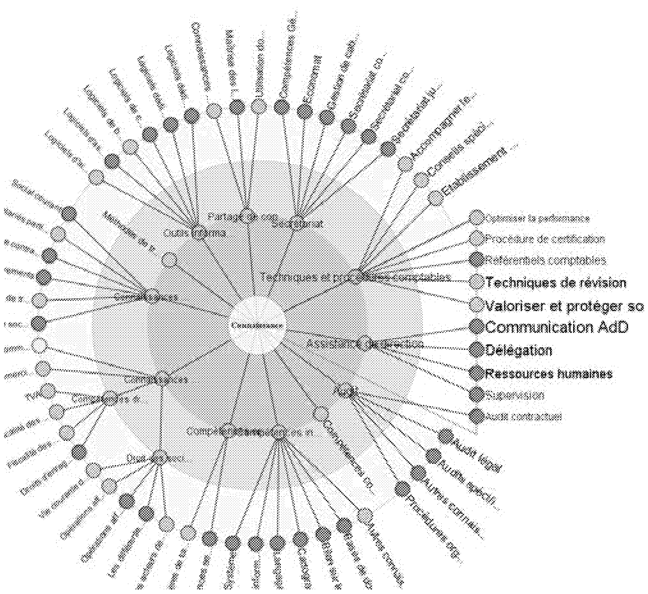


Figure 9 a simulation map (XML&Java)

CONCLUSION

Knowledge criticality analysis is a central issue for company. As a matter of fact, in an economic context which is more and more difficult and competitive, knowledge management is the new economic stake for enterprise. Enterprise must be able to have a global view and understanding of their knowledge and know-how as well as to be able to identify as quickly as possible what knowledge are, or could be in a close future, critical.

Our approach of knowledge criticality analysis relies on the knowledge cartography of the enterprise. The enterprise knowledge is modelling according to three ontologies: the ontology of competencies, of jobs and of projects. Then it is possible to index all information on these ontologies and to visualize the classification as so many semantic maps.

In order to qualify a knowledge as critical we have introduced different criticality criteria mainly based on the notion of competency. For example a competency is said critical if it is held by few collaborators or if it is mobilized by a lot of projects, etc. Managers and decision-makers can then visualize three different maps of critical knowledge for competencies, jobs and projects.

At last but not least, the maps of critical knowledge can be used for simulation in order to identify the knowledge which could be critical in a close future.

REFERENCES

Aubertin, G. ; Boughzala, I.. ; and Ermine, J-L. 2003. "La cartographie des connaissances critiques". Revue des sciences et technologies de l'information, série RIA-ECA - Volume 17 - n°1-2-3/2003, pages 495 - 502, Edition Lavoisier

Breuker, J. ; Van De Velde, W. 1994. "CommonKADS Library for Expertise Modelling, Reusable Problem Solving Components". IOS Press Amsterdam.

Card, S. K. ; Mackinlay, J. D. ; and Shneiderman B. 1999. "Readings in information visualization : using vision to think". San Francisco, Calif., Morgan Kaufmann Publishers.

Dieng-Kuntz, R. and al. 2001. " Méthodes et outils pour la gestion des connaissances, une approche pluridisciplinaire du knowledge management ". Dunod informatique. 2ème édition.

Duineveld A.J ; Stoter R. ; Weiden M.R. ; Kenepa B. ; and Benjamins V.R. 2000. "WonderTools? A comparative study of ontological engineering tools", International Journal of Human-Computer Studies, Volume 52, Number 6, June 2000, pp. 1111-1133(23).

Ermine, J-L ; Chaillot, M. ; Bigeon, P. ; Charreton, B. ; and Malavieille, D. 1996. "MKSM, a method for knowledge management". In J. F. Schreinemakers ed, *Knowledge Management: Organization, Competence and Methodology, Proc. of ISMICK'96*, Rotterdam, the Netherlands, Wurzburg:Ergon, Advances in Knowledge Management, vol. 1, October 21-22, p. 288 - 302.

Ermine, J-L. 2003. “ La gestion des connaissances ”. Hermès. Lavoisier.

Gruber, T.R. 1992. “A Translation Approach to Portable Ontology Specifications”. Knowledge Systems Laboratory September 1992- Technical Report KSL 92-71 Revised April 1993. Appeared in Knowledge Acquisition, 5(2):199-220, 1993

Gruber, T. 1995. “Towards principles for the design of ontologies used for knowledge sharing”. International Journal of Human Computer Studies, 43: 907-928.

Grundstein, M. ; Rosentham-Sabroux, C. ; and Pachulski A. 2002. “Reinforcing Decision Aid by Capitilizing on Company’s knowledge :Future Prospects”. EJOR 2002

Grundstein, M. and Rosenthal-Sabroux, C. 2003. “ A Way to Highlight Crucial Knowledge for Extended Company’s Employees ”. Annales du Lamsade, pages 173-192.

Guarino N. and Welty C. 2000. “A Formal Ontology of Properties”. Proceedings of 12th Int. Conf. on Knowledge Engineering and Knowledge Management. Lecture Notes on Computer Science, Springer Verlag, 2000.

Handschuh, S. and Staab S. 2003. “Annotation for the Semantic Web”, Volume 96 Frontiers in Artificial Intelligence and Applications.

Kiryakov, A. ; Popov, B. ; Terziev, I. ; Manov, D. ; and Ognyanoff , D. “Semantic Annotation, Indexing, and Retrieval”.Elsevier's Journal of Web Sematics, Vol. 2, Issue (1), 2005.

Roche, C. 2001.“The ‘Specific-Difference’ Principle: a Methodology for Building Consensual and Coherent Ontologies”. IC-AI’2001 2000 : Las Vegas, USA, June 25-28 2001.

Roche, C. 2003. “Ontology: a Survey”. 8th Symposium on Automated Systems Based on Human Skill and Knowledge, IFAC, September 22-24 2003, Göteborg, Sweden.

Roche, C. ; Foveau, C. ; and Reguigui S. 2005. “ La démarche ontologique pour la gestion des compétences et des connaissances”. EGC 2005, Volume 2, pages 359-364.

Schreiber, G. ; Akkermans, H. ; Anjewierden, A. ; de Hoog , R. ; Shadbolt, N. ; VandeVelde, W. ; Wielinga, B. 1999. “Knowledge Engineering and Management: The CommonKADS Methodology”. The MIT Press.

Tricot, C. and Roche, C. 2005. “Visual Information Exploration: A Return on Experience in Knowledge Base Management,”. ICAI'05 - The 2005 International Conference on Artificial Intelligence. Las Vegas, Nevada, USA, 2005.

Zacklad, M. and Grundstein, M. 2001. “ Ingénierie et capitalisation des connaissances”. Hermes. Informatique et Systèmes d’Information.

Ontologos corp. http://www.ontologos-corp.com

Protégé. 2005. http://protege.stanford.edu "The Protégé Ontology Editor and Knowledge Acquisition System," 2005.

SIMULATION IN BUSINESS AND RISK ANALYSIS

Evolving Collusion in a Simulation of an Oligopolistic Marketplace

Stephen Huxford and David Hunt
Faculty of Information Technology
Monash University
Wellington Road CLAYTON 3800,
Australia
E-mail: {Stephen.Huxford|David.Hunt}@infotech.monash.edu.au

KEYWORDS

Genetic Algorithm, Simulation, Marketplace

ABSTRACT

The time phased behaviour of participants with interdependent pricing strategies in a marketplace is simulated using a GA. Each member of the population embodies a pricing strategy and their fitness equates to their profitability. Population members are monitored for collusive behaviour which is illegal in many real marketplaces. For simple market share models simplistic strategies are evolved quickly which prevent the exploration of the space of all strategies by the GA. More complex market share models also produce simplistic strategies but with unpredictable transitions between these strategies in which the GA explores more of the solution space. Strategies that appear to be collusive are evolved and their origin and nature are examined.

INTRODUCTION

An oligopolistic marketplace is one in which relatively few sellers compete to maximize their profits by selling different brands of the same product. In addition the marketplace is such that the actions of any given seller will influence the actions of the remaining sellers. This interdependence makes it a non-trivial exercise to discover selling strategies that produce market-leading profits over time.

One promising approach to uncovering such strategies is to evolve a simulation of the marketplace. If a representation of a time-phased selling strategy can be codified then several of these strategies can be allowed to compete for the profits available from the marketplace. If we begin with random strategies the most profitable strategies can be selected after a marketplace simulation. These strategies can be subject to genetic operations and allowed to compete in a second marketplace simulation. Repeating this cycle of simulation, selection of the fittest (most profitable) and genetic operation should encourage the evolution of profitable strategies.

Although retail marketplaces often involve complexities such as buyer segmentation and the effects of advertising, some important markets are far simpler. For instance, in the Australian petroleum retail marketplace profits depend almost exclusively on a seller's current price point and this price point's related market share. In addition the marketplace is oligopolistic. To protect buyers, Government regulations forbid price collusion amongst petroleum sellers. Despite or because of these regulations more than 30 inquiries into the pricing strategies of the petroleum sellers have been held over the last twenty five years. Any insights that an evolving simulation of such a marketplace can provide, especially with respect to pricing collusion, would be valuable.

SIMULATION AND GENETIC ALGORITHM DETAILS

A marketplace simulation (henceforth known as a game) proceeds as a series of time steps. At each time step sellers (henceforth known as players) take as input their price rank at the last two time steps. Their output is a price movement which may or may not create a new price rank for them at the current step. A player's price rank at the current time step will subsequently influence their price movements for the next two time steps since it will be part of that player's price rank history. Players begin a game with a common initial price. For four players there are 20 possible price rankings and therefore 20^2 possible two step price rank histories. More generally, if there are n players then the number of price rank histories is:

$$\left\{ \sum_{k=1}^{n} {}^{n-1}C_{k-1} \times k \right\}^2 .$$

The example presented in Table 1 clarifies the concept of a price rank history where each price rank includes the relative position of a player's price and the ranking structure of all players' prices.

Table 1. An Example of a Price Rank History

Time Step	Prices (player 1, 2, 3 and 4)	Price Rank, player 2
t - 2	(59, 61, 61, 62)	2(1223)
t - 1	(60, 61, 62, 63)	3(1234)

In each time step of a game only three price movement responses are allowed namely: increment or decrement the current price by one unit or make no change.

Conversion of a player's remembered two step price rank history to a price movement response is performed by looking up the appropriate position on a chromosome consisting of a 3-nary byte string. This string is the representation of a player's strategy. The string is 400 bytes long and each byte can be considered as a gene that can be in one of three possible alleles: -1, 0 or +1, to indicate a price movement response. The position of a byte in the byte string indicates which of the 400 possible price rank histories it contains the response to. In practice several additional start up bytes are required to take care of responses in the first two time steps of each game when players have no remembered two step price rank history. Each gene represents an independent response component of a strategy so it can be mutated in isolation and crossover operations do not require a repair mechanism.

This representation of a strategy by a byte string suitable for processing by a GA, is an adaptation of one used by Axelrod (Axelrod 1987) to investigate the Iterated Prisoners' Dilemma (IPD) game and an extension of this representation by Yao and Darwen (Yao and Darwen 1994) to investigate the N-person IPD.

This representation has been used previously to explore aspects of an oligopolistic marketplace other than collusive behaviour (Cheung et al. 1997a; Bedingfield et al. 1997a; Cheung et al. 1997b, Bedingfield et al. 1997b). A related representation has been described by Cheung and Hinde (Cheung and Hinde 2000).

The genetic algorithm, used in this study, typically employs a population of 20 player strategies with moderate one-point crossover and mutation rates of 0.2 and 0.01 respectively. Fitness scaling is used to both dampen extreme fitness values and emphasize smaller differences. Selection is performed using the roulette wheel approach. The algorithm is based on one presented by Goldberg (Goldberg 1989).

During each generation the marketplace is simulated ${}^{20}C_4$ times as a game is played for each combination of 4 player strategies taken from the available population of 20. Typically a game proceeds for 22 time steps. A strategy's fitness (profit) is accumulated over every step of every game it participates in during a generation.

MARKET SHARE MODEL

Profit calculations are based on price and market share which, of course, are interrelated in most retail markets. One of the more popular price/market share models amongst marketing analysts is the Multinomial Logit (MNL) model (Cooper et al. 1988; Bell et al. 1975). According to this model the market share of a player with price p_i is:

$$S_i = \frac{e^{-\alpha p_i}}{\sum_{j=1}^{n} e^{-\alpha p_j}} \quad \text{where } \alpha > 0 \text{ is an adjustable constant.}$$

To help visualize the effect of this model 3D profit plots for $\alpha = 0.02$ and $\alpha = 0.2$ are shown in Figures 1 and 2 respectively. To make the plots visually practical one player's price is varied against the sum of the remaining three. In these plots the player 2, 3, 4 price axis is scaled as if players 2, 3 and 4 all have the same common price. Low values of alpha lead to simple market share and profit surfaces where high prices are punished only weakly with reduced market share and resultant profits. High values of alpha produce more complex surfaces that punish high prices more severely.

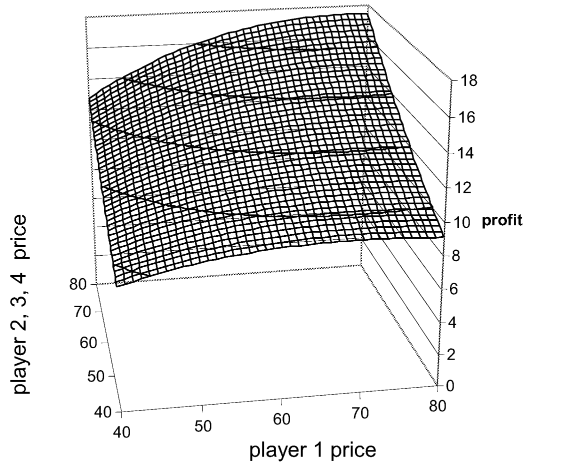


Figure 1: Profit Plot for MNL Market Share Model with $\alpha = 0.02$

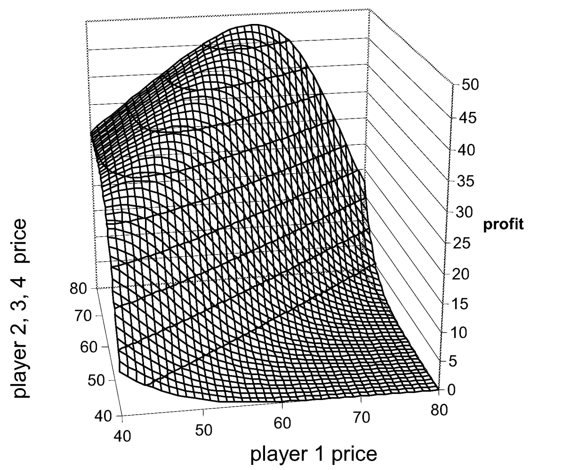


Figure 2: Profit Plot for MNL Market Share Model with $\alpha = 0.2$

MNL MARKET SHARE MODEL WITH LOW ALPHA VALUES

If alpha is low ($\alpha < 0.05$) the MNL market share model produces a profit surface that presents an unequivocal contour to an evolving strategy. No matter what the price of the other players the most profitable strategy for any given player is to increase its price. This leads to some unusual gene usage during strategy evolution.

The strategies in a population compete in many 4-player games over a generation to accumulate profit which is interpreted as their fitness which determines their probability of selection into the next generation. Because of the unequivocal pressure to increase prices brought to bear by low alpha MNL market share models the +1 allele should be increasingly selected for all genes as generation follows generation. Actually detailed and summary reporting of the genetic processes at work reveal only a few genes are actually biased in this way across the strategy population. When a gene is biased it means it is considered to be a “good idea” across the population. Specifically, this price movement response to this price rank history will contribute better than the other alleles of this gene to a strategy’s overall profitability.

Genes will not be meaningfully biased unless they are used. If they are not used they cannot contribute to a strategy’s profitability and therefore cannot influence the selection of a strategy and therefore indirectly the selection of themselves.

Since at the start of each game all players have the same price point, a frequently used gene is the 1(1111) gene i.e. first amongst four equal firsts. Once this gene has become biased in the strategy population all games follow a predictable pattern: all players increment their price at every step and remain locked with even prices. Only the 1(1111) gene is used. Additional genes used to decide on price movements in the first two steps of any game when players do not have a two-step price rank history to refer to complicate things a little but since these genes are also subject to evolutionary pressure they are quickly biased into their +1 allele.

Once the above situation has become established there is very little exploration of the space containing all possible strategies. All such exploration occurs in the transition from the initially random strategies to strategies with their start up and 1(1111) gene biased to their +1 allele. This transition usually takes less than 20 generations. In addition if mutation hits either of the start up genes or the 1(1111) gene there is a brief use of genes other than the 1(1111) gene. These genes have had little use and are therefore unbiased. A brief chaotic drop in fitness results but the constant evolutionary pressure to increase profits soon re-establishes the equilibrium situation.

A fitness profile over one hundred generation is shown for a marketplace based on the MNL market share model with α = 0.02 in Figure 3. It is an average over ten evolutions where each evolution is based on a different random seed for its stochastic processes including the generation of the original, random population. The fitness profile for an individual evolution is not shown since these do not show anything other than trivial deviations from the average. Fitness is shown as a percentage of standard profitability which is the profit made by a player if all players maintain their original price for the duration of all the games for a generation.

The initial transition and subsequent perturbations where some non-trivial part of the space containing all possible strategies is being explored has already been explored (Cheung et al. 1997a; Bedingfield et al. 1997a; Cheung et al. 1997b, Bedingfield et al. 1997b). However, at least two notes of caution should be sounded. The resulting strategies are not profitable in the face of any other strategy (i.e. they are not good predator strategies) but only profitable in the face of identical strategies. In essence, a fragile common strategy is evolved. This strategy could be seen as collusive as it relies on all players following an identical strategy of increasing a common price. Actually the individual players are only following their own best interests in this case and are not guilty of collusion. It should be noted that the actual chromosomes of each of these “identical” players may be quite different but if the chromosome genes that actually get used are identical this amounts to identical strategies. The non-identical genes are relatively unbiased as a result of the short initial transition and therefore produce chaotic results when called into play during subsequent perturbations.

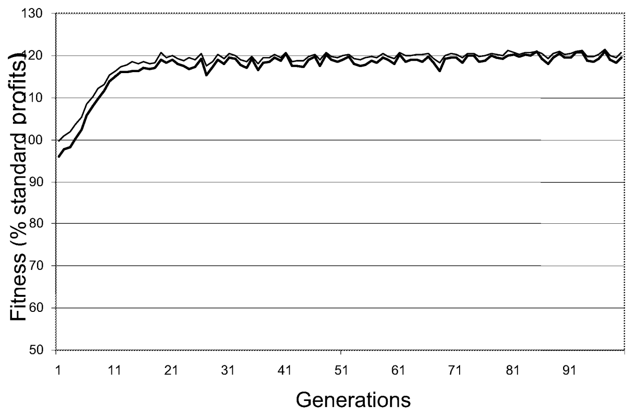


Figure 3: Best and Average Fitness over 100 generations (averaged over 10 evolutionary instances) using the MNL Market Share Model with α = 0.02

The second note of caution concerns the fact that members of the genetic algorithm population are pitted against each other to determine fitness. This has the effect of constantly changing the evolutionary background. In the first generation strategies emerge that best deal with other random strategies. In the second generation these same strategies are now pitted against the best strategies from the first generation. They were selected because they dealt well with random strategies but they are now expected to do well against a different set of “semi-smart” strategies. Again it must be remembered that what is being evolved is a profitable common strategy not a profitable predatory strategy.

This feature of the evolving simulation is not unrealistic since presumably in a real oligopolistic marketplace, players take note of successful pricing strategies used by their competitors and incorporate them into their own strategies in a never-ending, escalating competition.

MNL MARKET SHARE MODEL WITH HIGH ALPHA VALUES

If alpha is high (α > 0.05) the MNL market share model produces a profit surface that presents a more equivocal contour to an evolving strategy than the low alpha case. Unlike the low alpha case the most profitable strategy is not to always increment price but depends on the current price of this strategy and the current prices of the other three strategies in a game.

A fitness profile over five hundred generation is shown for a marketplace based on the MNL market share model with α = 0.2 in Figure 4. It is an average over ten evolutions where each evolution is based on a different random seed for its stochastic processes including the generation of the original, random population.

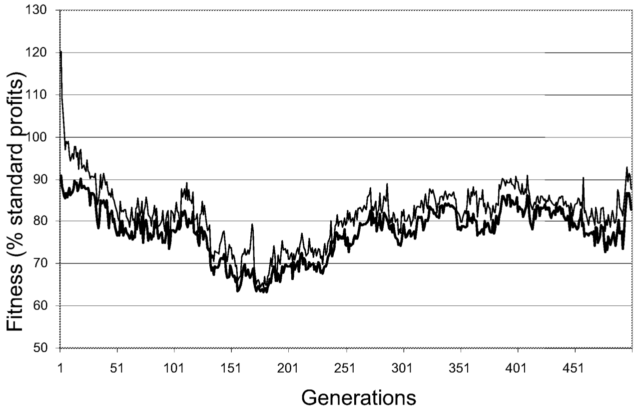


Figure 4: Best and Average Fitness over 500 generations (averaged over 10 evolutionary instances) using the MNL Market Share Model with α = 0.2

This is indeed a disappointing plot for an evolutionary process that hopes to evolve more profitable strategies over the generations. In fact it seems counterintuitive that after selecting for the most profitable strategies in a generation these strategies often become less profitable in the next generation. However it must be remembered, as discussed already, that the evolutionary background changes from one generation to the next. A good strategy in generation n may not be as appropriate in generation n + 1 since the opposing strategies have changed.

Examination of the evolutionary instances that contribute to the average plot reveals little reproducibility but some large scale similarities. These include sustained attraction to at least three fitness levels. In terms of standard profitability (defined above) the sustained fitness levels are at 114%, 92% and 54%. One or more of these fitness levels appears in all ten evolution instances that make up the average profile shown in Figure 4. A fitness profile for one of the evolutionary instances is shown in Figure 5 and discussed in detail below.

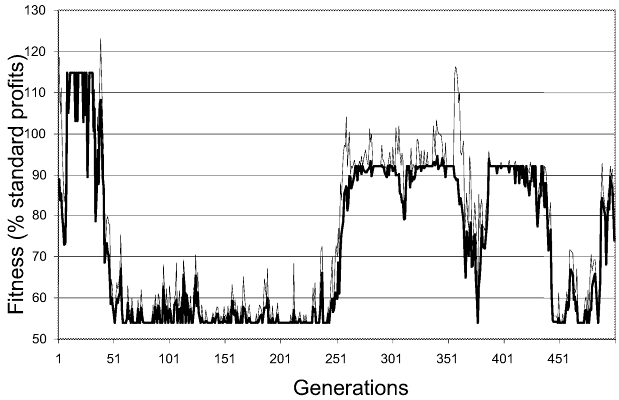


Figure 5: Best and Average Fitness over 500 generations (single evolutionary instance) using the MNL Market Share Model with α = 0.2

The levels themselves are simply explained. They correspond to extremely homogenous strategy populations where the few genes that are used are heavily biased into their +1 allele (114% level), 0 allele (92% level) or -1 allele (54% level). The transitions between levels are more interesting.

Both the transitions and the levels result from the market share model and evolutionary pressure. The MLN market share model with high alpha values punishes higher prices with very small market shares and consequently very small profits. The evolutionary pressure is not simply for greater profits but for a shared strategy that delivers greater profits to all strategies in a game. A profitable predatory strategy quickly becomes a victim of its own success when it faces itself in subsequent generations. Put simply, wolves soon starve when there are no sheep!

In the fitness profile for the evolutionary instance shown in Figure 5 the 114% level emerges within 10 generations. The initial price for all players in all games is arbitrarily set at 60 units. Since a price array of (59, 60, 60, 60) is more profitable for player 1 than (61, 60, 60, 60) it is somewhat surprising that the +1 allele is selected for. The reason for its selection is that the price array of (61, 61, 61, 61) is more profitable for all players than (59, 59, 59, 59). Having said this, the 114% level only emerges more or less immediately in 30% of evolutionary instances. The vagaries of the initial, random population often leads to an initial transition to one of the other two levels.

The 114% level collapses spectacularly to the 54% level at generation 31. The collapse can be traced to a mutation of the 1(1111) gene from +1 to -1. This is the first time this particular mutation occurs. The individual with the critical gene mutated invokes the use of many unused and therefore unbiased genes in its competitors during marketplace games. Prices become unlocked and the delicate balance that maintained the 114% level is broken. Strategies that cause lower prices tend to be more profitable and there follows a flooding of the population with such strategies. The result is the emergence of a homogeneous population where the

1(1111) gene is in the -1 allele and with few exceptions is the only gene used.

Unlike the 114% level the 54% level is relatively stable. Every game will begin with all players decrementing their price in step. Any player electing (via a mutation of the 1(1111) gene) not to follow the others down is immediately punished with reduced profits and less chance of selection into the next generation.

The transition from the 54% level to the 92% level is therefore puzzling. Its genesis is a mutation of one of the start up genes of a member of the population. This event leads, within a few generations, to the creation of an interesting and complex strategy which is the fittest strategy for generation 254. Variations of this strategy subsequently power this unexpected transition to a homogenous population in which the 1(1111) gene is biased to its 0 allele and more or less used exclusively.

Several plots that show the complexity of the fittest strategy of generation 254 are shown in Figs 6, 7 and 8. The plots of average price and average profit per game step over all games in a generation are, on the face of it, contradictory. Despite maintaining a higher price than the population average at each game step the strategy manages to also maintain a higher profit than the population average beyond the fourth game step. This would appear to contradict the market share model. The plot of frequency of final price rank hints that the price and profit plots are hiding some interesting detail in their averaging. In particular the strategy results in several final price ranks not just always first or always last for instance. The profits are apparently picked up in several different ways probably depending on the particular opponent strategies in a game.

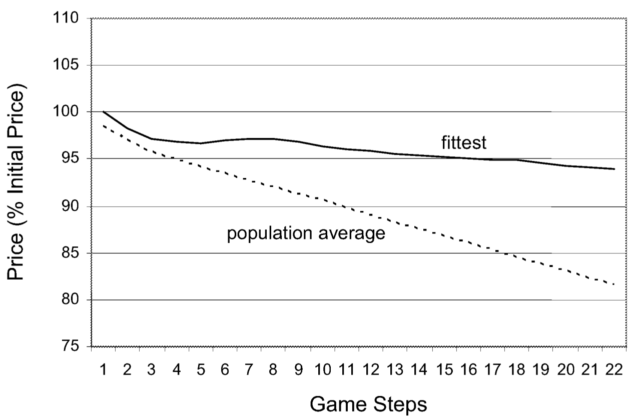


Figure 6: Average Price set by Strategies at each game step over all games played in Generation 254

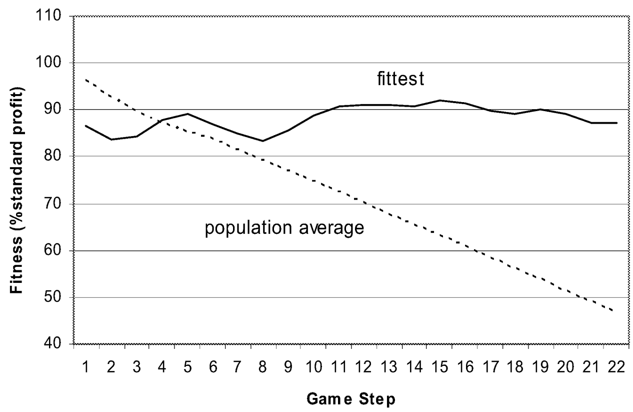


Figure 7: Average Profit Accumulated by Strategies at each game step over all games played in Generation 254

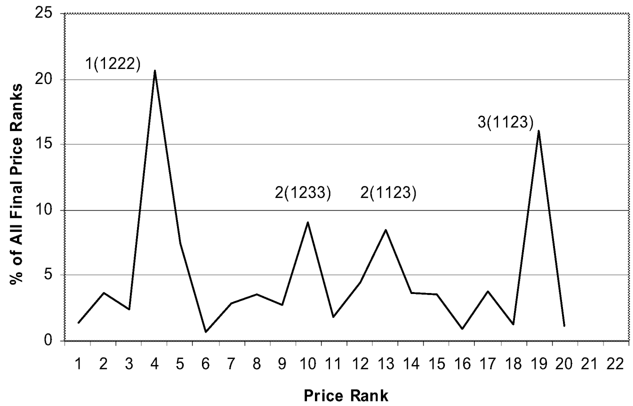


Figure 8: Frequency of Final Price Ranked achieved by the Fittest Strategy in Generation 254

COLLUSION

All three sustained fitness levels can be construed as collusion since all involve the maintenance of a fixed identical price by identical strategies. However, the 54% level differs from the 114% and 92% levels. Strategies that emerge to break the 54% strategy are usually unfit and quickly removed from an evolving population. Strategies that emerge to break the other levels are usually fit and quickly invoke a transition to a lower level. Although it requires a mutation to cause a transition in the evolving simulation in the real world this corresponds to a willful change of strategy. Therefore to attain and maintain one of the higher profit levels strategies must be willfully converged and maintained by members of an oligopoly.

ONGOING AND FUTURE WORK

Monitoring techniques that help understand how complex strategies, such as the one mentioned above, achieve their success are being developed. There is a delicate balance between being overwhelmed by fine detail and burying that same detail in averages.

Bonuses and penalties for collusion are being introduced into the evolving simulation. To mimic the real world bonuses are given after n time steps of identical pricing and penalties after m steps where m > n. Different bonus and penalty ratios are being tried along with different delays in the time steps between bonus and penalty.

The correspondence between Nash equilibria from Game Theory and the sustained fitness levels discovered in the current work is being investigated.

In parallel with the above work predatory strategies are being evolved by pitting members of a genetic population against a static but large number of arbitrary and semi-smart strategies rather than against each other.

REFERENCES

Axelrod, R 1987. "The Evolution Of Strategies In The Iterated Prisoner's Dilemma". In *Genetic Algorithms and Simulated Annealing*, L.Davis ed., Morgan Kaufmann, 32 - 41

Bedingfield, S.; S. Huxford; and Y. Cheung. 1997. "Simulating Pricing Behaviours Using a Genetic Algorithm". In *Lecture Notes in Computer Science*, Vol. 1305 Springer-Verlag 3-10

Bedingfield, S., Huxford S., Cheung, Y. 1997. "Evolving Oligopolistic Behaviour Using a Genetic Algorithm". In *AISB Proceedings of Workshop on Evolutionary Computing* (University of Manchester, UK April 7-8). 43-46

Bell, D. E.; R.L Keeney.; and J.D.C Little. 1975. "A Market Share Theorem", *Journal of Marketing Research*, XII (May), 136-141.

Cheung, Y.; S. Bedingfield; and S. Huxford 1978. "Oligopolistic Behaviour of the Australian Petroleum Industry". In *Proceedings of International Conference on Computational Intelligence and Multimedia Applications* (Gold Coast, Australia, Feb. 10-12). Ed. B Verma & X Yao, 161-165

Cheung, Y., Bedingfield, S., Huxford, S 1997. "Monitoring And Interpreting Evolved Behaviours In An Oligopoly". In *Proceedings of the IEEE International Conference on Evolutionary Computation* (Indiana University, Indianapolis, April 13-16). Sponsored by IEEE Neural Networks Council 692-701

Cheung Y. and C. Hinde 2000. "Economic Modelling with A General Purpose GA" In *Advances in Intelligent Systems: Theory and Applications* M. Mohammadian (Ed), IOS Press, 96-101

Cooper, L. and M.Nakanishi. 1988. *Market-Share Analysis, Evaluating Competitive Marketing Effectiveness*. Kluwer Academic Publishers

Goldberg, D. 1989. *GAs in Search, Optimisation and Machine Learning*, Reading Addison-Wesley.

Yao, X and P. J. Darwen. "An Experimental Study of N-Person Iterated Prisoner's Dilemma Games". In *Informatica 1994,* Vol 18, 435-450

AN INTEGRATIVE APPROACH TO INVESTMENT APPRAISAL: SIMULATING THE RISK OF AN OPTIMUM DECISION

Athanassios A. Rentizelas	Georgios G. Tziralis	Konstantinos A. Kirytopoulos
Sector of Industrial Management and Operational Research	Sector of Industrial Management and Operational Research	Financial and Management Engineering Department
Mechanical Engineering School	Mechanical Engineering School	Business School
National Technical University of Athens	National Technical University of Athens	University of the Aegean
Iroon Polytechniou 9, 15780, Athens Greece	Iroon Polytechniou 9, 15780, Athens Greece	Fostini 31, 82100, Chios Greece
E-mail: arent@central.ntua.gr	E-mail: gtzi@central.ntua.gr	E-mail: kkir@central.ntua.gr

KEYWORDS

Investment Appraisal, Genetic Algorithms, Monte Carlo Simulation, Risk Analysis.

ABSTRACT

The ever increasing complexity and variety of investment decisions has, sometimes, proved the commonly used investment appraisal approaches as insufficient. The point estimates for an uncertain future seem at least unconvincing. At the same time, the investors are in need of sound arguments, which will be able to shape the investment specifications and appraise their uncertain nature.

This paper proposes an innovative approach that attempts to fulfil the aforementioned needs. The two-step approach reaches an optimum through a Genetic Algorithm optimization and then models the environment's uncertainties with a Monte Carlo simulation. The approach, thus, offers the best investment characteristics, as well as its implied risk. The proposed method is further demonstrated through an extensive Case Study.

INTRODUCTION

The constantly increasing complexity of business ecosystems has rapidly expanded the variety, as well as the importance and risk, of potential investments. Investment decisions are now more crucial than ever for any kind of production companies, since the right investment in a global economy framework may aid the prosperity of the company, while the wrong investment may lead even to bankruptcy.

The investment's problem is as old as the economy itself. In order to succeed in optimum utilisation of the limited available investment resources, each decision should take into consideration a huge set of factors, which are either defined by the investor, or can be affected only by external factors. The decision to proceed or not to an investment is based upon the outcome of the dominant economic criteria, such as the Net Present Value (NPV), the Internal Rate of Return (IRR) and the Payback period (Van Groenendaal and Kleijnen 2002, Biezma and San Cristobal 2005).

The typical investment analysis forces the decision maker to assign values to every single variable of the investment model and to perform at best a sensitivity analysis afterwards. This approach will lead to a non-optimal outcome, not capable of giving the highest possible amount of information to the decision maker. An alternative approach is to take into account risk management principles while estimating the outcome of an investment (Pike and Neale 1993).

Risk management has always being regarded in the economic field as a potential fluctuation of the economic variables (Alexander C. 1999). The economic and technical risks (Hammer 1972) are the first structured approaches met in the scientific literature. However, there are many more scientific sectors where scientists try to investigate and exploit the benefits of risk management. Research is focused, but not limited, to the fields of human errors, occupational safety (Reason 1990), supply chain management (Ponis 2005) and quite recently to project risk management (Kirytopoulos et al 2001, Leopoulos et al 2003).

The paper is focusing on risks that might affect an investment, thus putting in jeopardy the capital invested by a company or an individual. In nearly any type of investment, risks are most of the times regarded to be external. External risks are those that the probability of their occurrence cannot be modified by the investor (Leopoulos and Kirytopoulos 2004). In other words, the investor knows that risks might happen but cannot do anything to change the probability of their occurrence. Pike (1986) has modelled the investment decision through the use of statistics. He adjusted a normal distribution on each variable of the Net Present Value (NPV) function affected by risks and by using statistics calculations he reached an NPV outcome described not just as a single value but a normal distribution, thus providing valuable information to the investor.

The first part of the paper is concerned with defining the problem and analyzing and presenting the conceptual model. The model aims at providing an alternative and innovative approach for investment appraisal, by applying optimization methods (Genetic Algorithms) while simultaneously considering investment risk.

In order to highlight the characteristics of the conceptual model, a Case Study is presented at the second part of the paper. The Case Study concerns an investment in a tri-generation power plant (electricity, heat and cooling), fuelled by renewable energy sources, namely several types of biomass. This particular type of example was chosen on the one hand due to the increased interest on renewable energy sources following the recent energy crisis and extreme oil prices and on the other hand because this specific case possesses all the characteristics that make it an ideal candidate for applying the conceptual model presented.

METHOD DESCRIPTION

Problem formulation

The fundamental scope of the problem is to assess and select an investment project by optimizing the investor's choices while in parallel considering the risk correlated with these choices. Among various economic criteria, NPV is selected, without loss of generality, as the one most commonly used.

The NPV of an investment is dependent on the values of a set of variables, which are either defined by the investor's decisions, or are fully determined by external to the decision system factors. A potential investor's aim should be therefore twofold: On the one hand, the best possible, in terms of NPV, investment should be selected by optimizing the investor's decisions. On the other hand, specific knowledge is needed on the volatility of the external factors and on how this volatility affects investor's decisions, in regard to his risk policy.

It is therefore apparent that the investor's defined parameters and the external variables should be considered separately. The paper presents an integrative approach, which attempts first to compute the NPV in a way that heuristically optimizes the decision-making process by shaping the best feasible investment and second reliably quantify its implied risk.

Conceptual model presentation

As stated above, NPV is a function composed of investor defined parameters, as well as external variables. In mathematical terms the NPV function is described in equation 1.

$$\text{NPV} = f(\mathbf{P},\mathbf{V}) = f(P_1,\ldots,P_n,V_1,\ldots,V_m) \qquad (1)$$

where:

$\mathbf{P} = [P_i]$, $i = 1,\ldots,n$, parameters that their value is defined by the investor and

$\mathbf{V} = [V_j]$, $j = 1,\ldots,m$, variables that their value is an outcome of external factors.

The common approach to modeling external variables is to assume a single value for each one of them (Savvides 1994). However, the deterministic consideration of a probabilistic reality is by definition erroneous. In this paper, a different view is proposed. Let

$$V_j \sim X_j(mean_j, std_j) \qquad (2)$$

that is, the variable V_j follows a statistical distribution X_j with known characteristics $mean_j, std_j$. By this way, the implied volatility of each external factor is not omitted, but modeled through a statistical distribution. The distributions' characteristics are expert-defined.

The first step of the approach is to assume that each external variable is equal to its most probable value, which, eventually, leads to the most probable scenario:

$$\forall j : V_j = V_{j,probable} \ \& \ \mathbf{V} = \mathbf{V}_{\mathbf{probable}}$$

Having determined the V value, the only thing that remains is to decide for the P value. The occurring optimization problem is solved via use of Genetic Algorithms

$$\text{NPV}=f(\mathbf{P},\mathbf{V}_{\mathbf{probable}})=\max \xrightarrow{GA} \mathbf{P} = \mathbf{P}_{GA} \qquad (3)$$

Genetic Algorithms (GA) comprise an optimization and search technique based on the principles of genetics and natural selection. A GA allows a population composed of many individuals to evolve under specified selection rules to a state that maximizes the selected criteria. GAs implement the biological concepts of genetic re-combination, genetic mutation and natural selection (Haupt and Haupt 2004).

As the name implies, GAs are analogous to the biological concept of a gene, the basic building block for individual traits of organisms, and are combined in chromosomes to form the total, inherited characteristics of an individual. Pursuing the biological analogy further, genes and chromosomes change from one generation to the next so successor generations are never exactly the same as their predecessors. Recalling the "survival of the fittest" theory, alterations in genes and in the composition of chromosomes lead to the creation of individuals with new characteristics. Some of these characteristics are favorable in a given environment and others may be less favorable or even deadly. Consequently, populations find individuals with favorable genes and chromosomal combinations increasing in number and those with less favorable genes and chromosomal combinations decreasing in number. Over time, populations of organisms begin to find an optimal niche in the environment in which the gene pool is well balanced and effective or else the population begins declining (Grupe and Jooste 2004).

Some of the advantages of GA include that it optimizes even non-continuous and non-differentiable functions, with continuous or discrete variables, it does not require derivative information, it simultaneously searches from a wide sampling of the cost surface and it deals with a large number of variables. Even more, GA may succeed in finding the global optimum, due to the fact that the method evaluates simultaneously a large population instead of a single point for most non-heuristic optimization methods. These advantages are intriguing and produce stunning results when traditional optimization approaches fall miserably (Haupt and Haupt 2004).

The solution accomplished by the use of GA, $\mathbf{P}_{GA}$, provides the optimum values for parameters defined by the investor, which define the optimum investment decision, provided that the external factors have values equal to their most probable value.
However, NPV=$f(\mathbf{P}_{GA}, \mathbf{V}_{\mathbf{probable}})$ is just a point-estimate of the final NPV that will occur under realization of the optimum investment scenario. But the $\mathbf{V}_{\mathbf{probable}}$ is just a case among many that can happen, since V_j values are external parameters. The method presented here, proposes a more advanced view by implementing a probabilistic NPV calculation, which is accomplished by the use of Monte Carlo Simulation.

Monte Carlo Simulation (MCS) constitutes a stochastic statistical methodology of quantitative solution and risk assessment for non-deterministic problems, using a pseudo-population of randomly produced alternative scenarios from prescribed statistical distributions. The MCS suggests the modelling of the range of possible values for each input variable and the following reproduction of an efficient number of scenarios, so that the depiction of the respectively large number of results in a density function diagram could attribute in a reliable manner the needed distribution of the output variable, showing in parallel the possibility of occurrence for each value and marking out extreme or probable results (Vose 2000).
A MCS is performed for $k = 1,...,q$ iterations, where q is typically larger than 1000, by picking randomly values from the statistical distributions X_j, such that

$$\forall j,k : V_{jk} \sim X_{jk}(mean_j, std_j) \qquad (4)$$

and consecutively computing $\forall k : \mathbf{V} = \mathbf{V}_k$ a value of

$$\text{NPV}_k = f(\mathbf{P}_{GA}, \mathbf{V}_k) \qquad (5)$$

The NPV is finally calculated not as a single figure but as a probability density function, which is based on optimum investor's decisions and represents the quantified volatility of external parameters.

The alternative representation of the k values of NPV derived from the MCS in a cumulative probability diagram also provides valuable information. The confidence level at the point of zero NPV could serve as a useful meter of risk. Moreover, the value of NPV at a selected confidence level provides a strong decision tool, which can serve as a decision criterion for the final acceptance of the project.

CASE STUDY

Case description

The Case Study concerns an investment analysis for a tri-generation power plant, given the demand of a specific customer for heat and cooling. Heat and cooling demand profiles are available for a specific residential sector of 500 customers. The geographical positioning of the plant and the customers is assumed to be fixed. The revenues of the power plant under consideration will stem from selling electricity to the grid as well as heat and cooling to the customers via a district heating network that will be constructed. The price of heat is assumed to be a fixed percentage of the cost of heat obtained by using oil (as this is the common practice) whereas the price of cooling is a fixed percentage of the cost of cooling obtained by electrical compression chillers.

The power plant will consist of a base-load co-generation module (gasifier and reciprocating engine) and a biomass boiler for peak-load heat production. There are four biomass types available in the region, therefore all of them are considered as potential fuel sources for the plant. The multi-biomass approach has been used, due to the significant cost reductions that can be obtained at the logistics and warehousing stages of the biomass supply chain, compared to the single-biomass approach (Rentizelas et al. 2005). The basic characteristics of the biomass sources are presented in Table 1.

Table 1: Biomass Characteristics

Biomass type	Type 1	Type 2	Type 3	Type 4
Price (€/kg)	0.05	0.01	0.03	0.01
Heating Value (KJ/kg wet)	10000	10500	12000	10200
Density (kg/m3)	400	350	450	350
Biomass availability (months)	Sept-Oct	Feb	Feb-Mar	May
Yield (tons/hectare)	2	4	5	3.5

The parameters included in the GA optimization are:

- Capacity rates for the base-load co-generation plant as well as the peak-load heat boiler
- Amounts of each biomass source to be procured every year

- Initial biomass inventory at the beginning of every year of operation.

The objective function (equation leading to the NPV value) of the optimization problem is the NPV of the investment, during its operational lifetime. Investment considered includes the power plant, the supply chain of the biomass, the construction of the District Heating and Cooling network and connection to the customers, as well as the electricity transmission line and connection to the grid. The financial data used for the NPV calculation are presented in Table 2.

Table 2: Financial data

Interest rate (most probable value)	6%
Inflation rate (most probable value)	3%
Investment lifetime (fixed value)	15 years
Public Subsidy (most probable value)	40%

The specific case is highly suitable for optimization with GA, as the objective function is non-linear and non-continuous, due to economies of scale of the biomass co-generation plant and analytical modelling of the biomass logistics network. Furthermore, several parameters may take a very wide range of values, therefore MCS is used to depict the effect that they may have on the NPV value. For example, oil price has undergone significant fluctuations lately, and has a significant dual effect on the NPV value: on the one hand it affects revenues from heat sales and on the other hand it affects reversely logistics cost for transporting and handling biomass. The characteristics of distributions X_i for all parameters V_i have been derived by experts and are presented in Table 3. The third column of the table refers to $\mathbf{V}_{\mathbf{probable}}$, which gives the values that have been used in the GA optimisation model.

It can be seen from Table 3 that various distribution types have been chosen to simulate the volatility concerning the values of various variables. For example, most of the variables that relate to construction work have been simulated with triangular distributions, as this is common practice in construction project management (Kirytopoulos et al 2001). Other variables that have a certain value today, but are expected to follow a particular trend in the future (e.g. increasing electricity prices) are simulated also by properly modified triangular distributions. Financial figures that may fluctuate in a totally unpredictable manner have been modelled by uniform distributions, with logical upper and lower bounds. The adoption of Monte Carlo simulation gets over the significant limitations of Pike's (1986) model that demands the use of the Central Limit Theorem. Thus, the limitation of using the same type of distribution to describe the totality of the parameters included in the NPV value can be omitted in the proposed approach, where the use of MCS is introduced.

Table 3: Characteristics of Parameters' Distributions

Parameter	*Distribution*	*Probable value*	*Standard deviation*	*Minimum value*	*Maximum value*
Income from electricity (€/kwh)	Triangular	0,06611		0,05	0,1
Income from electric power reimbursement (euro/kw/month)	Normal	1,7	0,2	1	
Power reimbursement factor (-)	Normal	0,9	0,05	0,7	1
Income from cool (euro/Kwh)	Triangular	0,036		0,03	0,05
Public subsidy on investment (%)	Uniform	40%		30%	50%
Total CHP efficiency rate	Triangular	0,85		0,75	0,88
Thermal CHP efficiency rate	Triangular	0,59		0,5	0,65
Thermal boiler efficiency rate	Triangular	0,8		0,7	0,85
Chiller efficiency	Normal	0,7	0,1	0,65	0,9
Inflation (%)	Uniform	0,03		0,02	0,05
Compound interest rate (%)	Uniform	0,06		0,05	0,08
Investment cost for CHP plant – reference case (€/kWel)	Normal	2000	100	1500	3000
Investment cost for Boiler – reference case (€/kWth)	Uniform	200	10	160	230
Diesel cost (€/lt)	Uniform	0,6		0,45	1
Biomass price – type I (€/ton wet)	Uniform	50		40	80
Biomass price – type II (€/ton wet)	Uniform	10		5	40
Biomass type – type III (€/ton wet)	Uniform	30		15	60
Biomass price – type IV (€/ton wet)	Uniform	10		5	30
Warehouse & equipment investment cost (% of base scenario)	Triangular	1		0,85	1,25
Warehouse & equipment operational cost (% of base scenario)	Triangular	1		0,9	1,15
Electricity transmission network investment cost (% of base scenario)	Triangular	1		0,9	1,2
District heating network investment cost (% of base scenario)	Triangular	1		0,85	1,3
Cooling equipment investment cost (% of base scenario)	Triangular	1		0,9	1,1

Results

The optimum solution for the set of parameters **P** obtained by GA optimization ($\mathbf{P}_{GA}$) is presented in Table 4. The optimum investment decision arises as clearly defined in terms of plant capacity, initial inventory and annual amount of biomass per type. The proposed approach provides a single optimum value for all the parameters that are determined by investor's choices.

Table 4: Optimized Variables ($\mathbf{V}_{GA}$)

Variables	*Value*
Thermal power capacity of CHP plant (kW)	4.388,56
Thermal power capacity of peak-load boiler (kW)	2.584,71
Total annual amount of biomass – type I (tons/year)	815,13
Total annual amount of biomass – type II (tons/year)	7.544,68
Total annual amount of biomass – type III (tons/year)	749,58
Total annual amount of biomass – type IV (tons/year)	7.004,80
Initial biomass inventory (m^3)	22.877,87

Following the adoption of these optimum prices, a MCS is performed to reveal the implied volatility that characterises the rest of the variables (refer to Table 3). The characteristics of the NPV distribution that occurred through the 10.000 iterations of the simulation are presented in Table 5 that follows.

Table 5: NPV Distribution Characteristics

Minimum (€*10^6)	-3,50	Mode (€*10^6)	12.09
Maximum (€*10^6)	37,47	Left X (€*10^6)	5,08
Mean (€*10^6)	13,28	Left P	5%
Median (€*10^6)	12,95	Right X (€*10^6)	22,54
Std Dev	5,31	Right P	95%
Variance	41,5	Diff X (€*10^6)	17,46
Skewness	0,327	Diff P	90%
Kurtosis	2,931		

An adequate amount of information is now available to the decision maker, in order to appraise in his own customised way the risk that is implied in the optimum investment. For example, the 'Left P' statistic indicates that there is a 95% possibility that the NPV of the proposed investment would exceed 5,08 million €. The probability density function diagram of the simulated scenarios is presented in Figure 1.

The figure reveals in a visual way the possibility of occurrence of each NPV, as long as it marks out the most probable or extreme results. It also comprises an advanced but convenient view for the comparative assessment of two or more projects.

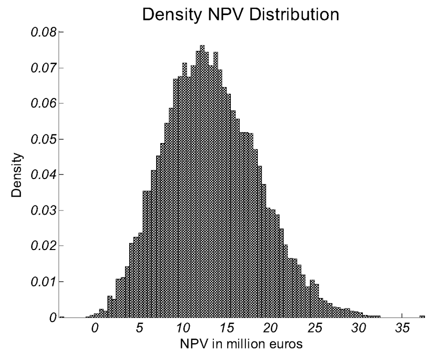


Figure 1: Density NPV Distribution for Optimum Investment

An alternative depiction of the NPV distribution is provided in Figure 2.

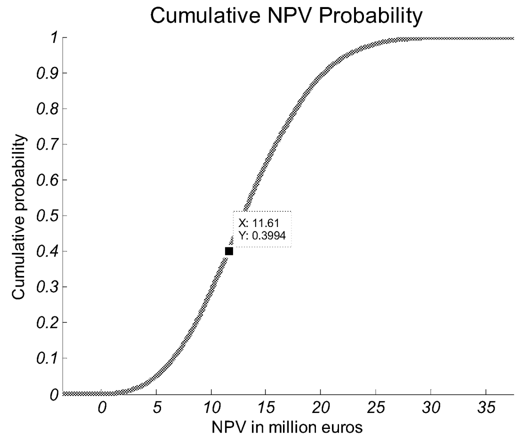


Figure 2: Cumulative NPV Probability Distribution for Optimum Investment

The cumulative graph stands also as a very utile tool. It reveals straightway that the likelihood of a deficit is practically zero. Likewise, it provides the minimum value of the NPV for a given confidence level. For example, as it is pointed on the diagram, it states that there is only 40% possibility that the project's NPV will be under 11,61 million €.

CONCLUSIONS

The paper proposes an innovative integrative approach to investment appraisal. The factors that affect the investment (NPV) under consideration are grouped into two sets. The first one contains the parameters that are determined by the investor's choices, such as the power plant capacity. This set substantially designates the investment's specifications. The other set is comprised of variables that their values are shaped by external to the investor decisions, such as the oil price. This type of variables is responsible for the risk that is embodied in the investment.

The approach proposed in this paper suggests that the parameters should be calculated by the use of a GA, while the volatility of the variables will be modelled, afterwards, though the use of a MCS. The probabilistic modelling of the volatility of external variables is achieved through the use of expert-defined statistical distributions. The most probable values of these variables are adopted, in order to perform an optimization of the selected economic criterion (here NPV). The optimization, performed using the advanced GA technique, results in the best selection of investor's defined parameters. The optimum investment specifications (indicated by the GA) are further used to perform a MCS for the external variable. The simulation results are finally used to form a cumulative probability diagram of the criterion's value distribution.

Hence, the decision maker is able to define the optimum investment and assess its risk in a reliable, convenient and customisable way.

The approach was demonstrated through a case study of an investment regarding a tri-generation power plant. The various external variables were stochastically modelled, and their most probable values were used to perform a GA optimization for the NPV, through which the optimum power plant's configuration and operational characteristics emerged. The MCS that followed revealed the distribution of the NPV, and thus enabled the decision maker to assess his expected returns in a reliable quantitative way. The investment was finally fully determined, in terms of its design specifications, as well as the probability distribution of its expected NPV.

REFERENCES

Alexander C. 1999, "Risk Management and Analysis: Measuring and Modelling Financial Risk", John Wiley & Son Ltd, 1999.

Biezma M.V. and San Cristobal J.R. 2005, "*Investment Criteria for the Selection of Cogeneration Plants – A State of the Art Review*", Applied Thermal Engineering, In Press.

Grupe F. and Jooste S. 2004, "*Genetic Algorithms: A Business Perspective*", Information Management & Computer Security, 12:3, pages 289-298.

Hammer R.W. 1972 "Handbook of system and product safety", Englewood Cliffs, Prentice-Hall: NJ, 1972.

Haupt R. and Haupt S.E. 2004, "Practical Genetic Algorithms", Willey-IEEE 2004.

Kirytopoulos K., Leopoulos V. and Malandrakis C. 2001, "Risk Management: A powerful tool for improving efficiency of project oriented SMEs", Proc. of the 4th SMESME International Conference, Denmark, pages 331-339, 2001.

Leopoulos V. and Kirytopoulos K., 2004, "*Risk Management: A competitive advantage in the purchasing function*", Production Planning & Control, Taylor and Francis, 15:7, pages 678-687.

Leopoulos V., Kirytopoulos K. and Malandrakis C. 2003, "An applicable methodology for strategic risk management during the bidding process", *International Journal of Risk Assessment and Management* 4:1, pages 67-80, Inderscience Publishers.

Pike R. 1986, "Investment Decisions and Financial Strategy", Humanities Press: USA, 1986.

Pike R. and Neale B. 1993, "Corporate Finance and Investment - Decisions and Strategies", Prentice Hall, 1993.

Ponis S.T. 2005, "A Reference Model to Support Knowledge Logistics Management in Virtual Enterprises: A Proposed Methodology", *International Journal of Knowledge, Culture and Change Management*, accepted paper under publication.

Reason J. 1990 "Human Error", Cambridge University Press, Cambridge, 1990.

Rentizelas A, Tolis A. and Tatsiopoulos I. 2005 "Multi-Biomass Supply Chain Modelling and Optimization", Proc. of the World Renewable Energy Congress, Aberdeen, UK, 2005.

Savvides S. 1994, "Risk Analysis in Investment Appraisal", *Project Appraisal*", 9:1, pages 3-18.

Van Groenendaal W. and Kleijnen J., 2002 "Deterministic versus Stochastic Sensitivity Analysis in Investment Problems: An Environmental Case Study", *European Journal of Operational Research*, 141:1, pages 8-20.

Vose D. 2000, "Risk Analysis: A Quantitative Guide", John Wiley & Sons, 2000.

BIOGRAPHIE

Mr. Athanassios RENTIZELAS holds a Mechanical Engineering Diploma from the National Technical University of Athens (NTUA), and an MSc in Operations Management with distinction from UMIST, Manchester, UK. He is currently a PhD Candidate in the Industrial Management & Operational Research Sector of NTUA Mechanical Engineering School. His research interests and expertise lie in the areas of supply chain management, logistics, renewable energy sources, investment analysis, optimization and quality management. He has been awarded various academic awards for his exceptional performance during his studies, as well as scholarships for postgraduate studies.

Mr. Georgios TZIRALIS has received his Mechanical Engineering Diploma (Bachelor plus MSc equivalent) from the National Technical University of Athens and is now a PhD Candidate in the Industrial Management and Operational Research Sector of the NTUA's Mechanical Engineering School. His research interests extend in the area of Computational Methods for Predictive Machine Learning, including subjects such as Data Mining, Neural Networks, Prediction Markets, Monte Carlo Simulation and Time Series Analysis. In parallel, he is a senior researcher in the Athens 2004 Olympic Games Global Impact project, which he shapes in co action with the International Olympic Committee.

Dr. Konstantinos KIRYTOPOULOS holds a Mechanical Engineering Diploma (Bachelor plus MSc equivalent) and a PhD on Project Risk Management. He is specialised in Industrial Management and is currently occupied as a Lecturer in the Aegean University (Greece), an external scientific associate of the National Technical University of Athens and as a freelance consultant. His main research interests are Project Management, Risk Management and Investment Analysis. He has participated in major risk management projects and has already published a significant amount of papers at International Conferences and Journals.

BUSINESS GAMING

PREDICTION MARKETS: ASPECTS ON TRANSFORMING A BUSINESS GAME INTO A VALUABLE INFORMATION AGGREGATION TOOL

Georgios G. Tziralis
Ilias P. Tatsiopoulos
Sector of Industrial Management and Operational Research
Mechanical Engineering School
National Technical University of Athens
Iroon Polytechniou 9, 15780, Athens
Greece
E-mail: {gtzi|itat}@central.ntua.gr

KEYWORDS
Prediction markets, forecasting, information aggregation

ABSTRACT

The importance of forecasting in transforming strategic objectives into decisions is apparent. Lately, traditional forecasting methods have been depicted as being inferior to Artificial Intelligence (AI) approaches (such as neural networks) which, by definition, are attempting to simulate the human decision making process and model the experts' knowledge.

However, the goal of identifying and accumulating the entire related knowledge and imitating the thinking processes of the ideal expert is hard to realise and might even be impossible to achieve. The depth of the correlated knowledge seems infinite and probably indeterminable, while the cognitive processes of the human mind remain a mystery. Therefore, both standard and modern approaches to forecasting are intrinsically problematic.

But, instead of trying to compensate the human expert's shortcomings by subtractively modeling their cognitive function and then performing the forecast, the decision maker could just let the experts make their forecasts and subsequently sum up the information in a way that confutes their inherent shortcomings. This paper advocates the usage of a market simulation game that uses expert forecasts and serves as an information aggregation tool producing as end result optimum valuable predictors.

THE PROBLEM IN RETROSPECT

Forecasting: The link between strategic thinking and decision making

The essential problem of management is to transform a company's strategic objectives into decisions and actions. The constantly increasing volatility of business ecosystems emphasises the critical importance of forecasting in decision making processes (Armstrong 1983, Waddell and Sohal 1994).

As Henri Fayol stated (Makridakis 1996) - already in 1916,

"*... and it is true that if foresight is not the whole of management at least it is an essential part of it.*"

Forecasting is defined as any statement about the future. Such statements may be well founded, or lack any sound basis; they may be accurate or inaccurate on any given occasion, or on average, precise or imprecise, model-based or informal. Forecasts are produced by methods as diverse as well-tested systems of hundreds of econometrically-estimated equations, through to methods which have scarcely any observable basis (Clements and Hendry 2001).

The vastness of the subject resembles the vagueness of its definition, a fact that is further proved by the existence of 35 English words defining the meaning of forecasting (One-look web dictionary, *http://www.onelook.com/?w=*%3Aforecasting&loc=cb*, accessed at 4/11/2005). The number of available forecasting techniques is even broader.

Traditional Techniques vs AI in Forecasting

During the extensive period of research into and assessment of business applications of various forecasting modeling techniques, a set of them has arisen as the most useful and of widespread applicability. Moving average, exponential smoothing, regression analysis, trend line analysis, decomposition analysis and ARIMA modeling are among the techniques that forecasting executives prefer, mainly due to their familiarity, satisfaction and usage (Mentzer and Kahn 1995).

Any forecasting model assumes that there exists an underlying – known or unknown – relationship between the input (the past values of the relevant variables) and the output (future value). In spite of their previously mentioned desirable characteristics, the ability of traditional forecasting methods to reproduce this underlying function is limited due to the complexity of the real system (Zhang et al 1998).

AI provides an attractive alternative to this problem. Artificial neural networks, mathematical models inspired by the organization and functioning of biological neurons and the human brain in particular, perform at least as well as common statistical methodologies (Hill et al 1994). Neural network modeling provides a data-driven, self-adaptive method that comprises a universal non-linear functional approximation and has an extensive ability to generalize (Zhang et al 1998). These characteristics make artificial neural networks an appealing approach to expert knowledge modeling.

Expert knowledge: Is modeling and inference feasible?

In the majority of cases, a specialist is unable to formulate all his knowledge and experience in an organised and transferable manner. Therefore, when trying to represent an expert's knowledge explicitly (e.g. by rules) the result seems confusing, indiscernible and rather inadequate. This sterile representation of knowledge is often not able to preserve and can even violate the original knowledge in such a way that the inference engine fails to draw the correct conclusions from the provided knowledge (Síma and Cervenka 2000). Moreover, human experts do not usually apply formal, logical and analytical skills to situations, rather they associate the new case with some old pattern to derive a solution. It should be taken into consideration though, that the simulations of even very simplified mathematical models of neural networks exhibit surprisingly 'intelligent' behavior, resembling human intelligence, e.g. the ability to learn new knowledge and to generalize previous experience (Peretto 2004). Therefore, the applicability of neural networks to expert knowledge modeling and forecasting emerges naturally.

However, the oversimplified modeling of human cognitive processes, whose complexity is vast and only partially understandable, remains an obstacle. The inability of fully extracting and explicitly representing the totality of the correlated variables derived from an expert completes the intrinsic problematic nature of neural network forecasting modeling.

Consequently, modeling does not seem as the best possible or even secure approach when attempting to compensate the undesirable, albeit human, characteristics of expert's opinions, namely biases and shortcomings (Armstrong and Brodie 1999). On the other hand, ignorance, inconsistency, irrelevance, inaccuracy, conflict and uncertainty are some of those problematic features (Ayyub 2001) that make expert judgments unattractive and of low worth.

So, how is it possible to use the whole implicit expert knowledge and simultaneously eliminate its inherent biases and shortcomings? The following part focuses on presenting a business game that serves as an optimum information aggregator by combining the expert's opinions in a way that negates their weaknesses.

PREDICTION MARKETS

A preamble

In 1945, Friedrick Hayek hypothesed that open markets efficiently and effectively facilitate the aggregation and transmission of information through prices (Hayek 1945). Twenty years later, Eugene Fama formulated the efficient market hypothesis, which states that an efficient market continuously reflects the sum of all available information about future events into security prices (Fama 1965). This implies that security prices reflect their true expected value and that no additional available information can be combined with efficient market prices to improve the market's forecast accuracy. In parallel, the economic theory of rational expectations, originating with Muth (Muth 1961) and most closely associated with Lucas (Lucas 1987), not only acknowledge the capacity of markets to aggregate information, but also their ability to convey information through the prices and volumes of the traded assets. More recently, Vernon Smith demonstrated that people behave rationally in experimental markets (Smith 2002) and was awarded the 2003 Nobel Prize in Economics for using experimental markets to prove and refine important theories about market behaviour.

In other words, a market functions by nature as an information aggregator that, under certain conditions, can produce the optimum prediction for the given information. So, why not organize a business game in the form of a virtual market so as to aggregate the agents' knowledge and information and extract optimized forecasts?

Definition and first applications

As 'prediction' are called the markets that run for the primary purpose of using the information content in market values in order to make predictions about specific future events (Berg and Rietz 2003). Their fundamental difference to commonly defined markets is therefore their scope of use as they consist a forecasting tool rather than a resource allocation mechanism. A prediction market can also serve as a decision support system by providing information about the current situation or by evaluating effects of decisions over time (Hanson 1999).

In literature, the term 'prediction markets' is not universally adopted. Alternative terms include decision markets, information markets, idea futures, forecasting markets, idea stocks, artificial markets, information aggregation mechanisms and virtual stock markets.

The longest known running set of prediction markets is the so-called Iowa Electronic Market, which established in 1988 and were designed to predict elections outcomes. The first business application however took place some years later. In 1997, Gerhard Ortner of the Technische Universität Wien developed and administered a prediction market used by Siemens to predict a large software project's completion date (Ortner 1997). In parallel, from 1996 to 1999, Kay-Yut Chen and Charles Plott of the California Institute of Technology administered prediction markets at Hewlett Packard in order to predict the sales volume of printers (Chen and Plott 2002).

A paradigm

A simple paradigm illuminates the conceptual use and operational principles of prediction markets. Suppose for example that the board of directors of a company needs reliable sales forecasts in order to reengineer the supply chain and minimize operational costs. For each product, the employees who have access to relevant information are given a virtual sum of money and access to the market. The shares of claims (stocks) that are traded in the market are directly connected to the height of sales volume of a given product. A sample claim might state for example that "Demand for *Product X* will fall between *y* and *z* units in the first quarter of the next year". If the claim proves to be true, then it pays €1, otherwise it pays nothing. Assume that sometime the claim trades at a price of *P* cents. That is the market, or the aggregated views of the employees, denotes that there is a *P*% likelihood that the claim will hold true. An employee who believes that there is at least a *P*% chance of it occurring would probably buy shares, while another one who believes the percentage to be smaller would sell the claim. The transaction volume is a useful indicator of the relative confidence that traders have at the current market value. Low volumes could indicate lack of significant dissent regarding the claim's value and consequently demonstrate that the market has reached a consensus concerning the current value of the claim (Schrieber 2004).

IMPLEMENTATION DIRECTIONS

Design Issues

The following provides a basic and synoptic framework of the key elements that should be taken into consideration during a prediction market's design.

Firstly, and foremost, the objective of the forecasts should be defined (Spann and Skierra 2003). The transformation of the forecasting goals into shares of claims should be accomplished in a way that the contracts are clear, easily understood and easily adjudicated (Wolfers and Zitzewitz 2004). Various types of claims structures and definitions are available (Schrieber 2004).

The second decision refers to the selection of market participants. Each employee who has access to information concerning the under study claims should be able to share his knowledge through the market's mechanism. Ideally, the available sum of money that each participant is able to trade in each share should be analogous to the extent and importance of the relevant knowledge that he possesses. In practice, all traders start with the same amount of money and the market, during the resource redistribution that takes place through its operation, decides the importance of each player and allocates its resources accordingly.

A decision of equal importance is the selection of trading mechanisms. The most widely used approaches are the continuous double-auction and the automated 'market-maker'. However, a dynamic pari-mutuel market has recently emerged as a more sophisticated and advantageous approach (Pennock 2004).

Another critical issue is the players' remuneration. Some participants may be sufficiently motivated to play simply out of an intrinsic employment of trading based on their knowledge. Others, however, may need some incentive to participation and truthful information revelation (Schrieber 2004, Wolfers and Zitzewitz 2004). The choice of such incentives may be crucial for the success of the market. The decision of whether to use monetary (for example real or play money) or non-monetary rewards (for example predetermined awards) is up to designer (Spann and Skierra 2003) and remains more an art than a science (Servan-Schreiber et al 2004).

Other considerations concern design of the financial market. Position and price limits (specific limits on portfolio selection, maximum and minimum prices for limit orders and quotes), trading hours (Spann and Skierra 2003) are also issues that the market designer has to include in his design.

Finally, these specifications have to be implemented in a user-friendly trading interface. The way transactional information is incorporated in the interface may have significant impact on trading activity and information sharing. Therefore, the amount of information about each employee's transactions that is published has to be thoroughly determined.

Benefits

The design process of prediction markets consists of a laborious multi-step process, which demands that various decisions are taken and lacks generally applicable guidelines. The benefits, however, derived from its use seem to compensate the development and operating costs.

Some of the most difficult steps in a typical forecasting application are to mine, namely to collect, merge and clean relevant data from human experts. The market games appear to serve by definition as a mechanism to negate these objections and handle effectually all these steps; this consist definitely a chief advantage.

The forecast accuracy of an efficient prediction market is, under the condition of efficiency, optimum. In practice, prediction markets usually tend to perform at least as well as the single best individual, without requiring knowledge of who that individual is in advance (Surowiecki 2004). The bias removal and constant reallocation of each employee's forecasting weight, in terms of portfolio value, that take place in such an information aggregation mechanism, are facts that also evidence the width of its accuracy.

Another major advantage of the implementation of such a market is its dynamic nature. The market continuously reflects the best possible forecast through its innate ability to constantly represent and immediately respond to new information (Hanson 1999). This characteristic is absent from the vast majority traditional approaches to the forecasting problem.

The enriched insight that a decision maker obtains from the price fluctuations is an additional benefit. A prediction market's forecast provides far more information than a point forecast. For example, the price fluctuation range is indicative of the amount of risk embedded in the claim, while the impact that the introduction of new information has on a forecast serves as a kind of sensitivity analysis.

The efficient adaptability of a prediction market also represents an advantage. In contrast to traditional approaches, the operation of the prediction market is not affected by possible changes in types and sources of information or even the number of inputs or participants. Prediction markets are by nature able to transform unlimited amounts of timely and locally dispread qualitative information into accurate quantitative forecasts about the future.

The previous benefits of predictions markets' use regard the extracted forecasts alone. However, the perhaps most important outcome of the implementation of such a business tool is wider and intangible. The motivation to share information, the weakening of personal and political impulse, the encouragement of honest assessment and the careful evaluation of each employee's decision making processes are attitudes that are cultivated, reinforced and taken advantage of by a prediction market. Such attitudes unambiguously boost the transformation of an enterprise's culture towards a more participial, creative and reliable role of its personnel, no matter how valuable the predictions are.

When to use

It is now apparent that the benefits of prediction markets are substantial and sui generis. While the tool in general appears as very attractive, its design and operational cost still remain considerable. In order to maintain its suitability over traditional forecasting approaches, additional factors have to be considered.

Factors, such as importance, quality, acceptance (positive influence) and effort (negative influence) (Schrieber 2004), heavily influence the value of a forecast. Importance refers to the extent of the impact that the forecast has on specific business decisions. Quality is directly related to the amount of information that is contained and revealed by the forecast. Acceptance declares the degree of confidence that the decision makers have in the forecast, while effort measures the amount of resources (people, time and money) spent in formulating the forecast.

The quality of a prediction market's forecast is high, also high is its acceptance, provided that its function and benefits are communicated in a correct way. However, the effort invested in it is usually of much higher level than that of traditional methods. As a consequence, the value of a market's forecast is significantly higher than those of more widespread used methods, only in the case that the importance of its outcomes is major and influential. In other words, the use of prediction markets' approach is lucrative and highly recommended only for the forecasting objectives that have a major impact on the business decisions.

CONCLUSION

The role of forecasting in transforming strategic objectives into decisions is crucially important. Traditional forecasting approaches as well as modern AI techniques have been proven to perform poorly in modeling and aggregating of experts' knowledge.

This paper proposes an opposite and novel point of view by introducing the use of prediction markets. Such business games function as an optimum information aggregator and are inherently able to transform the totality of expert's qualitative knowledge and information into accurate quantitative forecasts about the future.

Several design issues are discussed, such as the definition of the forecast's objectives, their transformation into shares of claims, selection of employees and participation incentives, etc. The uniqueness of the tool is further documented in terms of its accuracy, dynamic nature, adaptability and the enriched insight it provides for the forecasting problem in question.

This paper provides the context for the rationalistic adaptation, design and implementation of a valuable business forecasting tool. The application of this tool for the extraction of expert' aggregated estimations for the quantified assessment of Athens 2004 Olympic

Games impact is presently under consideration by the authors' team.

REFERENCES

Armstrong S. 1983, “Strategic Planning and Forecasting Fundamentals”. In Kenneth J. Albert, editor, *The Strategic Management Handbook*, pages 2-1 to 2-32. McGraw Hill New York, 1983.

Armstrong S. and Brodie R. 1999, “Forecasting for Marketing”. In Graham J. Hooley and Michael K. Hussey, editors, *Quantitative Methods in Marketing*, pages 92-119, International Thompson Business Press, London, 1999.

Ayyub B. 2001, “Elicitation of Expert Opinions for Uncertainty and Risks”, CRC Press, 2001.

Berg J. and Rietz T. 2003, “Prediction Markets as Decision Support Systems”, *Information Systems Frontiers*, 5:1, pages 79-93.

Chen K. and Plott C. 2002, "Information Aggregation Mechanisms: Concept, Design and Field Implementation for a Sales Forecasting Problem." California Institute of Technology, Social Science Working Paper 1131.

Clements M. and Hendry D. 2001, “An Overview of Economic Forecasting”. In Michael P. Clements and David F. Hendry, editors, *A Companion to Economic Forecasting*, Blackwell Publishing, 2001.

Fama E. 1965, “Behavior of Stock Market Prices”, *Journal of Business*, 38, pages 34-105.

Hanson R. 1999, “Decision Markets”, *IEEE Intelligent Systems*, 14:3, pages 16-19.

Hayek F. 1945, "The use of knowledge in society", *American Economic Review*, 35, pages 519-530.

Hill T., Marquez L., O'Connor M. and Remus W. 1994, “Artificial neural network models for forecasting and decision making”, *International Journal of Forecasting*, 10, pages 5-15.

Lucas R. 1987, “Models in business cycles”, London: Blackwell

Makridakis S. 1996, “Forecasting: Its role and value for planning and strategy”, *International Journal of Forecasting*, 12, pages 513-537.

Mentzer J. and Kahn K. 1995, “Forecasting technique familiarity, satisfaction, usage, and application”, *Journal of Forecasting*, 14, pages 465-476.

Muth J. 1961, “Rational expectations and the theory of price movements”, *Econometrica*, 29:6, pages 315-335.

Ortner G. 1997, “Forecasting Markets – An Industrial Application, Part I”, Working paper, version 0.42, Vienna, Austria, University of Technology Vienna, Department of Managerial Economics & Industrial Organization.

Pennock D. 2004, “A Dynamic pari-mutuel market for hedging, wagering, and information aggregation”. In *Proceedings of ACM Conference on Electronic Commerce*, pages 170-179, New York, May 2004.

Peretto P. 2004, “An Introduction to Modelling of Neural Networks”, Cambridge University Press, 2004.

Schrieber J. 2004, “The Application of Prediction Markets to Business”, Massachusetts Institute of Technology, submitted to the Engineering Systems Division for the Degree of Master of Engineering in Logistics.

Servan-Schreiber E., Wolfers J., Pennock D. and Galebach B. 2004, “Prediction Markets: Does Money Matter?”, *Electronic Markets*, 14:3, pages 243-251.

Síma J. and Cervenka J. 2000, "Neural Knowledge Processing in Expert Systems". In Ian Cloete and Jacek M. Zurada, editors, *Knowledge-Based Neurocomputing*, MIT Press, 2000.

Smith V. 2002, “Constructivist and Ecological Rationality in Economics”, Nobel Prize Lecture, December 8, 2002.

Spann M. and Skierra B. 2003, “Internet-Based Virtual Stock Markets for Business Forecasting”, *Management Science*, 49:10, pages 1310-1326.

Surowiecki J. 2004, “The Wisdom of Crowds”, Doubleday, 2004.

Waddell D. and Sohal A. 1994, “Forecasting: The key to Managerial Decision Making”, *Management Decision*, 32:1, pages 41-49.

Wolfers J. and Zitzewitz E. 2004, “Prediction Markets”, *Journal of Economic Perspectives*, 18:2, pages 107-126.

Zhang G., Patuwo E. and Hu M. 1998 “Forecasting with artificial neural networks: The state of the art”, *International Journal of Forecasting*, 14, pages 35-62.

BIOGRAPHIE

GEORGIOS G. TZIRALIS was born in Athens in 1982. He has obtained his Mechanical Engineering Diploma (Bachelor plus MSc equivalent) from the National Technical University of Athens in 2004 and is now a PhD Candidate in the Industrial Management and Operational Research Sector of the NTUA's Mechanical Engineering School. His research interests extend in the area of Prediction Markets and Computational Methods for Predictive Machine Learning, including subjects such as Data Mining, Neural Networks, Monte Carlo Simulation and Time Series Analysis. In parallel, he is a senior researcher in the Athens 2004 Olympic Games Global Impact project, which he shapes in co action with the International Olympic Committee.

ILIAS P. TATSIOPOULOS is a Professor in Production Planning and Control and Manufacturing Information Systems in Mechanical Engineering School of National Technical University of Athens. He has a Degree in Mechanical and Electrical Engineering with emphasis in industrial engineering from NTUA and a Ph.D. in Operations Management and O.R. from the University of Lancaster, England. His academic interests are computer-aided production management, logistics, forecasting and the design of production systems.

ACKNOWLEDGMENTS

This research work is part of a project funded by the Hellenic Ministry of Culture.

A MONTE CARLO ASSESSMENT OF SUSPENSE IN SPORTS AND GAMES

Emmanuel Dion
Marketing Professor at Audencia, Nantes School of Management
Audencia Nantes School of Management
8 route de la Jonelière F-44300 NANTES
E-mail: edion@audencia.com

KEYWORDS

Simulation, Scoring systems, Games, Sports, Suspense, Monte Carlo.

ABSTRACT

This paper analyses how the scoring systems actually and potentially used in a variety of sports and games affect the suspense they can generate. The experimental method relies on several Monte Carlo simulations applied on a comparative basis to basketball, table tennis and tennis. Results show that simple systems such as basketball scoring tend to generate limited uncertainty whereas stratified systems such as tennis scoring favour a higher level of unpredictability, hence suspense.

INTRODUCTION

Little if any research has been undertaken until now about how scoring systems affect the progress types of sports matches, along with their propensity to generate suspense. This fact is all the more surprising that huge issues relating to sports fans' interest and media coverage are at stake, resulting in potentially major financial consequences.

This point would be of little practical value if the sports scoring systems were absolutely fixed as part of the games' universal history. This assumption is reinforced by the fact that some scoring systems have their roots deep in their sport's history. For instance, the 15 – 30 – 40 – game progress in tennis games is said –without proof- to come from the presence of clocks next to some tennis courts (*jeu de paume*) halls in France centuries ago. But historical fixity does not hold if we observe that:

- volleyball has fundamentally changed its scoring system in 2000. With the former system, points could be scored only when a team had the service - side-out scoring- and all sets went up to only 15 points versus 25 points now.
- rugby has two major variations, along which even the number of players involved in the game changes. In 2005 (vs. the original 1905 rules), a try is worth 5 points (vs. 3), a conversion 2 (unchanged), a penalty 3 (unchanged) , a drop goal 3 (vs. 4).
- basketball is played with different rules in Europe and in the United States of America
- tennis has changed several times its scoring system during its history, the last change dealing with doubles: shortening of the sets, disappearance of the advantage rule, emergence of a super tie-breaker that replaces the decider. Not to mention the differences between professional tournaments, some matches being played in three or five sets, with or without a tie-breaker in the decider.
- Etc.

All those considerations are related to major sports that have a dominant position among entertainment activities offered to man and a strong history that tends to make them change-resistant. By contrast it is therefore easy to imagine how flexible scoring systems can be in general, which might sound counter-intuitive at first, because scoring tends to be unconsciously admitted as the most fundamental set of rules of a game, since it defines exactly the goal that players should aim at.

In fans' perception it is hard to define which comes first: the conscience of the score, or the attention to style. In fact, both seem to be intricately intertwined, as is being exemplified by the following excerpt of an internet chat (see Figure 1). Thousands of chats' excerpts would show the same thing as these lines that were written during the final match of the Tennis Masters Cup held in Shanghai in November 2005: suspense, tension, attention (and hence economic potential) mainly comes from the score line progress.

"Federer_Express_2000
20.11.2005 | 12:44: Its not looking good, this shouldnt have goned to 4, damn it!

Sally-Suzanne
20.11.2005 | 12:45: 15-0 Roger! [...]

mary91
20.11.2005 | 12:45: Oh,my God! I don't watdh Roger's match?but I will die.............soon

Melanija
20.11.2005 | 12:46: Let's all stay positive. Let's all believe he will make it in the end.

Sally-Suzanne
20.11.2005 | 12:46: beautiful shot from Rog! 30-0!

Sally-Suzanne
20.11.2005 | 12:47: just long from Roger . deuce now!

Federer_Express_2004
20.11.2005 | 12:47: So far Nalbandian has not given up nothing to Roger, no breakpoints this set.

Sally-Suzanne:
20.11.2005 | 12:47: i know its 4th set, whats the game score?"

Figure 1 – Excerpt from **http://www.rogerfederer.com**
(typing mistakes have intentionally not been corrected)

We can easily see in this example how comments about the game and style are mixed in depth with comments about the score. In this perspective, the scoring does not appear any more as a pure relative social construct, but as an intangible state of the world. In other words, it becomes in fans' minds part of a first-degree apprehension of facts.

This should not mislead sports' organizers and managers : scoring systems are not intangible, even though they have to keep this imaginary property in fans' minds. Scoring systems are efficient tools that contribute to define the potential of attention or interest that might arise from sports shows. Undertaking scientific research on the topic might therefore be of considerable economic value.

METHODOLOGY

The experiment relies on a Monte Carlo based simulation that reproduces the progress of scoreboards based on various scoring systems: to be able to compare scoring systems on a coherent basis, we have chosen to concentrate on major sports that:

- (1) rely on a succession of *n* independent phases (rallies in tennis, table tennis, or volleyball, offensive action in basketball, touches in fencing) that lead to score a large number of times throughout a match
- (2) with *n* usually ranging between 50 and 200 (*n* is never fixed in advance to preserve suspense although a minimum value can be set).

Note that many sports could be adapted to make this scoring system suitable for them without changing their fundamentals: shooting, archery, bowling, golf, etc. However, condition (2) excludes all sports and games for which scoring is rare (such as ice hockey or soccer) or unique (such as races or most combat sports).

For the purpose of clarity, we have chosen to concentrate on three sports that illustrate two extreme and one medium scoring options:

- **basketball** illustrates a simple method in which points are simply added from the start to the end of a match (shots may bring 1, 2 or 3 points but this does not change the core linearity of the method), the goal being to score more points than the other team. There is only one level of scoring.
- **tennis** illustrates a complex method in which points have to be capitalised in games and games have to be capitalised in sets to decide who is the winner (3 levels of scoring). Optionally, when games go to deuce, another minor level of capitalisation exists with the advantage rule. In this experiment we have only investigated the most current type that consists of three-set matches (the first player to win two sets wins the match).
- **table tennis** illustrates a medium option with two levels of scoring: points and sets.

A computerised Monte Carlo simulation tool has been developed (using Excel ® and VBA ®) and used to reproduce the repeat of a very large number of matches relying on those three scoring systems and examine results from a statistical perspective.

Various sets of probabilities have been used to test the model.

Set A ("Equal") – Set A gave each player (or team) a constant and equal point win probability of 0.5. This allows to measure the degree of suspense generated in matches between two players (or teams) of equal strength. Set A has been run 10,000 times to ensure a good level of histogram smoothing.

Set B (p) ("Constant") – Set B gave player 1 a constant point win probability of p (and hence player 2 a constant point win probability of 1-p), p being different from 0.5. This allows to measure the degree of suspense generated in matches between two players (or teams) of unequal strength, the difference between the two being integrated in the model as a parameter. Set B has been run 1,000 times for the following values of p: 0.4, 0.425, 0.45, 0.475, 0.5. For values smaller than 0.4, matches were so unbalanced that conclusions would not have been significant anyway.

Set C (p, p', t) ("Turnaround") – Set C gives player 1 a point win probability that swaps from p to p' at time t during the match. This allows to measure the capacity of the scoreboard to reflect the change in the momentum of matches that tend to turn around at a given point due to the rise of a player and/or the fall of his competitor.

The device had to be slightly improved for basketball simulation, since:

- winning shots have a value of one to three points depending on the circumstances (free throws, shots from inside or outside the 3 –point line).
- the probability of one team scoring is dependent on their being on the offensive side, which in turn results from the last occurred event (opponents scored or missed).

For this reason, a special simulator has been built that reproduces the full sequence of events occurring in a match, second after second. The simulator consists of 2880 lines of events (corresponding to the 2880 seconds of a match) for which any of the following events can occur, with an appropriate probability distribution:

- Team 1 is in control of the ball and nothing happens
- Team 1 is in control of the ball and scores 1 point. Team 2 takes control of the ball.
- Team 1 is in control of the ball and scores 2 points. Team 2 takes control of the ball.
- Team 1 is in control of the ball and scores 3 points. Team 2 takes control of the ball.
- Team 1 loses the control of the ball (the ball is intercepted, a shot is missed and the opposite team wins the rebound, or there is a foul)
- All of the above for team 2

The set of probabilities used for each line depends upon the set of probabilities used for the previous lines, thus solving the non-independency-of-events problem. All probabilities have been computed from empirical data collected from NBA statistics for the start of season 2005, widely available on the internet (see table 1. Source: Atlantic Division of Eastern Conference, start of season until 14. December, website used: http://www.allbasketball.com/nba/05-06/stat_cum/index.php). In case of simulations with sets B or C, the sets of probabilities used were changed to care for the impact of p and p'.

Table 1 – Empirical data used for probability sets definition

	Side change	Score 2 points	Score 1 point	Score 3 points	Lost rebound when offensive	Lost ball control	Matches played
Boston Celtics	950	613	399	104	606	344	20
New Jersey Nets	922	553	409	98	625	297	19
NY Knicks	955	580	409	59	588	367	19
Philadelphia 76ers	1000	699	489	91	712	288	21
Toronto Raptors	1000	626	409	120	708	292	21
Total	4827	3071	2115	472	3239	1588	100
Average occuring time in seconds	60	94	136	610	89	181	

For all three sports investigated, a number of indicators have been defined to measure suspense.

- **Length variability ("Length")**
- **Pivot point ("Pivot")**: position of the first time in the match when one player (or team) takes the lead and keeps it until the end of the match. This measure is expressed in terms of percentage of progress in the match (0% being the start, 100% the end).
- **Number of equalities ("Equalities")**: Number of times when the leader changes throughout the match, including ties.
- **Immediacy ("Immediacies")**: propensity of the match to come close to a possible end.

For all those indicators, no fixed optima can be determined other than by undertaking some kind of psychological experimentation. However, it is easy to understand that the best quality of suspense is generated for:

- Undetermined length, even though the match should always seem to progress.
- Usually late but fundamentally undetermined pivot point for set A. Appropriate dependency over p, p' and t for sets B and C.
- Average number of equalities in all cases, with a dependence on p-p' for set C
- Potentially long but fundamentally undetermined immediacies especially for set A and to a lesser extent for set C.

If the score progresses in a chaotic way, with the outcome of the match being unpredictable until the very last points, the attendance might feel perplexed and reluctant to focus their attention during the first and middle stages of the match. But on the contrary if the score progresses along a too predictable line, the attendance might get bored and lose their attention well before the end. A good scoring system is a scoring system that allows some limited turnarounds on the scoreboard until the end of the match, with a turnaround occurrence probability decreasing over time but always remaining dependent upon a possible shift in players' momentum. Only this combination can allow fans to write in their minds, while attending a match, an imaginary story that will keep their attention and memories open.

RESULTS

In this paper we will only discuss the results observed for sets A and B, and leave set C findings for further research and discussion.

A – Length variability

Note : In this section, length means number of scoring events (points in tennis and table tennis, winning shots in basketball), and not duration.

Tennis, table tennis and basketball have relatively different schemes in terms of length for set A. Statistical indicators and probability distributions are displayed on table 2 and figure 2. Although the length means are similar, the standard deviations vary considerably, basketball being much more predictable (and therefore compatible with television standards) than table tennis and tennis. The table tennis curve has three local maxima corresponding to matches that go to 3, 4 or 5 sets (the last two being slightly more frequent than the first). The same effect exists, but very much attenuated, for tennis matches that can go to 2 or 3 sets.

Table 2 – Length statistics for set A

	Tennis	**Table tennis**	**Basketball**
Mean	164,60	153,62	189,01
Std.	42,51	30,05	17,83

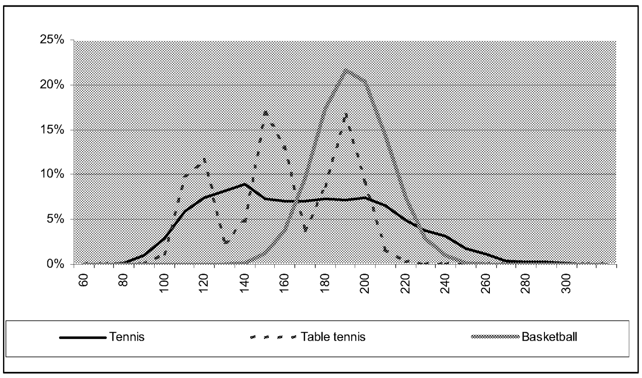


Figure 2 – Length distribution for set A

When point win probabilities change (see set B tables and figures, numerated from 3a to 3c for details), we can observe that length means and standard deviation decrease simultaneously for tennis and table tennis while being almost unchanged for basketball (technical note: values for set B and p=0.5 are different from those of set A because they have been computed again on a smaller basis to ensure compatibility with other values of p).

Table 3a – Length statistics for set B (Tennis)

	p = 0,5	p = 0,475	p = 0,45	p = 0,425	P = 0,4
Mean	163,22	158,67	142,04	121,67	109,11
Std	42,36	43,02	40,91	31,49	27,23

Table 3b – Length statistics for set B (Table tennis)

	p = 0,5	p = 0,475	p = 0,45	p = 0,425	p = 0,4
Mean	153,20	148,21	137,77	124,79	113,81
Std	30,77	30,87	30,11	27,28	21,44

Table 3c – Length statistics for set B (Basketball)

	p = 0,5	p = 0,475	p = 0,45	p = 0,425	p = 0,4
Mean	188,25	188,37	190,62	190,62	192,66
Std	17,84	16,75	18,23	17,16	17,49

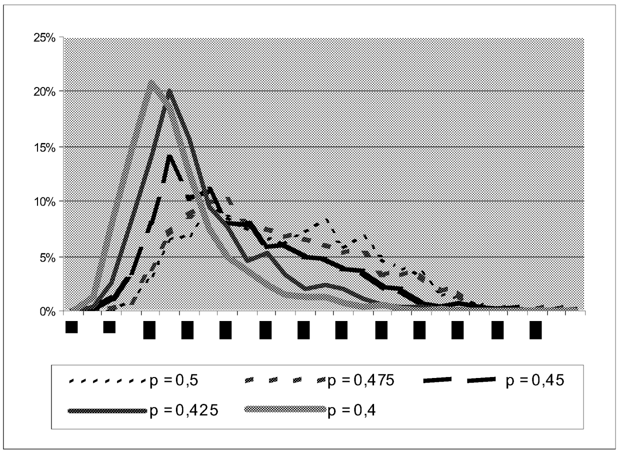


Figure 3a – Length distribution for set B (Tennis)

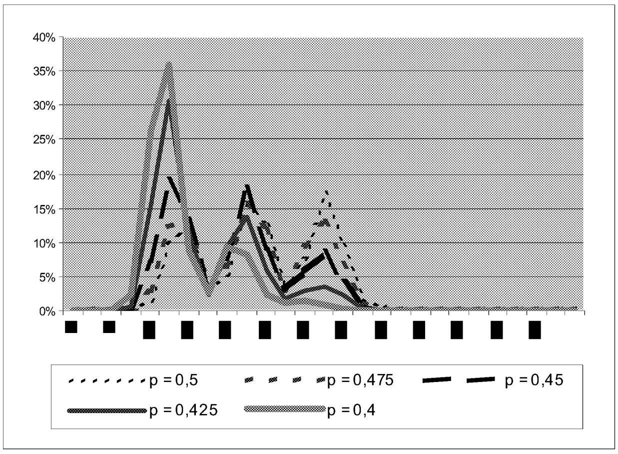


Figure 3b – Length distribution for set B (Table tennis)

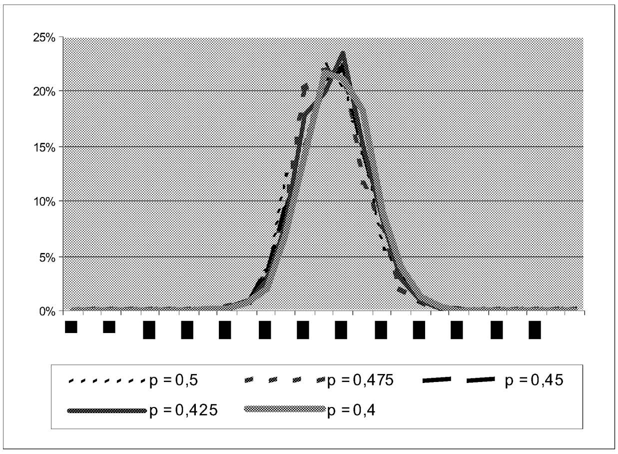


Figure 3c – Length distribution for set B (Basketball)

B – Pivot points distributions

In all three sports investigated, matches tend to pivot anytime when play is under way, with higher probabilities in the end though. Basketball tends to choose its winner right from the start slightly more often than others (see figure 4).

Globally, the basketball distribution curve in much smoother, and seems to show that the winner will appear most likely close to the beginning or the end of matches, while the other sports' curves look more chaotic. For table tennis, we can observe the three aforementioned peaks corresponding to the three possible number of sets. For tennis there is a slight alternance of relatively high or low probabilities of decision, the highest coming right at the end. Tennis is the sport for which in case of equal strength players, matches are decided most often during the last 10% of play.

Table 4 – Pivot points statistics for set A

	Tennis	Table tennis	Basketball
Mean	57%	56%	51%
Std	33%	34%	35%

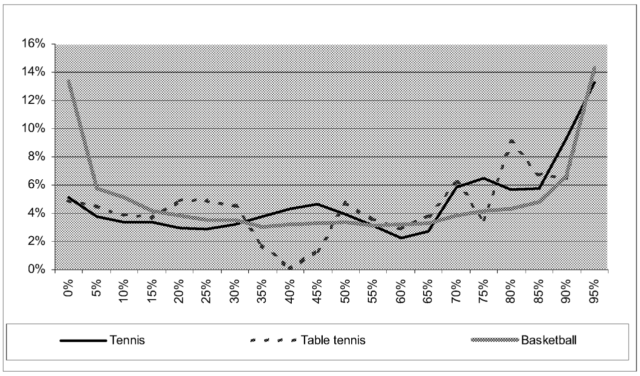


Figure 4 – Pivot points distribution for set A

The effects mentioned for set A remain observable for set B ans various values of p (see tables and figures 5a to 5c). Quite obviously, pivot points tend to appear all the more sooner in the matches that the point win probabilities is unequal between players.

Note that for some reasons, table tennis matches almost never turn around between 40% and 45% of play.

Table 5a – Pivot statistics for set B (Tennis)

	p = 0,5	p = 0,475	P = 0,45	p = 0,425	p = 0,4
Mean	0,56	0,52	0,42	0,29	0,22
Std	0,33	0,34	0,32	0,28	0,24

Table 5b – Pivot statistics for set B (Table tennis)

	P = 0,5	p = 0,475	P = 0,45	p = 0,425	p = 0,4
Mean	0,56	0,51	0,42	0,30	0,21
Std	0,34	0,35	0,33	0,30	0,24

Table 5c – Pivot statistics for set B (Basketball)

	p = 0,5	p = 0,475	P = 0,45	p = 0,425	p = 0,4
Mean	0,49	0,45	0,33	0,22	0,13
Std	0,35	0,35	0,31	0,25	0,17

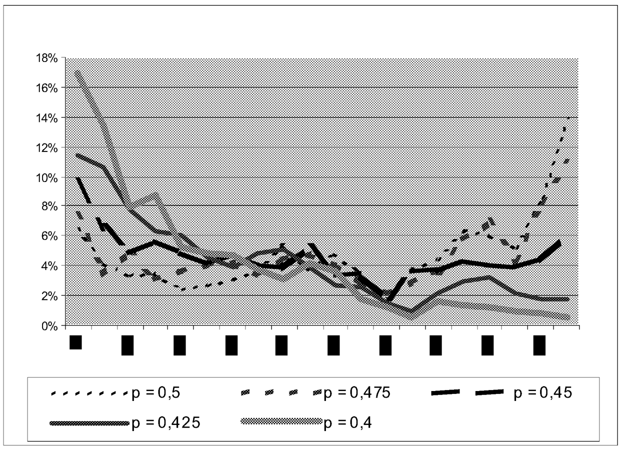


Figure 5a – Pivot distribution for set B (Tennis)

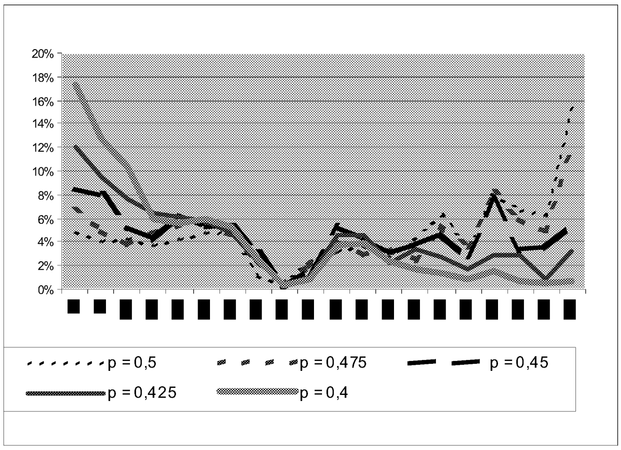


Figure 5b – Pivot distribution for set B (Table tennis)

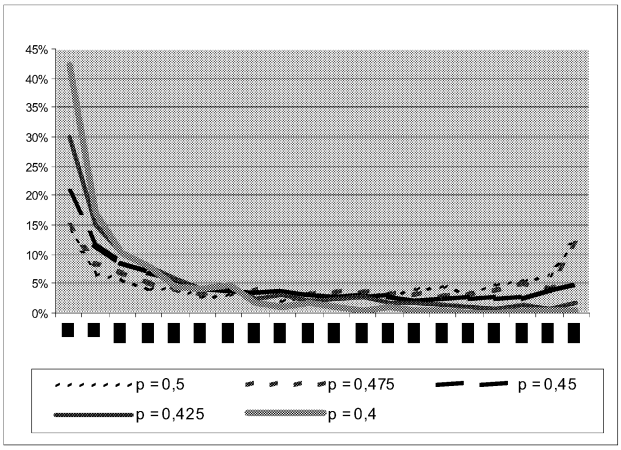


Figure 5c – Pivot distribution for set B (Basketball)

C – Equalities distributions

The equalities distribution show remarkable convergence between the three sports. Means, standard deviations and more noteworthily distribution curves look very much alike for set A (see table 6 and figure 6).

Table 6 – Equalities statistics for set A

	Tennis	Table tennis	Basketball
Mean	9,68	8,82	10,71
Std	6,91	6,14	7,57

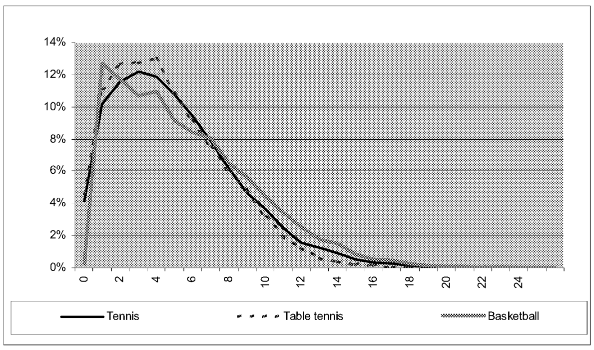


Figure 6 – Equalities distribution for set A

Aside from random variations due to the limited number of simulated matches, there is not much to comment about set B figures. Basketball, table tennis and tennis do not seem to vary significantly from one another on these criteria.

Table 7a – Equalities statistics for set B (Tennis)

	p = 0,5	p = 0,475	P = 0,45	p = 0,425	p = 0,4
Mean	9,58	8,97	7,33	5,21	4,18
Std	6,83	6,84	6,16	4,82	3,97

Table 7b – Equalities statistics for set B (Table tennis)

	p = 0,5	p = 0,475	p = 0,45	p = 0,425	p = 0,4
Mean	8,89	8,31	6,54	4,99	3,94
Std	6,17	6,25	5,45	4,73	4,10

Table 7c – Equalities statistics for set B (Basketball)

	p = 0,5	p = 0,475	p = 0,45	p = 0,425	p = 0,4
Mean	10,39	9,83	8,29	6,47	5,07
Std	7,43	6,98	6,59	5,61	4,43

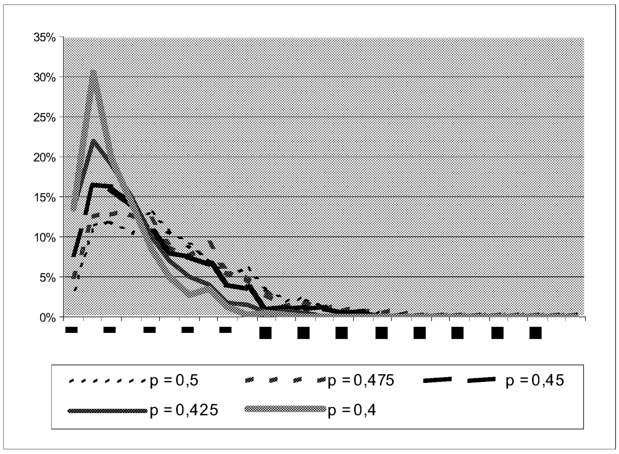


Figure 7a – Equalities distribution for set B (Tennis)

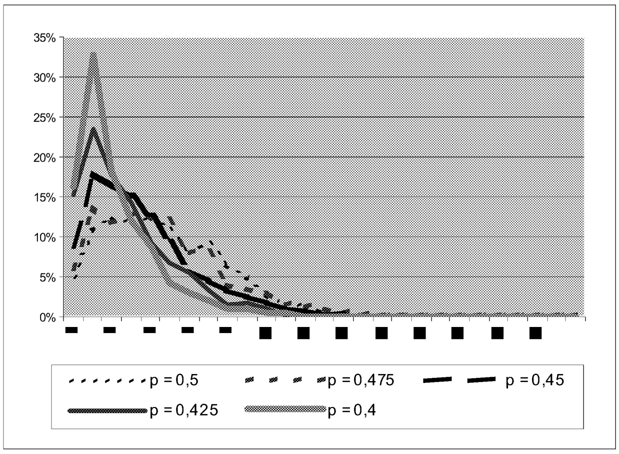


Figure 7b – Equalities distribution for set B (Table tennis)

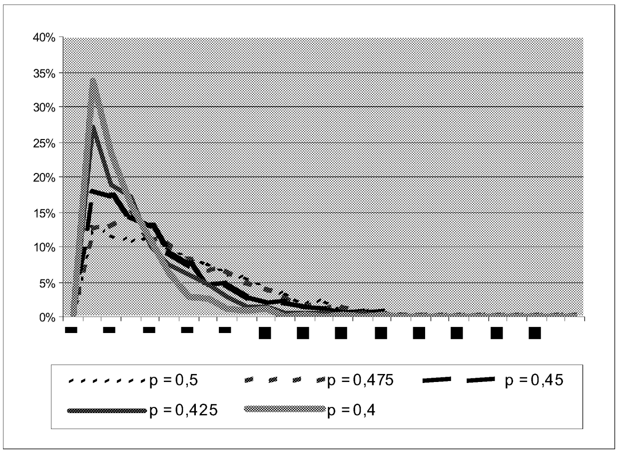


Figure 7c – Equalities distribution for set B (Basketball)

D – Immediacies distributions

Immediacies is a special indicator that need not be determined for basketball for in this case it is constant: the end of the match always happens at the official, known and predictable time.

As for tennis and table tennis, a match can stop more or less abruptly. In our Monte Carlo simulations, the shortest tennis match went to a total of 66 points whereas the longest lasted for 315 points, 4.8 times more (respectively 92, 223 and 2.4 for table tennis).

Simultaneously, immediacies seem a bit more frequent (and also more variable) in tennis than in table tennis. The mean values of 15.78 and 13.53 for set A (see table 8) should be compared to the minimum value of 5 (characteristic of the situation when the leader closes out the match as soon as he can).

These trends persist with the use of set B parameters. Tables and figures 9a and 9b show that even for an unequally balanced confrontation, a minimum level of suspense remains in that immediacies tend very slowly towards 5 (especially for tennis), resulting in a situation where anytime in the match, fans should not completely lose their hopes or attention.

Table 8 – Immediacies statistics for set A

	Tennis	Table tennis	Basketball
Mean	15,78	13,53	-
Std	10,73	8,02	-

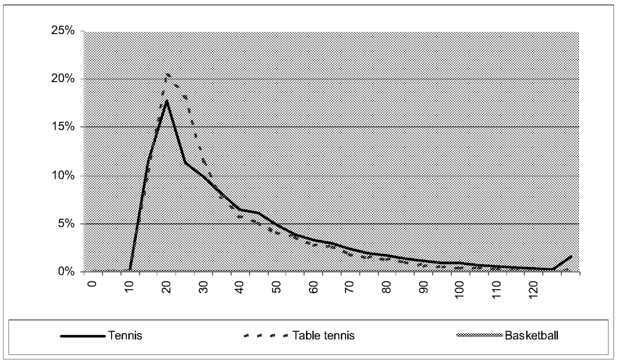


Figure 8 – Immediacies distribution for set A

Table 9a – Immediacies statistics for set B (Tennis)

	p = 0,5	p = 0,475	p = 0,45	p = 0,425	p = 0,4
Mean	15,37	15,61	13,54	11,58	10,41
Std	10,44	11,32	9,35	6,87	5,93

Table 9b – Immediacies statistics for set B (Table tennis)

	p = 0,5	p = 0,475	p = 0,45	p = 0,425	p = 0,4
Mean	13,62	12,56	11,16	10,05	9,00
Std	8,13	7,66	6,59	5,31	3,74

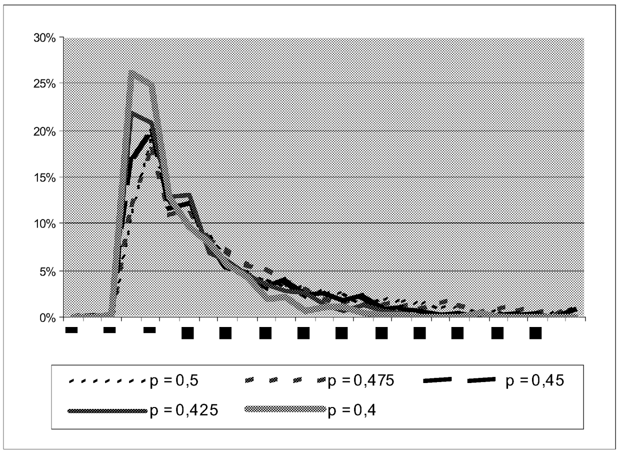


Figure 9a – Immediacies distribution for set B (Tennis)

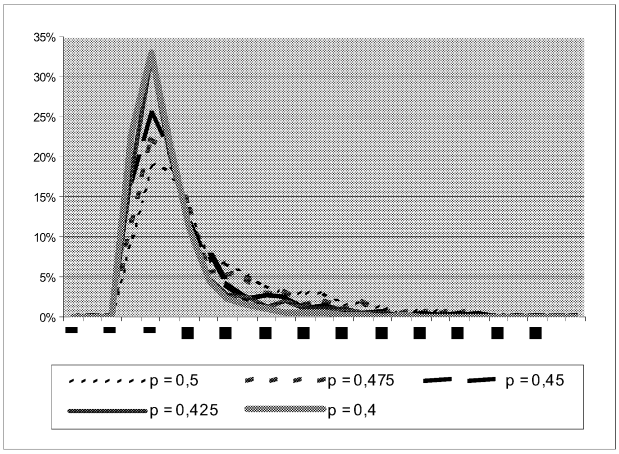


Figure 9b – Immediacies distribution for set B (Table tennis)

Conclusion

The conclusion we can draw from this series of experiments is that whereas basketball is more formatted, hence more suited for television programs, unpredictability in terms of length, capacity to turn around, and to come close to a possible end without necessarily actually closing are higher for table tennis and especially tennis.

This conclusion goes beyond a possible application to sports. It can lead to the development of a wider reflection about how suspense arises, not only in sports, but also in games, and especially in video games.

One of the obvious limits of this series of experiments is that it neglects some important factors such as:

- The advantage of serve in tennis or table tennis
- The feedback effect of the scoreboard on points winning probabilities (experience shows that better players tend to play well on important points)
- The extra-time situation in basketball
- Plus a number of psychological or physiological factors interfering with the rules in various sports.

In the future, we intend to present the results observed on the same measures operated on simulations run with parameters from set C, in order to investigate the capacity of score systems to reflect possible changes in matches momentums. After that, the next step could be either to replicate the research on a wider basis to assess the validity of the method and consolidate the findings, or to refine the simulating tools in order to take into account all the aforementioned factors that should be implemented into the model.

REFERENCES

Bennett, D. J. (1998), *Randomness*. Cambrige, Harvard University Press.

Fensch, T. (1995), *Sports Writing Handbook* 2nd Ed, Lea's Communication.

Kelly, G.A. (1955). *The Psychology of Personal Constructs*, New York, Norton.

Mancuso, J.C., J. Adams-Weber (1982). *Anticipation as a constructive process*, in .C. Mancuso & J. Adams-Weber (Eds.) The Construing Person, New York, Praeger, pp. 8-32.

Mandelbaum M. (2004). *The Meaning of Sports: Why Americans Watch Baseball, Football, and Basketball and What They See When They Do*, PublicAffairs.

Nadin, Mihai (1991). *Mind—Anticipation and Chaos*. Stuttgart: Belser Presse. (The text can be read in its entirety on the Web at http://www.networld.it/oikos/naminds1.htm.)

Nadin, Mihai (2000). *Anticipation—A Spooky Computation*. http://www.code.uni-wuppertal.de/de/computational_design/who/nadin/lectures/anticipa.html

Pitts, B. G., Stotlar, D. K. (2002). *Fundamentals of Sport Marketing*, Sport Management Library.

Taleb, N. N. (2004). *Fooled by Randomness*, Thomson Texere.

DISCRETE EVENT SIMULATION

QUEUEING NETWORKS IN EQUILIBRIUM AND MARKOV CHAINS: NUMERICAL SOLUTION METHODS

Dimitar Radev
University of Rousse
8 Studentska Street
7017 Rousse, Bulgaria
E-mail: dradev@abv.bg

Izabella Lokshina
SUNY Oneonta
324B Netzer Administration Bldg.
Oneonta, NY 13820, USA
E-mail: lokshiiv@oneonta.edu

Vladimir Denchev
University of Rousse
8 Studentska Street
7017 Rousse, Bulgaria
E-mail: vndenchev@abv.bg

KEYWORDS
Mobile communication networks, Markov chains, performance evaluation, steady state analysis, numerical solutions.

ABSTRACT

This paper is devoted to research and development of the computational procedures to obtain the steady-state probability vectors in the queueing networks. The numerical methods for the Markov chains performance evaluation and analysis, such as direct, iterative and aggregative methods, are considered and compared. The enhanced computational procedures, developed with use of Gaussian elimination direct approach, iterative power approach, Courtois's approximation method and Takahashi's iterative method are implemented and explored. Numerical examples for the steady state solutions based on the proposed algorithms are demonstrated; and comparative analysis of the obtained results is made.

INTRODUCTION

Queueing networks, which consist of several service stations, are usually used in modeling to represent the structures of various systems with large number of resources, for instance, computer and communication systems. In queueing networks at least two service stations are connected to each other. The stations, i.e., the nodes in the queueing networks, represent resources in the real systems. In general, jobs can be transferred between any two nodes of the networks; in particular, a job can be directly returned to the node it has just left. The queueing networks are *open* when jobs can enter the network from and leave the network to the outside. A queueing network is *closed* when jobs can neither enter nor leave the network. The number of jobs in closed network is constant. Performance measurements of the queueing networks could be obtained while computing the state probabilities. Creating the computational procedures for obtaining the steady-state probabilities for all possible states of the networks is a central issue in the queueing theory. The means for many other important network performance measurements could be easily obtained after having these algorithms developed. The Markov processes provide flexible, powerful, and efficient means for evaluation and analysis of dynamic systems. Performance and reliability measures of the queueing networks can be derived and evaluated with the steady-state analysis of Discrete-Time Markov Chains (DTMC) and Continuous-Time Markov Chains (CTMC). In general, each queueing system can be considered as a particular case of the Markov process, and then mathematically evaluated in terms of the process. Direct and iterative methods can be used for obtaining numerical solutions in the steady-state analysis of the Markov chains (Bhalai 2002). Direct methods, as they are applied, can modify the parameter matrix. They can also require fixed amount of computational time, independently to the parameter values. At the same time, the direct methods are subject to accumulation of round-off errors, and create significant difficulties with sparse storage (Mehdi 2003). Iterative methods are based on the property of successive convergence to the desired solutions, and the evaluation process can be terminated as soon as iterates are sufficiently close to each other, i.e. the final/the exact values. The main advantage of iterative methods, compared with direct methods, is based on the fact that iterative methods maintain the parameter matrix (Stewart 1994), since efficient sparse storage schemes and efficient sparsity-preserving algorithms are applied. On the other hand, the main disadvantage of iterative methods is that convergence is not always guaranteed, and it is highly dependent on the applied method. The rate of convergence is very sensitive to the values of entries in the parameter matrix (Trivedi 2001). The important task here is to obtain exact or approximate numerical solutions for global and local balance equations of the queueing systems with the Markov chains. This is a focal point in modeling the closed and open queueing networks, especially with heavy tailed traffic. This is a reason why we developed and compared the computational procedures to obtain the equilibrium numerical solutions with use of direct, iterative and aggregative techniques in the steady state analysis of the Markov chains.

MARGINAL PROBABILITIES AND GLOBAL BALANCE EQUATION

Obtaining the steady-state probabilities $\pi(k_1, k_2,\ldots,k_N)$ for all possible states in the network is very important, as the mean values of other performance measures can be then calculated. We deal with the multidimensional state spaces, and $\pi(k_1, k_2,\ldots,k_N)$ will denote the steady-state probabilities of states $(k_1, k_2,\ldots,k_N)$. The most important performance measures for queueing networks are marginal probabilities $\pi_i(k)$. For the closed queueing networks the marginal probabilities $\pi_i(k)$ indicate that the i-th node contains exactly $k_i = k$ jobs that can be calculated according to (1):

$$\boldsymbol{\pi}_i(k) = \sum_{\substack{\sum_{j=1}^{N} k_j = K}}^{k_i = k} \boldsymbol{\pi}(k_1,\ldots,k_N) \qquad (1)$$

The marginal probabilities $\pi_i(k)$ are joint probabilities for all possible states $(k_1, k_2,\ldots,k_N)$, $0 \le k_i \le K$, which satisfy the condition (2),

$$\sum_{j=1}^{N} k_j = K \qquad (2)$$

while a fixed number of jobs, k, is specified for the i-th node. The normalization condition for the joint probabilities is shown in (3):

$$\sum_{\sum_{j=1}^{N} k_j = K} \boldsymbol{\pi}(k_1,\ldots,k_N) = 1 \qquad (3)$$

For the open networks we have the marginal probabilities calculated according to (4), with use of normalization condition (5):

$$\boldsymbol{\pi}_i(k) = \sum_{k_i = k} \boldsymbol{\pi}(k_1,\ldots,k_N) \qquad (4)$$

$$\sum \boldsymbol{\pi}(k_1,\ldots,k_N) = 1 \qquad (5)$$

Using the marginal probabilities we can obtain other significant performance measures for the open and closed networks, such as utilization, throughput, mean number of jobs, mean queue length, mean response time and mean waiting time.

ALGORITHMS FOR STEADY STATE SOLUTIONS OF MARKOV CHAINS

Based on fundamental equations $\mathbf{v} = \mathbf{vP}$ and $\mathbf{0} = \boldsymbol{\pi}\mathbf{Q}$, the model (6) can be used for obtaining the steady-state probability vector in the ergodic Markov chains:

$$\mathbf{0} = \mathbf{v}(\mathbf{P} - \mathbf{I}) \qquad (6)$$

Consequently, for both discrete-time and continuous-time Markov chains, the linear system (7) should be solved:

$$\mathbf{0} = \mathbf{xA} \qquad (7)$$

Based on the entries that represent the parameters of the Markov chain, the matrix **A** is singular, and it has a rank of n-1 for any Markov chain of size $|S| = n$. In order to obtain a unique positive solution, we directly impose the normalization condition into (7) with use of (8):

$$\mathbf{x1} = 1 \qquad (8)$$

This can be defined as substitution of one column (say, the last column) of matrix **A** with the unit vector. The resulting linear system of non-homogeneous equations can be written according to (9).

$$\mathbf{b} = \mathbf{xA}, \quad \mathbf{b} = [0,0,\ldots,0,1] \qquad (9)$$

For any given ergodic CTMC, a DTMC can be developed, which obtains the identical steady-state probability vectors as in case of CTMC. Let’s consider the generator matrix $\mathbf{Q} - [q_{ij}]$ for CTMC, as shown in (10):

$$\mathbf{P} = \mathbf{Q}/q + \mathbf{I} \qquad (10)$$

where $q > \max_{i,j\in S}|q_{ij}|$. The situation $q = \max_{i,j\in S}|q_{ij}|$ should be avoided, in order to assure aperiodicity of the resulting DTMC [2]. The obtained matrix **P** can be used to determine the steady-state probability vector $\boldsymbol{\pi} = \mathbf{v}$, with solving $\mathbf{v} = \mathbf{vP}$ and $\mathbf{v1} = 1$. The method that can be applied in order to reduce a CTMC to a DTMC is called randomization, and sometimes, uniformization [3]. On the other hand, the transition probability matrix **P** for the ergodic DTMC is given, and the generator matrix **Q** of the CTMC can be defined according to (11).

$$\mathbf{Q} = \mathbf{P} - \mathbf{I} \qquad (11)$$

After resolving $\mathbf{0} = \boldsymbol{\pi}\mathbf{Q}$ considering condition $\boldsymbol{\pi}\mathbf{1} = 1$, the desired steady-state probability vector $\boldsymbol{\pi} = \mathbf{v}$ can be obtained.

DIRECT METHODS FOR NUMERICAL SOLUTIONS

The closed-form methods are applicable for the solutions when the Markov chains possess the particular structure. For the Markov chains with more general structures, we should avoid using numerical methods. Many direct methods are known to solve the system of linear equations. Some of them are restricted to the certain regular structures of the parameter matrices that are of less importance for the Markov chains, since these structures in general cannot be applied with the Markov chains. Among the most frequently applied techniques are the Gaussian elimination procedure, as well as Grassmann, Taksar, and Heyman techniques, which are based on the renewal arguments of the Gaussian elimination. The Gaussian elimination approach suffers sometimes from numerical difficulties created by subtractions of nearly equal numbers. These particular properties, as well as their variants through the reformulations on the base of the regenerative properties of the Markov chains, can be avoided when applying Grassmann, Taksar, and Heyman algorithms. The basic idea of the Gaussian elimination approach is transforming the original system (12) into another equivalent system with use of elementary operations on the parameter matrix, which preserve the rank of the matrix.

$$\begin{aligned} &a_{0,0}x_0 + a_{1,0}x_1 + \cdots + a_{n-1,0}x_{n-1} = b_0, \\ &a_{0,1}x_0 + a_{1,1}x_1 + \cdots + a_{n-1,1}x_{n-1} = b_1, \\ &\vdots \\ &a_{0,n-1}x_0 + a_{1,n-1}x_1 + \cdots + a_{n-1,n-1}x_{n-1} = b_{n-1}. \end{aligned} \tag{12}$$

As a result, the equivalent system of linear equations with a triangular matrix structure can be derived as (13), and the desired solution **x**, which is identical to the solution of the original system, can be obtained.

$$\begin{aligned} &a_{0,0}^{(n-1)}x_0 = b_0^{(n-1)}, \\ &a_{0,1}^{(n-2)}x_0 + a_{1,1}^{(n-2)}x_1 = b_1^{(n-2)}, \\ &\vdots \\ &a_{0,n-1}^{(0)}x_0 + a_{1,n-1}^{(0)}x_1 + \cdots + a_{n-1,n-1}^{(0)}x_{n-1} = b_{n-1}^{0}. \end{aligned} \tag{13}$$

As soon as the original system (12) is transformed into a triangular structure, the final results can be obtained applying a straightforward substitution process. In order to obtain (13), the elimination procedure needs to be performed on the original system. Approximately, the algorithm can be described as follows: first, the n$^{\text{-th}}$ equation in (12) is solved for x_{n-1}, and then x_{n-1} is eliminated from all other n-1 equations. Next, the $(n$-1$)^{\text{th}}$ equation is solved for x_{n-2}, and, again, x_{n-2} is eliminated from the remaining n-2 equations; and so forth. Finally, the system (13) is resolved, where $a_{i,j}^{(k)}$ denotes the coefficient of x_i in the $(j+1)^{\text{-th}}$ equation, which is obtained after the k$^{\text{-th}}$ elimination step. The Gaussian elimination procedure is based on the elementary matrices operations that preserve the rank of the matrix. Such elementary matrices operations are corresponding to interchange and multiplication of equations by a real-value constant. In matrix terms, the essential part of the Gaussian elimination is provided by the factorization of the parameter matrix **A** into components of the upper triangular matrix **U** and the lower triangular matrix **L**. Finally**,** the computation of consequential vector **x** can be made applying two simple steps according to (14):

$$\mathbf{b} = \mathbf{xA} = \mathbf{xUL} = \mathbf{yL} \tag{14}$$

At first, equation **yL** = **b** is resolved for the vector of unknowns **y,** and then equation **xU** =**y** is resolved for the vector of unknowns **x**. The enhanced computational procedure, developed according to the Gaussian elimination approach, is demonstrated as Algorithm 1.

Algorithm 1:

Step 1: *Construct the parameter matrix* **A** *and the right-hand-side vector* **b** *according to:* **b** = **xA**, $\mathbf{b} = [0,0,...,0,1]$.

Step 2: *Carry out elimination steps, or apply the standard algorithm to split the parameter matrix* **A** *into upper triangular matrix* **U** *and lower triangular matrix* **L** *such that* **A** = **UL** *holds. Note that the parameters of* **U** *can be computed with:*

$$a_{ij}^{(k)} = \begin{cases} 0, & \begin{array}{l} j = n-k-1, n-k-2, ..., 0 \\ i = n-1, n-2, ..., n-k, \end{array} \\ a_{ij}^{(k-1)} - a_{n-k}^{(k-1)} \dfrac{a_{n-k,j}^{(k-1)}}{a_{n-k,n-k}^{(k-1)}}, & otherwise \end{cases},$$

as the computation of **L** *can be deliberately avoided.*

Step 3: *Compute the intermediate values* **y** *according to:* **yL** = **b** *, or compute the intermediate values according to:*

$$b_j^{(k)} = b_j^{(k-1)} - b_{n-k}^{(k-1)} \frac{a_{n-k,j}^{(k-1)}}{a_{n-k,n-k}^{(k-1)}},$$

$$where \quad j = n-k-1, n-k, ..., 0.$$

Step 4: *Perform the substitution to yield the final result* **x** *according to:* **xU** = **y** *applying the formula:*

$$x_0 = \frac{b_0^{(n-1)}}{a_{0,0}^{(n-1)}}, x_j = \frac{b_j^{(n-j)}}{a_{j,j}^{(n-j)}} - \sum_{k=0}^{j-1} \frac{a_{k,j}^{(n-j)}}{a_{j,j}^{(n-j)}} x_k, \quad j = 1,2,...,n-1.$$

ITERATIVE METHODS FOR NUMERICAL SOLUTIONS

Convergence is a very important issue for iterative methods. Heuristic approach can be applied for choosing appropriate techniques of decision-making on convergence, but there are no general algorithms for the selection of such techniques. An estimate of the error must be made to determine convergence, because the desired solution vector is not known. The tolerance level ε must be specified in order to provide a measure of how close the current iteration vector $x^{(k)}$ is to the desired solution vector **x**. Some distance measures are often used to evaluate the current iteration vector $x^{(k)}$ in relation to some earlier iteration vectors $x^{(l)}, l < k$. If the current iteration vector is "close enough" to the earlier iteration vectors with respect to ε, then this condition is taken as an indicator of convergence to the final result. If ε is too small, convergence could become very slow, or not take place at all. If ε is too large, accuracy requirements could be violated, or, worse, convergence could be wrongly assumed. Some appropriate norm functions have to be applied in order to compare different iteration vectors. Size and type of the parameter matrix should be taken into consideration for the right choice of such a norm function. Concerning the right choice of ε and the norm function, we can say that components x_i of the solution vector can differ significantly from each other. The power method is a reliable iterative technique for computing the steady-state probability vector of the finite ergodic Markov chains. It sometimes tends to converge slowly, and the solely condition needed for convergence is the transition probability matrix **P** to be aperiodic, and then irreducibility is not necessary. The power method follows the transient behavior of the underlying discrete-time Markov chains, until some stationary, not necessarily steady-state, convergence is obtained. Therefore, it can also be used as a technique for computing the transient state probability vector $\mathbf{v}(n)$ of a DTMC. Equation $\mathbf{v} = \mathbf{vP}$ recommends to start with the initial guess of some probability vector $\mathbf{v}^{(0)}$, and repeatedly multiply it by the transition probability matrix $\mathbf{P}$ until convergence to v is obtained, with $\lim_{i\to\infty} \mathbf{v}^{(i)} = \mathbf{v}$. Since ergodicity, or at least aperiodicity of the underlying Markov chain are assumed, this procedure is guaranteed to converge to the desired fixed point of the unique steady-state probability vector. A single iteration step is the following according to (15):

$$\mathbf{v}^{(i+1)} = \mathbf{v}^{(i)}\mathbf{P}, \quad i \geq 0 \tag{15}$$

The relation between the iteration at step i and the initial probability vectors can be presented as (16).

$$\mathbf{v}^{(i)} = \mathbf{v}^{(0)}\mathbf{P}^{i}, \quad i \geq 0 \tag{16}$$

To obtain the final result of the steady-state probability vector ***v*,** only the renormalization remains to be executed. The speed of convergence of the power method depends on the relative sizes of the eigenvalues. The closer non-dominant eigenvalue is equal to 1, which slower the convergence. The enhanced computational procedure, developed according to the iterative power method, is demonstrated as Algorithm 2.

Algorithm 2:

Step 1: *Select q appropriately:*

$$\mathbf{A} = \begin{cases} \mathbf{P}, \\ \mathbf{Q}/q + \mathbf{I}; \end{cases} \qquad \mathbf{v}^{(0)} = \left(\mathbf{v}_0^{(0)}, \mathbf{v}_1^{(0)}, \ldots, \mathbf{v}_{n-1}^{(0)}\right)$$

Select convergence criterion ε, *and let* $n = 0$.

Define accurately some vector norm function $f\left(\left\|\mathbf{v}^{(n)}, \mathbf{v}^{(l)}\right\|\right), n \geq l$.

Set convergence = false.

Step 2: *Repeat until convergence:*

Step 2.1: $\mathbf{v}^{(n+1)} = \mathbf{v}^{(n)}\mathbf{A}$;

Step 2.2: *If* $f\left(\left\|\mathbf{v}^{(n+1)}, \mathbf{v}^{(l+1)}\right\|\right) < \varepsilon, l \leq n$,

Then convergence = true.

Step 2.3: $n = n + 1, \quad l = l + 1$

Step 3: $\left.\begin{matrix}\boldsymbol{\pi} \\ \mathbf{v}\end{matrix}\right\} \approx \mathbf{v}^{(n)}$.

APPROXIMATION METHODS FOR NUMERICAL SOLUTIONS

The two basic approximation methods for the steady-state analysis of Markov chains (Bhalai 2002) can be used to analyze the communication systems with the Markov chains. The Courtois's approach is usually applied in order to approximate computing the $v^{\approx}$ of the desired state probability vector v. The efficiency of the Courtois's method is based on the fact that instead of solving one linear system of equations with size of the state space *S*, several much smaller linear systems are solved independently. One system is solved for each subset S_j of the partitioned state space *S,* and one more system for the aggregated chain (Radev et al. 2005). Takahashi's method differs substantially from Courtois's approach with respect to both the used methodology and the applicability conditions. While the Courtois's method is non-iterative and applied for the approximate computations of the steady-state probability vector $v^{\approx}$ for the given ergodic DTMC (or $\pi^{\approx}$ in the case of a CTMC); the Takahashi's iterative method allows computations of the exact state probability vector. To perform a straightforward comparison of these two methods, any given ergodic CTMC can easily be transformed into the ergodic DTMC. The Courtois's approach is based on the decomposability properties of considered models, when the aggregation procedure is executed that uses independently computed sub results as constituent parts for composing the final results. The applicability of the method must be verified in each case. If the Markov chain has the tightly coupled subsets of states, where the states within each subset are tightly coupled to each other and weakly coupled to states outside the subset, it provides a strong intuitive indication of the applicability of the approach. Such subsets of states might then be aggregated to form the macro states as the base for the further analysis. The macro state probabilities, together with the conditional micro state probabilities within the subsets, can be composed to obtain the micro state probabilities of the initial model. Let **P** be the transition probability matrix. The state space, associated with **P,** is partitioned into groups, often referred to as lumps, such that any state within the group contains the same number of customers in a specific portion of the state descriptor. The transition probability matrix **P**, obtained according to these groups, consists of blocks of transition probabilities along the diagonal, and of other blocks along the off-diagonals above and below the diagonal. **P** is assumed to be nearly completely decomposable if the sum of the non-zero transition probabilities on each row that is positioned within the diagonal block is close to 1. That means that the sum of the off-diagonal probabilities along a row is extremely small. This occurs when the most of transitions are between the states of the same group, with very few transitions between the states of different groups. The error, induced by the method, could, generally, be bounded, but a formal error bounding is often omitted. The enhanced computational procedure, developed according to the Courtois's approximation method, is demonstrated as Algorithm 3.

Algorithm 3:

Step 1: *Create the state space and arrange it suitably according to a pattern of decomposition.*

Step 2: *Build the transition probability matrix* **P** *(use a randomization technique* **P**=**Q**/*q*+**I** *if the starting point is a CTMC), and partition* **P** *into* $M \times M$ *number of sub- matrices* $\mathbf{P}_{\mathrm{IJ}}^{*}, 0 \leq \mathrm{I}, \mathrm{J} \leq M - 1$.

Step 3: *Verify the nearly complete decomposability of* **P** *according to:*

$$\varepsilon < \frac{1 - \max_{\mathrm{I}}\left|\lambda_{\mathrm{I}}^{*}(2)\right|}{2} \quad \textit{with the chosen value of } \varepsilon.$$

Step 4: *Decompose* $\mathbf{P}$ *such that* $\mathbf{P} = \mathbf{P}^* + \varepsilon \cdot \tilde{\mathbf{C}}$. *Matrix* $\mathbf{P}^*$ *contains only stochastic diagonal sub matrices* $\mathbf{P}^*_{II}$, *and* ε *is a measure of the accuracy of Courtois's method. It is defined as the maximum sum of the entries of the non-diagonal sub-matrices* $\mathbf{P}^*_{IJ}, I \neq J$ *of* $\mathbf{P}$.

Step 5: *For each* $I, 0 \leq I, J \leq M-1$, *solve equation* $\mathbf{v}^*_I \cdot \mathbf{P}^*_{II}, = \mathbf{v}^*_I$, *with* $\mathbf{v}^*_I \cdot 1 = 1$ *to obtain the conditional state probability vectors* $\mathbf{v}^*_I$.

Step 6: *Compute the coupling between the decomposed macro states:*

Step 6.1: Generate the transition probability matrix $\mathbf{\Gamma} = [\mathbf{\Gamma}_{IJ}]$ *according to equation* $\mathbf{\Gamma}_{IJ} = \sum_{i \in S_I} \left(v^*_{Ii} \sum_{j \in S_j} p_{IJij} \right)$.

Step 6.2: Solve $\boldsymbol{\gamma}\mathbf{\Gamma} = \boldsymbol{\gamma}$, *to obtain the macro steady-state probability vector* $\boldsymbol{\gamma}$.

Step 7: *Compute the approximate steady state probability vector* $\mathbf{v}^{\approx}$ *of the micro states applying the conditional state probability vectors* $\mathbf{v}^*_I, 0 \leq I \leq M-1$, *according to:*

$$\mathbf{v}^{\approx}_{Ii} = \gamma_I \mathbf{v}^*_{Ii}, \quad 0 \leq I \leq M-1 \quad \textit{and} \quad \forall i \in S_I .$$

Step 8: *From* $\mathbf{v}^{\approx}$ *compute steady state performance and dependability measures along the lines of a specified reward structure.*

Compared to the Courtois's method, the Takahashi's approach partitions the state spaces S into M disjoint subsets of states $S_I \subset S, \quad 0 \leq I \leq M-1$, such that each subset S_I is aggregated into the macro state I. The criteria, however, used to the cluster states, are different in these two approaches. The calculations of the transition probabilities $\mathbf{\Gamma}_{IJ}$ among the macro states I and J are performed on the base of conditional micro-state probability vector $\mathbf{v}^*_I$ and the originally given transition probabilities p_{ij}, or p_{IJij}. With the Courtois's method, the conditional probabilities vectors $\mathbf{v}^*_I$, for the given partition element I, can be separately obtained for each subset of the micro states $S_I \subset S, 0 \leq I \leq M-1$, if the original model is nearly completely decomposable. By contrast, in the Takahashi's method, the partitioning of the state space is performed on the base of approximate lumpability, and it was presented in terms of DTMCs. However, this choice implies no limitations, since the ergodic DTMCs and CTMCs are equivalent with respect to their steady-state probability computations that applying transformations. The Takahashi's method is more convenient to be applied directly on the generator matrix $\mathbf{Q} = [q_{ij}]$ of the CTMC, and it allows obtaining the exact state probability vector. In the Takahashi's method the partitioning of the state space is performed on the base of approximate lumpability. The enhanced computational procedure, built according to the Takahashi's iterative method, is shown as Algorithm 4.

Algorithm 4:

Step 1 *Create the state space S and arrange it suitably according to a pattern of decomposition along the lines of approximate lump-ability into transition probability matrix.*

Step 2: *Initialization:*

Step 2.1 n := 0.

Step 2.2 Estimate $\mathbf{v}^0$.

Step 2.3 Choose ε *and* $0 < c < 1$.

Step 2.4 Choose some vector norm function f(||…||).

Step 3: $f(\|\mathbf{v}^{(n)} - \mathbf{v}^{(n)}P\|) \geq \varepsilon$.

Step 4: *Geometric convergence: from Y to N:* $n \geq 1$

residual error ($\mathbf{v}^{(n)}$ *)>c*residual error(* $\mathbf{v}^{(n-1)}$ *)*

$\mathbf{v}' = \mathbf{v}^{(n)}$;

$\mathbf{v}^{(n)} = \mathbf{v}'\mathbf{P}$ *n:=n+1.*

Step 5: *Aggregation for all:* $0 \leq I,J \leq M$ -1.

Step 5.1 Create appropriately

$$\mathbf{\Gamma}^{(n)} = [\mathbf{\Gamma}^{(n)}_{IJ}] = \left[\sum_{i \in S_I} \sum_{j \in S_J} \frac{\mathbf{v}^{(n-1)}_i p_{ij}}{\sum_{k \in S_I} \mathbf{v}^{(n-1)}_k} \right];$$

Step 5.2 Solve $\boldsymbol{\gamma}^{(n)} = \boldsymbol{\gamma}^{(n)} \cdot \mathbf{\Gamma}^{(n)}$

according to: $\boldsymbol{\gamma} = \boldsymbol{\gamma} \cdot \mathbf{\Gamma}, \quad \boldsymbol{\gamma} \cdot 1 = 1$;

Step 6: *Disaggregating for all:* $0 \leq I \leq M-1$.

Step 6.1 Calculate properly

$$\mathbf{\Gamma}^{(n)}_{Ij} = \sum_{i \in S_I} \frac{\mathbf{v}^{(n-1)}_i p_{ij}}{\sum_{k \in S_I} \mathbf{v}^{(n-1)}_k} \qquad \forall j \in S \textit{ according to:}$$

$$\mathbf{\Gamma}_{Ij} = \sum_{i \in S_I} \frac{\mathbf{v}^{\approx}_i p_{ij}}{\sum_{k \in S_I} \mathbf{v}^{\approx}_k}$$

Step 6.2 Calculate $\mathbf{v}^{(n)}_I = [v_{Ij}]$ *by solving the system of equations:*

$$\mathbf{v}^{(n)}_i = \sum_{j \in S_I} \mathbf{v}^{(n)}_j p_{ji} + \sum_{K=0, K \neq I}^{M-1} \gamma^{(n)}_K \mathbf{\Gamma}^{(n)}_{K\ i}, \qquad \forall i \in S_I ;$$

Step 7: *With* $\mathbf{v}^{(n)} = [v^{(n)}_I]$ *from the previous step solve* $\mathbf{v}^{(n)} \cdot 1 = 1$.

COMPARISON OF NUMERICAL RESULTS

Let's consider the implementations of the algorithms with the illustrations. At first, let's consider the example of the closed queueing network that contains three nodes (*N*=3) and four customers (*K*=4), as shown in Fig. 1(a). The number of evident states of the system is equal to the possible variation and is calculated as binomial coefficient:

$$\binom{N+K-1}{N-1} = \binom{6}{2} = \frac{6!}{2! \quad 4!} = 15 \qquad (17)$$

The state transition rate diagram for the queueing model is presented with the Markov chain that is shown in Fig. 1(b). The diagram demonstrates possible transitions between the Markov chain nodes. The transition rates between the states are equal to $\mu_1 = 0{,}5;$ $\mu_2 = 0{,}4;$ $\mu_3 = 0{,}1$. The numeration of the states represents the total number of the customers in each node. To define the system we have fifteen global-balance equations, and for each of them the left-hand side represents the flow out of the state and the right-hand side represents the flow into the state.

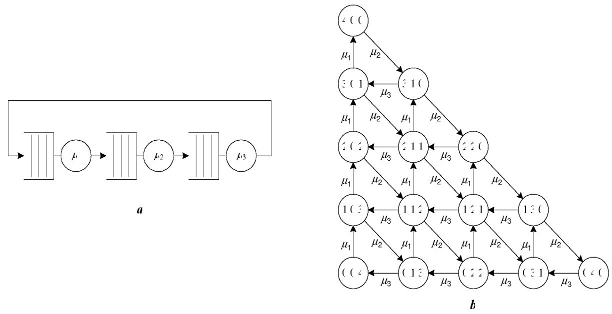


Figure 1: Queueing model (a) and transition rate diagram (b)

First, we write down the local balance equations and search the solutions, applying the substitution. This is how we get the exact steady-state probabilities, as shown in Table 1. Next, we get the results for the steady-state probabilities with the Gaussian elimination approach (Algorithm 1), the iterative power method (Algorithm 2), and the Courtois's approximation method (Algorithm 3). The desired probability vector is not known in advance, that's why an estimate of the relative error is used. A good convergence for the iterative power method is received for ν=90 iterations. The worst results are received for the Courtois's approximation approach, since the included aggregation procedure is dependent on the matrix type. Comparing the obtained results with exact ones, we prove that the enhanced algorithms, based on direct and iterative methods, provide excellent results for the closed networks (with relative error $<10^{-6}$). The approximation algorithm, based on the aggregation and desegregation, is not suitable for the demonstrated example, since the sum of elements around main diagonal is close to 1, and the elements, positioned far from the main diagonal, are not close to 0. This makes obvious that the aggregation and desegregation could be applied to only computations with the limited number of the queueing networks. The basic principles of the decomposition, implemented in Algorithms 3 and 4, are demonstrated in the example with the CTMC on Fig. 2(a). The closed queueing network with two customers and three server stations is considered. Each arbitrary pattern of the customers distributed among the stations is represented by a state. Consequently, in this example with N=3 and K=2 there are 6 states according to (17). For instance, in state (2,0,0) two customers are at service station one, while service stations two and three are both empty. After an exponentially distributed time period, a customer travels to service station two. The transition behavior is controlled with the transition rate μ_{12}. The transition behavior between the other states can be explained similarly. The lines point the direction for the possible transitions, since only counter clockwise direction is allowed in the network. Further, we assume that the customers preferably stay at two service stations, and relatively rarely relocate to the third service station, which is the most isolated. The decomposition is shown in Fig. 2(b). The solid lines emphasize the tightly coupled states, while the dotted lines represent the loose coupling. There are parameters with strong interaction between state 1 and 2, while state 3 interacts less with the others and their transition rates are μ_{12}=4; μ_{21}=0; μ_{13} = 0; μ_{31}=0,2; μ_{23}=0,4; μ_{32}=0.

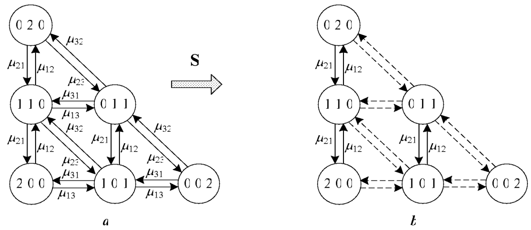


Figure 2: CTMC of queueing network (a) and its decomposition with regard to service station three (b)

The corresponding infinitesimal generator matrix **Q** is determined as the following:

$$\mathbf{Q} = \begin{pmatrix} -4 & 4 & 0 & 0 & 0 & 0 \\ 0 & -4.4 & 4 & 0.4 & 0 & 0 \\ 0 & 0 & -0.4 & 0 & 0.4 & 0 \\ 0.2 & 0 & 0 & -4.2 & 4 & 0 \\ 0 & 0.2 & 0 & 0 & -0.6 & 0.4 \\ 0 & 0 & 0 & 0.2 & 0 & -0.2 \end{pmatrix}.$$

The decomposition and values for generator matrix **Q** are shown symbolically in the following configuration:

State	(2,0,0)	(1,1,0)	(0,1,0)	(1,0,1)	(0,1,1)	(0,0,2)
(2,0,0)	$-\sum$	μ_{12}	0	μ_{13}	0	0
(1,1,0)	μ_{21}	$-\sum$	μ_{12}	μ_{23}	μ_{13}	0
(0,2,0)	0	μ_{21}	$-\sum$	0	μ_{23}	0
(1,0,1)	μ_{31}	μ_{32}	0	$-\sum$	μ_{12}	μ_{13}
(0,1,1)	0	μ_{31}	μ_{32}	μ_{21}	$-\sum$	μ_{23}
(0,0,2)	0	0	0	μ_{31}	μ_{32}	$-\sum$

This decomposition is used in the numerical example, while computing the steady-state probability vector with use of Algorithms 3 and 4. The relative error between the obtained and exact values was used to estimate the precise solutions. Additional comparative analysis of Algorithms 2, 3 and 4 is also made. Good convergence of numerical results is obtained after ν =200 iterations, as presented in Table 2. Better results could be obtained after using more levels of decomposition, as the state probabilities may vary significantly. Then, it is obvious that the Takahashi's approach (Algorithm 4) provides very good results. After the first iteration step $\nu^{(1)}$, the results are not near to the exact values, but after the second iteration step $\nu^{(2)}$, they resemble the exact values up to the third decimal digit. Comparison of running times of the decomposition and iterative algorithms can be obtained only in the case of significant number of the steady states (not less then 100).

Table 1 Numerical Results for Steady-State Probabilities Computed with Different Algorithms

State	Exact Value	Algorithm 1		Algorithm 2		Algorithm 3	
		Computed	Error $\times 10^{-6}$	$\nu^{(90)}$	Error $\times 10^{-6}$	$\nu^{\approx}$	Error
(0,0,4)	0.000962837	0.000962837	0.252	0.000962837	0.287	0.044678333	0.9784
(0,1,3)	0.004814184	0.004814184	0.037	0.004814184	0.065	0.0340617	0.8587
(0,2,2)	0.024070919	0.024070919	0.01	0.024070919	0.014	0.029638102	0.1878
(0,3,1)	0.120354595	0.120354595	0.002	0.120354594	0.011	0.065026881	0.8508
(0,4,0)	0.601772974	0.601772974	0	0.601772969	0.009	0.218083349	1.7594
(1,3,0)	0.150446243	0.150446243	0	0.150443245	0.011	0.125959943	0.1944
(2,2,0)	0.037510811	0.037510811	0.006	0.037610813	0.057	0.142925268	0.7368
(3,1,0)	0.009402703	0.009402703	0.052	0.009402704	0.126	0.115046814	0.9183
(4,0,0)	0.002350674	0.002350674	0.186	0.002350676	0.971	0.062154431	0.9622
(3,0,1)	0.00188054	0.00188054	0.069	0.001880541	0.415	0.021921065	0.9142
(2,0,2)	0.001504433	0.001504433	0.166	0.001504432	0.352	0.023430384	0.9358
(1,0,3)	0.001203546	0.001203546	0.174	0.001203546	0.067	0.020034416	0.9399
(1,1,2)	0.00601773	0.00601773	0.049	0.00601773	0.048	0.006432901	0.0645
(1,2,1)	0.030088649	0.030088649	0.015	0.030088649	0.005	0.06432901	0.5323
(2,1,1)	0.007522162	0.007522162	0.02	0.007522163	0.075	0.026277403	0.7137

Table 2 Computed Steady-State Probabilities for Example 2

State	Exact Value	Algorithm 3		Algorithm 4			Algorithm 2	
		$\nu^{\approx}$	Error	$\nu^{(0)}$	$\nu^{(2)}$	Error	$\nu^{(200)}$	Error $\times 10^{-5}$
(2,0,0)	0.00136	0.0049	2.56	$0.16\overline{6}$	0.0015	0.0714	0.0013679	0.731
(1,1,0)	0.01368	0.0139	0.017	$0.16\overline{6}$	0.0137	0	0.013679	0
(0,2,0)	0.13679	0.1391	0.017	$0.16\overline{6}$	0.1378	0.0073	0.1367993	0.241
(1,0,1)	0.02736	0.0382	0.398	$0.16\overline{6}$	0.0273	0.0036	0.0273597	0.183
(0,1,1)	0.27359	0.2679	0.02	$0.16\overline{6}$	0.2739	0.0011	0.273598	0
(0,0,2)	0.54719	0.5358	0.02	$0.16\overline{6}$	0.5457	0.0026	0.547195	0.007

CONCLUSIONS

New approach to the steady-state numerical solutions for the queueing networks is recommended in this paper, which is based on computing the stochastic probability vectors of the Markov chains. The computational procedures for the steady state analysis of the ergodic Markov chains, based on the Gaussian elimination direct approach, the power iterative approach, the Courtois's approximation method, and the Takahashi's iterative method are developed and explored. The numerical examples for the closed queueing networks with three nodes and exponentially distributed service rates are demonstrated. The obtained numerical results are analyzed and compared with exact results. The obtained outcome can be summarized as follows:

1. The algorithms, based on the direct approach, are effective; they modify the parametric matrix and require fixed amount of computational time. With increasing number of the steady states of the Markov chains the convergence is expected to decrease. Direct methods provide excellent results (with relative error less then 10^{-6}) for the limited number of the network states.
2. The convergence of iterative algorithms, based on the power approach, is more dependent on performance parameters (throughput, response service time, and queueing length), and less dependent on the structure of the queueing networks.
3. The computational precision and convergence of the algorithms, based on the aggregation and desegregation of the Markov chains, is considerably dependent on selected decomposition and values of the infinitesimal generator matrix. That's why the implementation of these algorithms is limited to the queueing networks with the small queue length and a small number of the steady states.

REFERENCES

Bhalai, S. (2002), *Markov Decision Processes: the control of high-dimensional systems*, Universal Press, Amsterdam, The Netherlands.

Hayes, J., Ganesh Babu, T., (2004), *Modeling and Analysis of Telecommunications Networks*, New York, John Wiley and Sons.

Mehdi, J. (2003), *Stochastic Models in Queueing Theory*, Academic Press.

Krieger, U., B. Muller-Clostermann, A. Sczittnick, (1990), Modeling and Analysis of Communication Systems Based on Computational Methods for Markov Chain, *IEEE Journal on Selected Areas in Communications*, Vol. 8. (9), 1630-1648.

Radev, D., Radeva, S. (2003), Artificial intelligence modeling of stochastic processes in digital communication networks, *Journal of Electrical Engineering*, Vol. 54, N: 9-10, 255-259.

Radev, D., Lokshina, I., Radeva, S. (2005), Clustering Procedure for Performance Analysis of Mobile Communication Networks, *Proceedings of the Industrial Simulation Conference,* Berlin, Germany, 83-87.

Radev, D., Lokshina, I., Radeva, S., Insinga, R. (2004), Neural Modeling of Link Occupancy Distribution for Broadband Telecommunication Transmission, *Proceedings of the DSI: 35-th Annual Meeting of the Decision Sciences Institute*, Boston, USA, 3001-3009.

Radev, D., Lokshina, I., Radeva, S. (2005), Estimation of the Local Balance of Queueing Networks with Clustering and Neural Modeling. *Proceedings of the International Conference on Telecommunication Systems, Modeling and Analysis: ICTSM'2005*, Dallas, TX, USA, 187-194.

Radev, D. (2005), Fluid Flow Analysis for Long-Range Dependent Traffic, *Proceedings of the Computer Science'2005*, Chalkidiki, Greece.

Radev, D., Denchev, V. Rashkova, E., (2005), Steady State Numerical Solutions of Markov Chains. *7th Balkan Conference on Operational Research*, Bucharest, Romania.

Rubinstein, R., Melamed, B., (1998), *Modern Simulation and Modeling,* New York, John Wiley & Sons.

Stewart, W. (1994), *Introduction to Numerical Solution of Markov Chains*, Princeton University Press, N.J., USA.

Trivedi, K. (2001), *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*, New York, John Willey & Sons.

BIOGRAPHY

DIMITAR RADEV, PhD is from University of Rousse, Bulgaria. In 2001 he became Associate Professor at Department of Communication Technique and Technologies at University of Rousse. His main research interests are Teletraffic Theory, Simulation and Modeling of Communication Networks, and Performance Analysis of Queuing Systems.

dradev@abv.bg

IZABELLA LOKSHINA, PhD is an Assistant Professor of Management Information Systems and Academic Coordinator of Management, Marketing and Management Information Systems at Division of Economics and Business at SUNY Oneonta. Her positions included Senior Scientific Researcher at the Moscow Central Research Institute of Complex Automation and Associate Professor of Automated Control Systems at Moscow State Mining University. Her main research interests are System Modeling and Artificial Intelligence (Fuzzy Systems and Neural Networks).

lokshiiv@oneonta.edu

VLADIMIR DENCHEV, PhD student is from University of Rousse, Bulgaria. His main research interests are Simulation and Modeling of Communication Networks, and Performance Analysis of Queuing Systems.

vndenchev@abv.bg

EVALUATION OF THE QUEUEING NETWORK EQUILIBRIUM BASED ON CLUSTERING ANALYSIS AND SELF-ORGANIZING MAP

Dimitar Radev
University of Rousse
8 Studentska Street
7017 Rousse, Bulgaria
E-mail: dradev@abv.bg

Izabella Lokshina
SUNY Oneonta
324B Netzer Administration Bldg.
Oneonta, NY 13820, USA
E-mail: lokshiiv@oneonta.edu

Svetla Radeva
University of Architecture,
Civil Engineering and Geodesy
1 Hr. Smirnensky Blvd. 1046
Sofia, Bulgaria
E-mail: svetla_fce@abv.bg

KEYWORDS

Mobile communication networks, Markov reward models, neural networks.

ABSTRACT

In this paper an approach to the network local-balance estimation is presented on the base of clustering and neural modeling, in which the cluster centers can replace the steady states of the Markov chain. The equilibrium of queueing networks is a mean for performance evaluation of real communication systems presented as Markov chains. The direct, iterative and approximation techniques for numerical steady state solutions are appropriate, but not always. Those techniques do not work for non-Markovian queues with non-exponential service times. The new approach was developed for local-balance performance evaluation, which was based on calculations of the steady-state probability vector with use of neural modeling and clustering. The experimental results were obtained with use of clustering analysis and self-organizing map, and the proposed method was compared with direct techniques on the base of obtained results.

INTRODUCTION

Traditional analysis of the real communication networks is made with use of Markov chain queueing models. Markov processes provide very flexible, powerful, and efficient means for the description and analysis of dynamic system properties, because performance and dependability measures can be easily derived and evaluated with steady-state analysis of Discrete-Time Markov Chains (DTMC) and Continuous-Time Markov Chains (CTMC) (Radev et al. 2005). Both direct and iterative methods can be used to make the steady-state analysis of Markov chain numerical solutions (Radev and Radeva 2003). When direct methods are applied, the parameter matrix is modified; and they are subject of round-off errors accumulation, which creates significant difficulties with limited storage (Bolch et al. 1998). Iterative methods are based on the property of successive convergence to the desired solution; and the process can be terminated if iteration results are sufficiently close to the final value. In fact, each queueing system can be mapped with the Markov process, and then it can be mathematically evaluated in terms of that process. Queueing networks, which consist of several service stations, are very suitable to represent the structure of various systems with large number of resources, for instance communication and computer systems. In the queueing networks at least two service stations are connected, while a station, i.e., a node in the network, represents a resource in the real system. As a rule, jobs can be transferred between any two nodes in the networks. The queueing networks are closed while jobs can neither enter nor leave the networks. The number of jobs in the closed networks is constant. The queueing networks are open when jobs can enter from and leave the networks to the outside. The basic characteristics of the networks' performance can be defined using the arrival rates, the departure rates, the queue length and the steady-states in particular networks, in which the birth - death processes can be observed. The birth – death processes are the Markov chains, in which the transitions are allowed only between the neighboring states. One of the important tasks here is the estimation of the network equilibrium.

MARKOVIAN QUEUES IN EQUILIBRIUM

The behavior of various queueing system models can be described using CTMC, which are defined on the base of the transition rates between the states. If the CTMC is ergodic, then the unique steady-state probability vector exists, which is different from the initial probability vector. The system of equations to determine the steady-state probability vector π is given on the base of $\boldsymbol{\pi}\mathbf{Q}=\mathbf{0}$, where $\mathbf{Q}$ is the infinitesimal generator matrix of the CTMC. According to this equation, the flux out of the state is equal to the flux into that state for each state of a queueing network in equilibrium. This conservation of flow in the steady state can be presented as (1),

$$\sum_{j\in S}\pi_j q_{ji} = \pi_i \sum_{j\in S} q_{ij}, \qquad \forall i \in S, \tag{1}$$

This quantity q_{ij} is the transition rate from state i to state j. After subtracting $\pi_i \cdot q_{ij}$ from both sides in (1), and considering that $q_{ij} = -\Sigma_{j\neq i}\, q_{ij}$, we can obtain the global balance equation (2),

$$\forall i \in S: \qquad \sum_{j\neq i}\pi_j q_{ji} - \pi_i \sum_{j\neq i} q_{ij} = 0, \tag{2}$$

which is correspondent to the matrix equation $\boldsymbol{\pi}\mathbf{Q}=\mathbf{0}$. Numerical techniques based on solving the global balance equations can be used, but for large networks this technique is very expensive and time-consuming, because the number of equations can be extremely large. In case of such large networks we should avoid generating and solving global balance equations, and we should better look for alternative techniques of solutions. If all nodes of networks fulfill certain assumptions concerning the distributions of the inter-arrival and service times and the Queueing discipline, then it is possible to derive local balance equations, which describe the system behavior in an unambiguous way. These local balance equations allow the essential simplification with respect to the global balance equations, because each equation can be split into the set of single equations, related to each individual node. Product-form networks are Queueing networks, which have an unambiguous solution of the local balance equations. The steady-state solutions for such networks' state probabilities consist of multiplicative factors, and each factor relates to a single node. The concept of equilibrium is that theoretical basis, which should be used in analysis. Different approaches should be used while solving the linear system in the form $\mathbf{0} = \mathbf{xA}$, which defines the steady-state probabilities of Markov chains. In this case direct, iterative and approximation numerical methods are known, and use of these techniques can lead to closed-form results. While direct methods yield exact results and are not efficient at all, iterative methods are generally more efficient, both in time and space. Since iterative methods are considerably more efficient in solving Markov chains, they are commonly used for larger models. For smaller models with less than a few thousand states, direct methods are reliable, accurate and efficient. Though closed-form results are highly desirable, they can be obtained for only a small class of models that have some structure in their matrix. Our starting point is equation $\boldsymbol{\pi}\mathbf{Q}=\mathbf{0}$, which expresses the equilibrium conditions for a general ergodic discrete-state continuous-time Markov process; let's recall that $\boldsymbol{\pi}$ is the row vector of equilibrium state probabilities and that $\mathbf{Q}$ is the infinitesimal generator, which elements are the infinitesimal transition rates of Markov process. In Queueing theory more standard notation is used, when vector $\boldsymbol{\pi}$ is replaced with the row vector $\mathbf{p}$, which k^{th} element is the equilibrium probability p_k of the system being found in state E_k. Our task is to solve $\mathbf{pQ} = \mathbf{0}$, with the additional conservation relation according to (3).

$$\sum_k p_k = 1 \tag{3}$$

This vector equation describes the "equation of motion" in equilibrium. Let's consider a closed network, in which the state variable consists of the vector ($k_1, k_2, \ldots, k_N$), and k_i is the number of jobs in the i^{th} node (including the jobs in service). Let's associate the equilibrium probability with this state, denoted by $p(k_1, k_2, \ldots, k_N)$. Similarly, the marginal distribution of finding k_i jobs in the i^{th} node is denoted by $p_i(k_i)$. The joint distribution for all nodes factored into the product of each of the marginal distributions according to (4),

$$p(k_1, k_2, \ldots, k_N) = p_1(k_1)\, p_2(k_2) \cdots p_N(k_N) \tag{4}$$

The quantity $p_i(k_i)$ is given as solution of the classical M/M/m system. Let's consider Markovian network with fixed and finite number of jobs K, while the unique equilibrium probability distribution exists for $p(k_1, k_2, \ldots, k_N)$, and the number of distinguishable states of the system is equal to the number of ways, in which K jobs can be placed among the N nodes. It is equal to binomial coefficient (5), which determines number of states in Markov chain.

$$\binom{N+K-1}{N-1} = \frac{(N+K-1)!}{(N-1)!\,K!} \tag{5}$$

As an illustration of numerical techniques to determine network equilibrium, a simple model of closed queueing network can be considered, which consists of $N = 3$ nodes with service time rates μ_1, μ_2, μ_3. There is K=2 jobs in the network, and the routing probabilities are p_{12}, p_{13}, p_{21}, and p_{31} (see Fig.1*a*). We define a local-balance equation (with respect to the given network state and the network node i) as one, which equates the rate of flow out of that network state due to departure of a job from node i, to the rate of flow into that network state, due to the arrival of a job to node i. Example with the three-node network is demonstrated. The service rate in the i^{th} node is given as μ_i, and it is independent to the number of jobs in that node. The number of states with state transitions rates is determined according to (5), and the state-transition-rate diagram of the underlying continuous-time Markov chain is presented in Fig.1*b*. The following states are possible in the network: (2,0,0), (0,2,0), (0,0,2), (1,1,0), (1,0,1), (0,1,1). We have six global-balance equations for this system, and for each of them left-hand side represents the flow out of a state and the right-hand side represents the flow into that state, as shown in (6).

$$\begin{gathered}
\pi(2,0,0)(\mu_1 p_{12}+\mu_1 p_{13}) = \pi(1,0,1)\mu_1 p_{31} + \pi(1,1,0)\mu_2 p_{21} \\
\pi(0,2,0)\mu_2 p_{21} = \pi(1,1,0)\mu_1 p_{12} \\
\pi(0,0,2)\mu_3 p_{31} = \pi(1,0,1)\mu_1 p_{13} \\
\pi(1,1,0)(\mu_2 p_{21}+\mu_1 p_{13}+\mu_1 p_{12}) = \pi(0,2,0)\mu_2 p_{21} + \pi(2,0,0)\mu_1 p_{12} + \pi(0,1,1)\mu_3 p_{31} \\
\pi(1,0,1)(\mu_3 p_{31}+\mu_1 p_{12}+\mu_1 p_{13}) = \pi(0,0,2)\mu_3 p_{31} + \pi(0,1,1)\mu_2 p_{21} + \pi(2,0,0)\mu_1 p_{13} \\
\pi(0,1,1)(\mu_3 p_{31}+\mu_2 p_{21}) = \pi(1,1,0)\mu_1 p_{13} + \pi(1,0,1)\mu_1 p_{12}
\end{gathered} \tag{6}$$

While service times are exponentially distributed, the equations (6) can be solved using direct numerical methods, as it is shown in (Hayes and Ganesh Babu 2004), (Radev et al. 2005). As we can see, the local balance of three-queueing closed network has to be searched in two-dimensional space. The steady-state probability vector can be determined with use of clustering and neural modeling, while cluster centers are correspondent to steady-states of Markov chain. The waiting time in queueing network is considered as such continuous stochastic time-series, in which mean value $\overline{W}$ is correspondent to mean queue length $\overline{Q}$. The ground for this suggestion is given in the Little's theorem, presented in (7).

$$\begin{aligned} \overline{K} &= \lambda \overline{T} \\ \overline{Q} &= \lambda \overline{W} \end{aligned} \tag{7}$$

The quantity $\overline{K}$ is mean number of jobs in queueing system, $\overline{T}$ is mean response time, and λ is throughput.

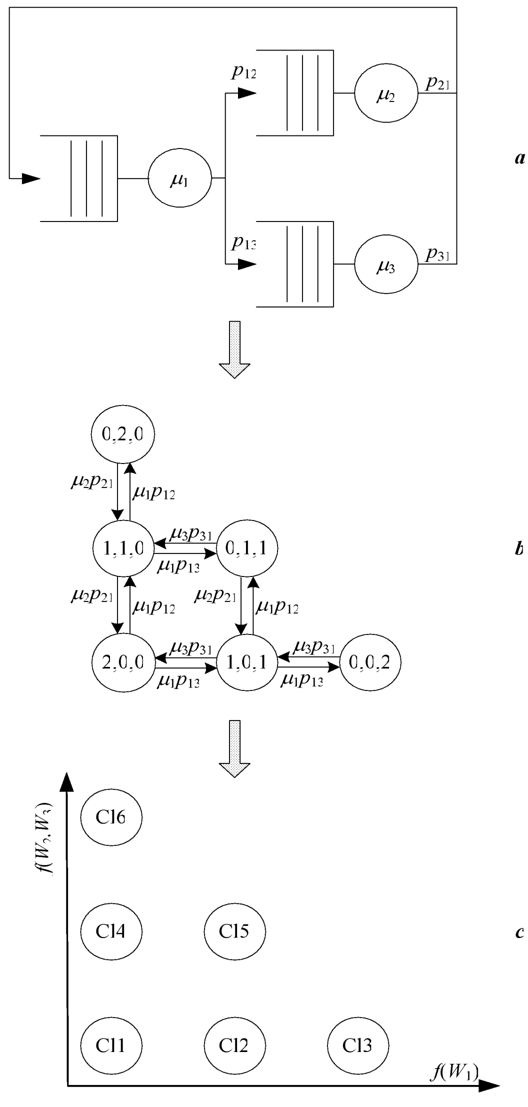


Figure 1: A Three-Node Closed Queueing Network (*a*), State Transition Diagram (*b*), and Clustering Diagram (*c*)

Since the departure rate is equal to arrival rate λ for a queueing system in statistical equilibrium, the throughput is given as $\lambda = m \cdot \rho \cdot \mu$. The response time T is waiting time plus service time, and its mean value is calculated as $\overline{T} = \overline{W} + 1/\mu$. The distribution functions of the waiting time and the response time are also important. As demonstrated in Fig.1*c*, the stochastic network is arranged with asymmetric nodes, therefore it makes sense to use self-organizing map model for clustering and to take advantage of the fact, that the solution for local balance can be obtained for non-Markovian queues as well, e.g., for non-exponential service times.

SELF-ORGANIZING NEURAL MAP AND CLUSTERING PROCEDURE

Self-organizing vector quantifications, which conserve topographic relations between centers, are particularly useful in communications, because noise added to the coded vectors may corrupt the representation on some level; and the topographic mapping ensures that a small change in code vector is decoded as a small change in attribute space, and hence, the small output change. Self-organizing neural networks have one-layered neural competitive structure, which can learn how to detect regularities and correlations in the input patterns. The neural maps learn both the distribution and the topology of the input vectors, to be capable to recognize neighboring clusters in the attribute space. Kohonen's network algorithm (Kohonen 1997) provides a tessellation of the input space into patches with corresponding code vectors. It has an additional feature that the centers are arranged in a low dimensional structure (usually a string, or a square grid), such that nearby points in the topological structure (the string or grid) map to nearby points in the attribute space. The Kohonen learning rule is used when the winning node represents the same class as a new training pattern, while a difference in class between the winning node and a training pattern causes the node to move away from training pattern at the same distance. In training, the winning node of the network moves towards that training pattern, and this leads to a smooth distribution of the network topology in the non-linear subspace of training data. In two-dimensional output space a map is expected, being correspondent to the k-dimensional array of output neurons C_j, which can be one- or two-dimensional. The connection between n-dimensional input vector and k-dimensional output neural vector is created through the weight matrix **W**. That output neuron j^* is selected as the winner in the competitive learning, for which the weight vector is closer to the current input according to (8).

$$\begin{gathered} \left| W_{jm}^{*} - X_m \right| \le \left| W_{jm} - X_m \right| \\ \forall j \in [1,\ldots,k], \forall m \in [1,\ldots,n] \end{gathered} \tag{8}$$

The Kohonen learning rule is different from vector quantification rule, and it is determined according to (9).

$$\Delta W_{jm} = \xi \wedge \left(j, j^*\right)\left(X_m - W_{jm}\right) \quad (9)$$

The neighborhood function $\wedge(j, j^*)$ is equal to 1 if $j=j^*$, and it decreases with increase of distance between neurons j and j^* in input space. The neurons, which are close to the winner j^*, change their weights more quickly than remote neurons, for which the neighborhood function is very small. The topological information contents in the fact that close neurons are changed almost in the same way; and this corresponds to the neighboring input patterns. The learning rule (9) is pulling the winner's weight vector to the point X_m. The self-organizing map is supposed to represent an elastic set in the input space, which can never be moved maximum close to the input values (Radev and Radeva 2004). The set has topology of attribute space, and its points have coordinates similar to weight vectors. Then we use a clustering procedure, which consists of following steps:

Step1 Definition of SOM and cluster centers

- For each cluster class Cj neurons-winners j^* are determined. That output neuron j^* should be selected as the winner, for which the weight vector is the closest one to the current input.
- In Kohonen learning rule the neighborhood function (j, j^*) should be equal to 1, if $j=j^*$, and it should decrease with increase of distance between neurons j and j^* in input space.

Step2 Iterative clustering of patterns

- Implemented axial and distance clustering are consecutive. Primary calculation of the number of patterns n_{j^*}, which belong to the cluster zone, should be made.
- Verification of classes patterns number $n_{j^*} < n_j$. If $n_{j^*} < n_j$ then calculation of Manhattan distances M_j for the rest of patterns should be made according to (10).

$$M_j = \sum_{i=1}^{n} \left\| x_{ij} - s_j \right\| \quad (10)$$

Step3 Optimal SOM clustering

- Recalculation of target values in each cluster should be made without restricting their number.
- Target values in optimal clusters should be proportional to probability density of clusters.

In the process of learning and training the neural maps learn both the distribution and the topology of the input vectors, and become capable to recognize neighboring clusters in the attribute space. In training, the winning node of the network, being the closest node in the input space to the given training pattern, moves towards that training pattern, while dragging with its neighboring nodes in the network topology, and this leads to its smooth distribution (Radev et al. 2004). Let's consider closed queueing network from Fig. 1, in which the sum of queueing waiting times $W_1 + W_2 + W_3$ is proportional to the steady-state distribution, and $W_1 + W_2 + W_3 = \overline{W} < 7$s. Two inputs of one-layered neural structure with 6 nodes are specified in time series with 300 discrete values. The waiting time W_1 with normal distribution with coefficient of variation $\sigma = 1.05$ is set in the first input. The difference $\overline{W} - W_1$ with exponential distribution with random variable parameter $\lambda = 2.5$ is set in the second input. This defined self-organizing map has 6 cluster centers, which are shown in Fig. 2 after 1000 epochs of training. According to described procedure, Fig. 3 demonstrates iterative clustered samples, which correspond to steady-state values of Markov chain. We find that the accuracy of clustering procedure increases significantly with increasing number of values in time series. When the Markov chain with great number of states is studied, a great increase in number of samples is needed. We should keep in mind, that for each position in the value of steady state probability at least 10 samples are needed. For example, for N = 3 and K = 8 users, the number of steady states is 45; and 4500 steady states should be investigated and clustered to obtain degree of accuracy $\varepsilon = 0.01$.

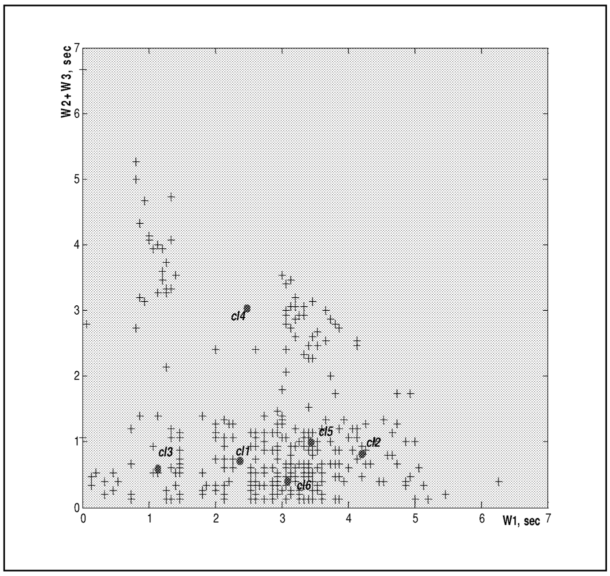


Figure 2: Definition of SOM and Cluster Centers

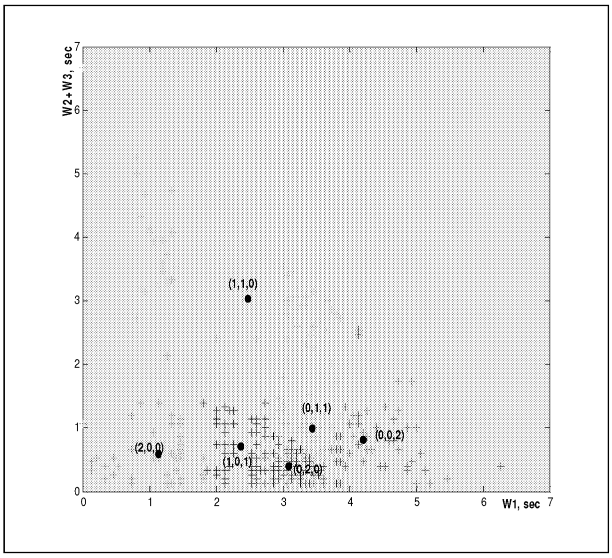


Figure 3: Optimal SOM Clustering

STEADY-STATE PROBABILITY VECTOR OF MARKOV CHAIN

Clustering procedure provided us with capability to calculate a set of discrete values of probabilities in cluster centers, which are called probability mass function (PMF). Each cluster is correspondent to strictly defined steady state, and (11) can be written according to (3).

$$\pi(1,0,1)+\pi(0,0,2)+\pi(2,0,0)+\pi(1,1,0)+\pi(0,1,1)+\pi(0,2,0)=1 \qquad (11)$$

The results of the example, which was considered in the paper, are shown in Table 1. The solution of system (6) is given for the following service rates: $\mu_1 = 4\text{sec}^{-1}$, $\mu_2 = 1\text{sec}^{-1}$, $\mu_3 = 2\text{sec}^{-1}$ and the routing probabilities $p_{12} = 0.4$, $p_{13} = 0.6$, $p_{21} = p_{31} = 1$, and this is made with purpose of comparison.

Table 1: Steady-State Probability Vector Distribution

Clusters	State	PMF	Exact
Cl 1	**(1,0,1)**	0.153333	0.123
Cl 2	**(0,0,2)**	0.16	0.148
Cl 3	**(2,0,0)**	0.143333	0.103
Cl 4	**(1,1,0)**	0.166667	0.165
Cl 5	**(0,1,1)**	0.16	0.198
Cl 6	**(0,2,0)**	0.216667	0.263

Exact comparative estimation of probability distribution cannot be obtained because the service time with exponential distribution was used for calculations. Nevertheless, we can see that suggested approach gives an opportunity for improving the qualitative analysis, receiving the steady-state solutions, and estimating the queueing networks equilibrium. The significance of this approach should increase with increasing number of jobs in queueing network. In the large-scale networks it is impossible to obtain numerical solution, and the cluster analysis with SOM neural model must be applied as soon as the input time series size is growing.

CONCLUSIONS

The estimation of steady-state probability vector of queueing networks requires complex computational procedures, and for the most of real large-scale communication networks this is impossible without simplification and restriction of the Markov chain. For the most of networks with non-Markovian queues or semi-Markov processes this task cannot be solved at all. The new approach to two-dimensional local-balance performance evaluation of queueing networks with use of clustering of one-layered self-organizing neural map was developed, while time series of waiting or response service times were used as inputs. The demonstrated technique, based on optimization of SOM clusters and calculation of steady-state probability mass function, provides satisfactory results, and this technique can be also used for non-Markovian queues and non-exponential service times.

REFERENCES

Bolch, G., Greiner, S., Meer, H, Trivedi, K. (1998). *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, John Wiley & Sons, New York.

Hayes, J., Ganesh Babu, T. (2004). *Modeling and Analysis of Telecommunications Networks*, John Wiley & Sons, New York.

Kohonen, T. (1997). *Self-Organizing Maps*, Second Edition, Springer Verlag, Berlin, Germany.

Radev, D., Radeva, S. (2003). Artificial Intelligence Modeling of Stochastic Processes in Digital Communication Networks, *Journal of Electrical Engineering,* vol. 54, N:9-10, 255-259.

Radev, D., Radeva, S. (2004). Kohonen Networks for Self-organizing Performance of Two Queues Markov Chains, *Ulmer Informatik-Berichte 2004-03: Poster Proceedings of the 27-th Annual German Conference of Artificial Intelligence,* Ulm, Germany, 189-203.

Radev, D., Lokshina, I., Radeva, S., Insinga, R. (2004). Neural Modeling of Link Occupancy Distribution for Broadband Telecommunication Transmission, *Proceedings of the DSI 2004: 35-th Annual Meeting of the Decision Sciences Institute*, Boston, USA, 3001-3009.

Radev, D., Lokshina, I., Radeva, S. (2005). Clustering Procedure for Performance Analysis of Mobile Communication Networks, *Proceedings of the EUROSIS-ETI 2005: 3rd Industrial Simulation Conference (ISC'05) at Fraunhofer IPK,* Berlin, Germany, 83-87.

Radev, D., Denchev, V., Rashkova, E. (2005). Approximations Algorithms for Steady-State Solutions of Markov Chains. *Proceedings of International Conference CompSysTech'2005*, Varna, Bulgaria.

BIOGRAPHY

DIMITAR RADEV, PhD is an Associate Professor at Department of Communication Technique and Technologies at University of Rousse, Bulgaria since 2001. His mainly research interests are connected with Teletraffic Theory, Simulation and Modeling of Communication Networks, Performance Analysis of Queueing Systems.

IZABELLA LOKSHINA, PhD is an Assistant Professor of Management Information Systems and Academic Coordinator of Management, Marketing and Management Information Systems at SUNY Oneonta. The positions she held before include Senior Scientific Researcher at the Moscow Central Research Institute of Complex Automation and Associate Professor of Automated Control Systems at Moscow State Mining University. Her main research interests are in area of Artificial Intelligence: Fuzzy Systems and Neural Modeling.

SVETLA RADEVA, PhD is an Associate Professor at Department of Computer-Aided Engineering at University of Architecture Civil Engineering and Geodesy, Bulgaria since 2001. She is mainly interested in Artificial Intelligence, Process and Product Modeling.

SIMULATION IN PROJECT SCHEDULING: WHY PERT IS NOT ENOUGH

Viktor Diamantas	Konstantinos Kirytopoulos	Vrassidas Leopoulos
National Technical University of Athens	University of the Aegean	National Technical University of Athens
Department of Mechanical Engineering	Department of Financial and Management Engineering	Department of Mechanical Engineering
Iroon Polytechniou 9	Fostini 31	Iroon Polytechniou 9
Athens, Greece	Chios, Greece	Athens, Greece

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

KEYWORDS
Project Management, Risk Management, Scheduling, PERT, Monte Carlo Simulation.

ABSTRACT

Various methods for modelling project duration have been developed so far. The Critical Path Method (CPM) is the mostly applied tool but it cannot handle uncertainty. Thus Program Evaluation and Review Technique (PERT) and Monte Carlo Simulation (MCS) are the most frequently used methods, when uncertainty should be taken into account. This study compares the results of the standard MCS with that of PERT as far as the estimation of project duration is concerned. Moreover, risk handling by the two approaches is also addressed. As a result the study defines the advantages and disadvantages of each one. Although theoretical oriented, the findings of the method are illustrated through a case study. The findings reveal that the modelling of risk is more robust when the MCS is used, thus leading to the conclusion that simulation may prove more efficient than the other stochastic methods.

1. INTRODUCTION

Construction projects are complex in nature and have many inherent uncertainties, that derive from the unique characteristics of each project, the diversity of resources and activities, as well as external factors. The project management community has defined all these uncertainty factors as risks (PMI 2004). These risks can affect the criteria against which the success of a project is usually measured, namely time (schedule), cost and quality (or performance). This paper focuses on the first of these three criteria and provides a comparative study, in order to come up with certain conclusions concerning the available methods for modeling and incorporating risks in project scheduling.

The core problem of schedule development is about defining each activity's duration as well as the right dependencies among project activities. Having achieved that, the project manager may then determine the start and finish dates of each project activity. The project is unlikely to finish on schedule if either the duration of activities or the definition of dependencies are not realistic.

The most important methods developed for project scheduling are the Critical Path Method (CPM - Maylor 2003), the Graphical Evaluation and Review Technique (GERT - Pritsker 1996), the Program Evaluation and Review Technique (PERT - Malcolm et al. 1959), and some simulation methods (Van Slyke 1963).

PERT is a well known method with proven value in managing complex projects. It was developed in 1958 by the US Navy Special Projects Office as part of the Polaris mobile submarine launched ballistic missile project (Malcolm et al. 1959). The development of PERT was based on the assumption that the duration of any activity within a project can be adequately modelled by the beta distribution. Initially, the mean and the variance of the duration, for each activity i, were estimated as:

$$\mu_i = \frac{1}{6} \cdot \left(a_i + 4 \cdot m_i + b_i\right) \quad (1)$$

$$\sigma_i^2 = \frac{1}{36} \cdot \left(b_i - a_i\right)^2 \quad (2)$$

Where a, m and b are the optimistic, most likely and pessimistic durations of activity i. Based on the Central Limit Theorem, the distribution describing project's duration is approximately normal, with the mean value of the project duration equalling to the sum of the means of the critical activities determined by calculations made according to the CPM. The variance of the project duration is the sum of the variances of the same CPM critical activities.

Several authors (Keefer and Verdini 1993; Premachandra 2001) have either modified the original PERT estimators or proposed new ones in order to estimate the activity time more precisely. Furthermore, authors like Johnson (1997) have proposed the use of the, convenient, triangular distribution, in order to adequately model activity duration. In fact, any distribution can be used to model activity duration, however, the beta or triangular distributions are the most commonly used, due to the, frequent, lack of sufficient data required to define a distribution with more complex parameters (Nasir et al. 2003).

An alternative to PERT is the Monte Carlo Simulation (MCS) which seems to be able to handle more complex project scheduling problems. Van Slyke (1963) was the first to suggest the use of MCS to find the cumulative distribution function of project network completion times. He also introduced the use of non-beta duration distributions. MCS is a stochastic method used to solve mathematical problems. It is now applied to a wide range of problems, such as econometrics, stock market forecasting, project duration estimation, etc. In MCS, the random selection process is repeated many times so as to create multiple scenarios. Each time a value is randomly selected for every variable of the objective function, a possible scenario is formed that leads to a certain outcome. This process is called an iteration The synthesis of all the iterations gives a range of possible outcomes, some of which are more and some other less probable. When many iterations are performed, thus many scenarios have been created, the MCS will come up with a distribution indicating all possible outcomes.

In its implementation, MCS can use the three point duration estimation approach to establish a statistical distribution of the duration of each activity (Wang 2005). Random numbers, representing duration, are generated for every activity from the selected distribution.

PERT and MCS can be enhanced by risk management, so that they take into account the effect of "expected" risks. Risk management was introduced in the 1990s as a structured process of project management. Small and bigger projects are characterised by the engagements the contractor undertakes and the particular profits that the customer expects upon the completion of the project. However, either due to the size or the complexity of the project, these engagements and profits are continuously revised during project's execution. Unexpected events that appear anytime during the project's life-cycle make these revisions necessary, and render risk management process essential (Koller 1999). The literature as far as the stages of project risk management are concerned is very extensive. In the end, every risk management approach provides an estimation of the effect of identified risks on activity or project duration. This effect of identified risks on the schedule of the project is a vital add-on to traditional scheduling.

2. PROBLEM FORMULATION AND METHODOLOGY

In order to illustrate the theoretical findings, a five step methodology was followed. The first step is the literature review of scheduling focusing on PERT and MCS. The outputs of this step are the theoretical findings summarised in **Section 3**.

The second step is the development of the "project model", which is, in our case, a construction project where the two scheduling tools would be used in order to validate the theoretical findings. In order to develop the model the major activities of the project and their dependencies must be identified by the project team. A Gantt chart of the project is created by entering these data in a project scheduling software. The final task within model development is the estimation of each activity's duration. The three point estimate approach is used, requiring the estimation of an optimistic (minimum), a most likely, and a pessimistic (maximum) value for the duration of each activity. These values are estimated by the project team, based on its knowledge, past experience acquired from similar projects, and the use of the corporate memory. The development of the project model is also an important step, since an inaccurate model will generate misleading results.

Two project durations are calculated. The first one, called "total duration" in this study, does not include the effect of discrete risks identified during the risk identification phase of risk management, while the second one, called "overall duration", does include them.

The third step is the conduction of risk management. The process suggested by Kirytopoulos et al. (2001; Leopoulos et al., 2002) is used. According to this process risk management consists of four basic phases, risk identification, risk assessment, risk mitigation and follow up. Therefore, any potential risk that may affect the project is identified by the project team, based on past experience and the corporate memory. The, most applicable, methods that may be used for risk identification are checklists, Risk Breakdown Structure (RBS), and brainstorming. The authors recognise the dual nature of risks as threats and opportunities, but identify only threats in order to follow the threat-oriented philosophy of the company which offered the case study. The second phase of the process, risk assessment, reveals risks' expected impact and probability of appearance. Based on their exposure (usually calculated as impact * probability) risks are prioritised. The third phase, risk mitigation, deals with the development of adequate response actions to confront the identified risks. Risk assessment was repeated after the mitigation phase so that the risks' expected impact and probability of appearance would take into account the mitigation actions developed.

Risk management is succeeded by the fourth step, conduction of the PERT analysis and MCS. The standard PERT approach is followed. Based on the developed model, the critical path of the project is identified, the mean value and variance of the duration for each activity that belongs to the critical path of the project is calculated, and finally the total duration's mean value and variance is calculated. The only, possible, way to take into account the effect of the identified risks, beyond the inherent uncertainty of each activity's duration, when PERT is used, is to add their exposure to the mean value of the total duration and thus to calculate the overall duration of the project. However, this approach suffers from almost unacceptable scientific redundancies.

Once PERT analysis has been completed the MCS model, based on the project model and results of the project risk management process, is created. This model takes into account the inherent uncertainty of each activity duration, the effect of the risks identified during risk management, and the correlation between activity durations. The triangular distribution, three point estimation, is used to describe the inherent uncertainty in the duration of each activity, while the effect of the identified risks follows the discrete distribution. For this reason these risks are called "discrete risks" in this study. Once the MCS model is fully developed the simulation is ran for 10,000 iterations and the results are recorded in a spreadsheet.

The final step of this method is the comparison of the results generated by PERT against the results generated by MCS, in order to identify whether the results validate the theoretical findings or not.

3. Theoretical Findings

The authors conducted a detailed literature review of scheduling focusing on PERT and MCS. Scientific papers, conference proceedings, case studies and books dealing with scheduling have been collected and conclusions regarding the use of PERT and MCS have been drawn. The following paragraphs summarise, the theoretical conclusions.

The main conclusion is that while PERT is a simple and straight forward method, requiring less time and resources than MCS, it suffers from major limitations. The removal of the paragraph regarding PERT from the 2004 version of the PMI PMBOK (PMI 2004) can be taken as an indication that practitioners show less interest to its use than in the past. The PMBOK cannot be considered a scientific bible of project management, but it clearly shows the shift of the practitioners' interest from PERT to simulation methods.

Furthermore, it was noted that PERT is limited by the constraints of the Central Limit Theorem (Nasir et al., 2003). PERT is obliged to model the duration of every activity of a project with the same statistical distribution and the duration of each activity should be an independent variable. It also ignores all "sub-critical" paths (Douglas, 1978; Nasir et al., 2003), taking into account a unique critical path as it is determined by deterministic calculations. In addition, PERT has difficulties in dealing with risks beyond the inherent uncertainties of activity duration. In fact, there is not any reference for the use of PERT in discrete risks handling. The authors tried to incorporate discrete risks in PERT and the project model came up with similar results with those generated by the MCS. However, the inclusion of discrete risks in PERT has a very weak scientific foundation. At last, PERT produces an overall expectation of the project duration without any indication regarding the causes of delays or the effect of identified risks.

As far as the MCS is concerned, it has been concluded that it solves most of the limitations that PERT faces. It is not based on the Central Limit Theorem, therefore activity durations may be modelled by different distributions (Van Slyke, 1963) allowing for the definition of the most applicable distribution for each activity instead of defining the most convenient distribution for a group of activities. Likewise, no statistical independence is assumed for activity durations. However, a major methodological deduction in traditional MCS is the assumption of statistical independence for individual activities which share risk factors with other activities. The problem was addressed by Van Dorp and Duffey (1999). MCS, also, takes into account potential changes in the critical path, successfully modelling the effect of "sub-critical" paths on the duration of the project (Douglas, 1978). Another critical advantage of MCS is that it can model the effect of risks beyond the inherent uncertainty of each activity's duration with a scientifically accepted approach. Finally, MCS clearly indicates the effect that each risk may have on the duration of the project.

MCS simulation is not a panacea though, as it is far more complex than PERT, requiring more time, special software, trained individuals and computer power, especially for large projects. Additionally, the modelling of activity durations and, especially, risks is not an easy task, and it can be considered almost impossible without the use of a risk registry and a corporate memory. Inaccurate modelling of activity duration and risks may result in erroneous results that can hinder the successful execution of the project.

4. CASE STUDY

4.1 Description of the project

The presented project was a medium scale construction project located in the Greek island of Crete and undertaken by a major international contractor based in Greece. It dealt with the construction of several independent residential buildings.

The initial Gantt chart of the project plan consisted of 85 activities, using multiple levels of detail mainly for the estimation of the project cost. The pessimistic (maximum), most likely and optimistic (minimum) duration of each activity has been estimated by the project team, based on their knowledge and past experience acquired from similar projects. The deterministic duration of the project, based on the critical path and the estimated most likely duration for each activity, was calculated to be 281 days.

4.2 Project Risks

As described in the methodology section a risk management process was applied to the project. The identification of possible risks relied primarily on the use of checklists and nominal group sessions (Chapman 1998). Furthermore, the process was enhanced by the use of the corporate memory which has the form of a Risk Breakdown Structure (RBS).
The RBS is defined as "A source oriented grouping of project risks that organises and defines the total risk exposure of the project. Each descending level represents an increasingly detailed definition of sources of risk to the

project" (Hillson 2002). A part of the RBS is shown in **Figure 1**.

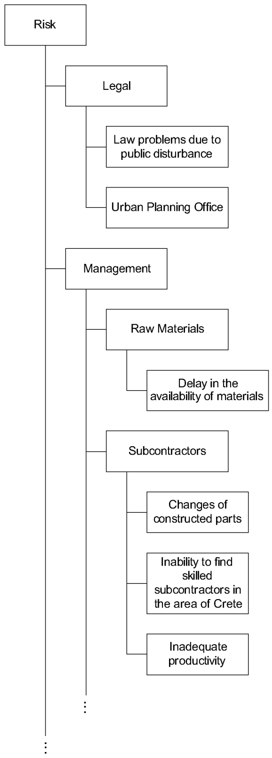


Figure 1: Risk Breakdown Structure

The probability of appearance and impact in days of each identified risk, were initially quantified during the risk assessment phase. The risk assessment phase was repeated after the development of mitigation actions and a selection of the final results are shown in **Table 1.**

4.3 Analysis of the results

The results of the deterministic approach, 281 days, correspond to a 10% confidence level in PERT total duration, a less than 5% confidence level in MCS total duration, and is lower than the minimum value of PERT and MCS overall duration.

In **Figure 2** the cumulative density function that derives from PERT (continuous line) is compared with the cumulative density function that derives from the MCS, without taking consideration of the discrete risks (dashed line). This outcome is in line with Diaz and Hadipriono (1993) conclusions that the MCS produces more conservative, results than PERT.

Table 1: Discrete Risk Quantification

Risks	Prob	Impact (Days)	Exp (P*I)
Earthquake	15%	7	1,05
Inability to find skilled subcontractors in the region of Crete	30%	7	2,1
Delay in the availability of materials	35%	14	4,9
Law problems due to public disturbance	60%	7	4,2

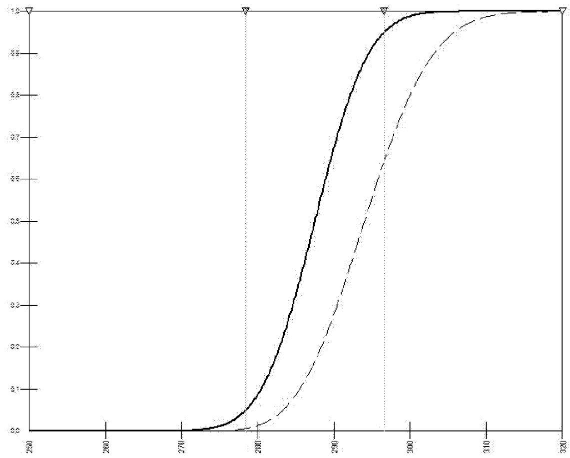


Figure 2: PERT vs. MC (dashed) Total Duration

Figures 3 and **4** show that scheduling of a project without taking into account the discrete risks may result in a very optimistic estimation of the project duration, both in PERT and MCS.

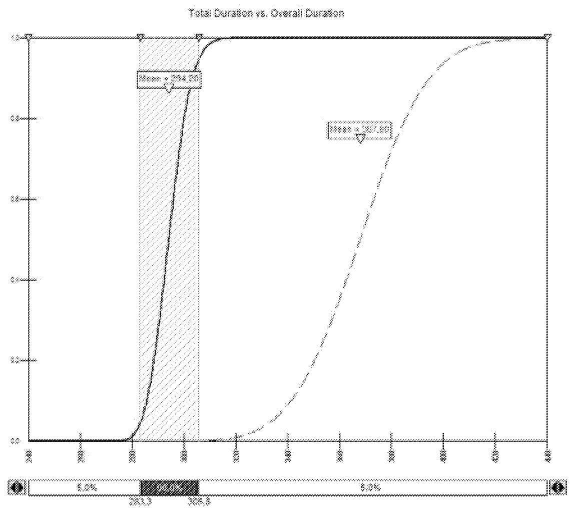


Figure 3: MC Total vs. MC Overall (dashed)

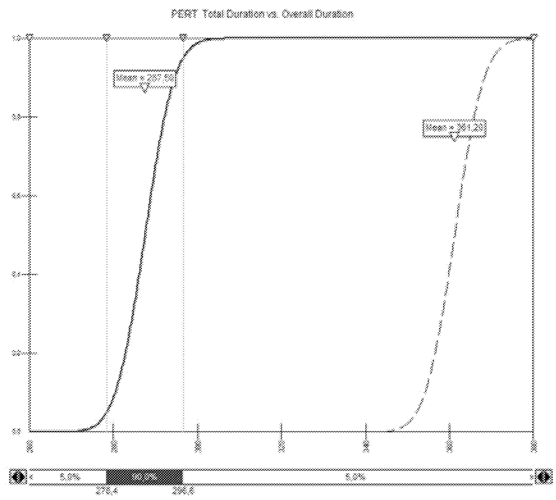


Figure 4: PERT Total vs. PERT Overall (dashed)

Figure 5 shows that the inclusion of risks in PERT analysis by adding the exposure of risks to the mean value of the duration distribution gives same results with the MCS, as far as the mean value of the duration is concerned. However, PERT has provided a much lower standard deviation than MCS which may prove misleading.

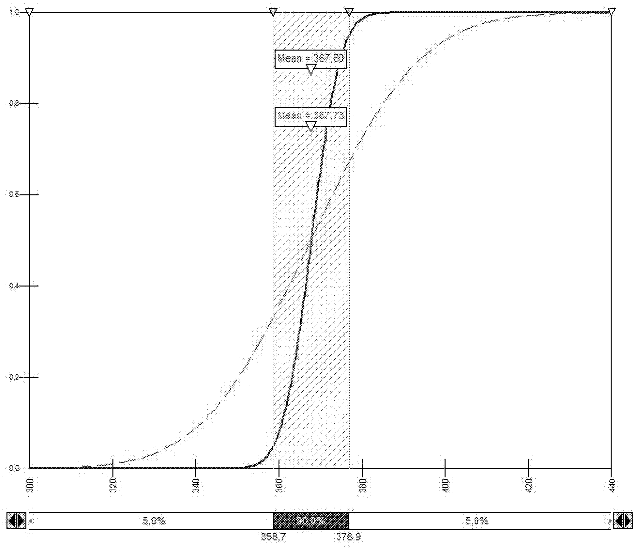


Figure 5: PERT vs. MC Overall Duration (dashed)

5. CONCLUSIONS – FURTHER RESEARCH

The objective of this study was to compare the capabilities of PERT with those of standard MCS as far as the estimation of the project duration is concerned. Analysis took into account risk modelling, as well. Risk modelling is considered an important scheduling factor so its inclusion in the study was of great importance. The literature study conducted about PERT and MCS lead to the conclusion that the modelling of risk is more robust when the MCS is used and that MCS gets over PERT's limitations. The theoretical findings were validated and illustrated through a case study.

Although the findings of the paper can be supported by a pure theoretical approach, it would be interesting to second the arguments presented here with more case studies. The outcome of this new research is expected to validate theory, as it happened with the case study that was presented in this paper.

Another area for further research are the new simulation methods based on MCS for the estimation of activity and project duration developed by several authors (Dawood, 1998; Oztas and Okmen, 2005). The results of PERT and traditional MCS in the new case studies can be compared with the results of the new methods in order to identify advantages and disadvantages and validate the results of those new methods.

REFERENCES

Chapman R., 1998, "The effectiveness of working group risk identification and assessment techniques", International Journal of Project Management, Vol. 16, No 6, pp. 333-343.

Dawood N., 1998, "Estimating project and activity duration: a risk management approach using network analysis", *Construction Management and Economics*, Vol. 16, pp. 41-48.

Diaz C.F. and Hadipriono F.C., 1993, "Non deterministic networking methods", *ASCE Journal of Construction Engineering and Management*, Vol. 119, No. 1, pp. 40-57.

Douglas D.E., 1978, "PERT and simulation", *Proceedings of the 10th conference on Winter simulation Miami Beach, Florida*, 4-6 December, 1978, pp. 89-98.

Hillson D., 2003, "Effective Opportunity Management for Projects: Exploiting Positive Risk", Dekker, USA.

Johnson D., 1997, "The triangular distribution as a proxy for the beta distribution in risk analysis", *The Statician*, Vol. 46, No. 3, pp. 387-398.

Keefer DL, and Verdini WA., 1993, "Better estimation of PERT activity time parameters", *Management Science,* Vol. 39, pp. 1086-1091.

Koller G., 1999, "Risk assessment and decision making in business and industry: a practical guide", CRC Press.

Kirytopoulos K., Leopoulos V., Malandrakis Ch., "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.

Leopoulos V., Kirytopoulos K. and Bellos V., 2002, "A model for product risk estimation through corporate memory and techniques integration", Advances in Multimedia, Video and Signal Processing Systems, WSEAS Press, pp. 302-307.

Malcolm DG, Rooseboom JH, Clark CE, and Fazer W., 1959, "Application of a technique for research and development program evaluation", *Operations Research*, Vol. 7, pp. 646-669.

Maylor H. 2003, *Project Management*, Pearson Education, Harlow.

Nasir D., McGabe B., and Hartono L., 2003, "Evaluating Risk in Construction – Schedule Model (ERIC-S): Construction Schedule Risk Model", *Journal of construction engineering and management*, Vol. 129, No. 5, pp. 518-527.

Oztas A., and Okmen O., 2005, "Judgemental risk analysis process development in construction projects", *Building and Environment*, Vol. 40, pp. 1244-1254.

PMI, 2004, "A Guide to the Project Management Body of Knowledge", Project Management Institute, USA.

Premachandra I.M., 2001, "An approximation of the activity duration distribution in PERT", *Computers & Operations Research*, Vol. 28, pp. 443-452.

Pritsker A.A.B., 1966, "GERT graphical evaluation and review technique", Memorandum RM-4973-NASA, Rand Corporation.

Van Dorp J.R., and Duffey M.R., 1999, "Statistical dependence in risk analysis for project networks using Monte Carlo Methods", *International Journal of Production Economics,* Vol. 58, pp. 17-29.

Van Slyke R.M., 1963, "Monte Carlo methods and the PERT problem", *Operations Research*, Vol. 11, No. 5, pp. 839-860.

Wang W.C., 2005, "Impact of soft logic on the probabilistic duration of construction projects", *International Journal of Project Management*, Vol. 23, pp. 600-610.

SUSTAINABLE DEVELOPMENT

MODELLING AND SIMULATION OF CAR DISMANTLING FACILITIES.

G.M.Acaccia, R.C.Michelini, L.Penzo, N.Qualich
PMAR*lab*- Laboratorio di Progettazione e Misure per l'Automazione e la Robotica
DIMEC- Dipartimento di Meccanica e Costruzione delle Macchine - Università di Genova
Via all'Opera Pia 15A - 16145 Genova – Italy
{acaccia|michelini|penzo|qualich}@dimec.unige.it

KEYWORDS.
Discrete, decision support, computer-aided analysis, dynamic modelling, event-oriented

ABSTRACT.
Reverse logistics is addressed as new business issue promoted by the compulsory targets regulation enacted by the EU, to lower pollution and to reduce raw stuffs consumption. The paper gives a survey of the existing legal frame, and suggests the modelling and simulation paths to assess the dismantling process performance.

INTRODUCTION.

The EU environmental policy is conditioning spur to aim at preserving the quality of life of future people. Recovery is fostered as antidote of the affluent society, with the inherent task of educating consumers, towards more conservative behaviours. Therefore, broadband actions are chosen, such as the ones dealing with end-of-life vehicles, ELV, mass-produced goods, typically, feeding the replacement market, with high eco-impact. Conservativeness is achieved through the carmakers responsibility, setting out regulations with enforceable targets and visible fees, easy to run and control. The ELV regulation provides explanatory hints to this trend. The 2000/53/EC Directive, enacted 18.09.2000, and acknowledged by the Member States before April 2002, defines the rules to be followed by national organisations for vehicles dismissal, and by automotive producers for selling cars respectful of legal requests. The Commission further modified and integrated the rules by special specifications, e.g. the 2003/138/EC, to establish standard codes for parts and materials. It needs to be mentioned that the EU is concerned by *final issues*, not about *how* these achievements are obtained by Member States. The documentation leads to a few remarks, such as:

- regulation addresses environment protection, by explicitly conferring responsibility to the producers;
- design for disassembly, for recover, for recycle, etc. practices shall become standard requirement;
- extensive resort to re-manufacturing, to re-use of parts, materials, etc. is promoted and stimulated;
- appropriate aids (modularity, identifying codes, etc.) are suggested to make dismantling easier, etc.;
- fixed lifecycle monitoring and reporting disclose on-duty conformance-to-specification checks;
- not-justified high-impact and polluting behaviours are heavily taxed or totally forbidden.

The remarks show that business paradigms changes are fostered, based on new design patterns, with concern on *point-of-service* performance and full commitment for *withdrawal*. The focus on growth sustainability is basic driver, and, obviously, it leads to reverse logistics entrepreneurship, with larger role for competitiveness played by operators in the backward cycle. The present study is concerned with the recover, reuse and recycle opportunities, and it specifically addresses topics in end-of-life vehicles, *ELV*, processing, as they follow from the environmental policy enacted by the EU.

THE REFERENCE FRAME.

Eco-consistency by reverse logistics entrepreneurship does not develop by itself, as the backward chain outcomes do not have explicit purchasers, lacking links of actual needs to them. Then, to separate backward, from forward chain leads to deviating issues, keeping benefit within a set of parties, but damaging third (and future) ones. The cost/profit ratio is to be assessed balancing the profit of the consumers side (producers and users), *against* the protection of all people not involved by the specific value chain. In that sense, regulation for eco-consistency is peremptory task that governments need undertake, enacting rules to charge the consumers side for environmental impact linked to the all supply chain, including dismissal and recovery. The automotive field is paradigmatically noteworthy, with widely spread market of *registered* items: falls-off affect large amount of users; end-of-life vehicles are individually acknowledged and recorded. Thereafter, the EU approach specifies mandatory incumbents, by sets of directives, notably:

- member states shall establish collecting systems for exhausted vehicles and parts, at authorised sites, where preliminary treatments will grant safety and security fitting out, by removing noxious and harmful parts;
- withdrawal needs to be assured without charge on the final owners (prescription to be fully enabled from 01.01.2007), but included in the product-service delivery, as inherent attribute;
- users co-responsibility could be invoked, for special non-conservative behaviours, when critical pieces are damaged, removed or modified, altering the original setting of the supply;
- the dismantling and destruction incumbents ought to be certified, with full assessment of recovery parts,

re-cycled materials, thermo-recovery and residuals dumping, to be notified to the EU.

The enacted compulsory targets are quite severe, say:

- from 01.01.2006: 85% by weigh of the vehicle ought to be recovered or recycled, 10% can be dumped to landfills (after suitable neutralisation) and 5% can be used as auxiliary fuel;
- from 01.01.2015: figures are modified, allowing 10% for fuel use, but only 5% to landfills.

By now, these scenarios are hardly outlined by the EU directives, basically enacted for pollution remediation. The mandatory targets distinguish: *recovery* (up to 90-95 %, by contrast with *dump* residues, e.g., as low as 10-5 % in weigh), from *reuse* (second-hand market, possibly, meaningful fractions) and from *recycle* (*secondary* materials, replacing *raw* provisions). The recovery, which establishes on such targets, favours the expansion of the materials, rather than energy, markets, due to quite severe limits on landfill dumping.

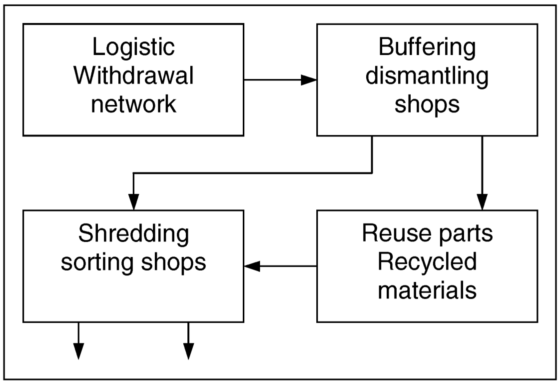


Fig 1. Reverse logistics material flow.

The approach, then, highlights the critical nature of the disassembly process, after *ELV* withdrawal, for removal of dangerous liquids and parts and safety fitting of the whole material flows. On these facts, recovery characterises by, Fig.1:

- the need of a logistic network, with collecting and transportation equipment, storage points, handling and processing fixtures, and joint information flow for acquisition and recording;
- the establishment of proper dismantling shops for safe pieces/parts withdrawal, and of suitably located shredding facilities, to grind the left-out hulks to tiny pieces;
- the resort to specialised sorting plants, to separate the different metals (ferrous alloys, stainless steels, brass, aluminium alloys, etc.), from glass, plastics, etc.;
- the development of reuse-parts and recycle-materials tracks, with enhancement of the design, manufacture and maintenance by second-hand provisioning practice.

THE *ELV* DISMANTLING.

Developing effective car dismantling and recycling facilities is urgent challenge in the carmakers field, with, probably, the most demanding innovations on the automotive systems to overcome the world-wide competition. Several investigation lines have been undertaken these recent years, aiming at:

- quantitative assessment of recover/reuse/recycle processes, under different operation plans, for existing situations (vehicles not produced with *recycling in mind*) and alternative organisations and technologies;
- establishing proper recommendations and warnings to improve (by design-for-dismantling, d*f*d, design-for-recovering, d*f*r, etc., practices), the future, more eco-conservative, automotive market;
- balancing *forward* and *backward* chain incumbents to optimise return on investment of the manufacture business, based on lifecycle engagements, driven by voluntary agreements and compulsory targets;
- reconsidering the reverse logistics as mandatory requirement for growth sustainability, in front of natural resources ceaseless depletion and human surroundings progressive pollution.

The direct investigation and performance assessment of the dismantling facilities will, thereafter, become standard routine. The well-established approach, for shop-floor selection, goes across implementing suitable software aids, to describe the equipment behaviour and simulate actual operation achievements. The process physical model shall, first, move from defining work-cycles and work-stations. Indeed: to *dismantle* means to take to pieces an artefact, or to strip/deprive of its outfits; while: to *disassemble*, means to systematically separate an artefact into its constitutive pieces. Efficiency would lead to prefer the second process.

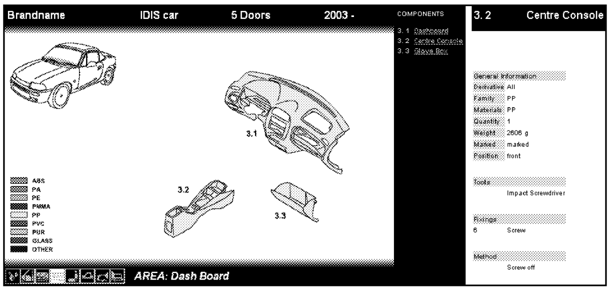


Fig 2. Example IDIS output.

Thereafter, the disassembly process should consider different features: - manual *vs*. automatic (*also*: mixed mode operation); - partial (a subset of parts are not removed), selective (a set of parts are removed the first); - parallel (two or more parts are jointly removed), sequential (parts are removed one by one); - non destructive (unbroken parts are withdrawn) *vs*. destructive (parts are cut/broken during removal); - and so on. Today, basic knowledge of car-wreckers comes from IDIS, International Dismantling Information System, http://www.idis2.com, a data-base of some 1 000 vehicle types of 25 carmakers, listing about 46 000 pieces. This helps displaying the part solid models and describing texts (in 21 languages), with instructions and tool/fixture requirements, Fig. 2.

For environmental protection, the enacted regulations require special care to handle hazardous materials and to safely avoid contamination. This leads to sets of preliminary *ELV* treatments, to remove potentially harmful or noxious parts (containing lead, mercury, etc.), fluids (for brakes, lubrication, conditioning, etc.), or items (air-bags, exhaust silencer, etc.), etc., which are subject of special restrictions. The subsequent steps address the disassembly-for-reuse duty, looking after separating pieces to be restored or remanufactured for second-hand marketing; for competitiveness, the safety and reliability figures need comply properly high standards. At this stage only, the dismantling-for-recycle should start. The process sequencing aims at clearly separating items into homogeneous blocks as for material properties and recovery potentials, namely:

- outer outfits and glasses (windscreen, window-panes, etc.), for special purpose recover/recycle;
- special pieces (doors, front/rear bonnet, etc.) and groups (gearbox, engine, etc.), for possible recovery;
- plastic parts, possibly, properly ordered according included markers or acknowledgement identifiers;
- internal outfits and fabrics (covers, upholstery, stuffing, etc.), following scheduled requirements;
- metallic parts, possibly, keeping (or re-establishing) the ordering trim of the pertinent data-base;
- residual crock-skeleton/wreck, for the subsequent (appropriate) compacting/shredding operations.

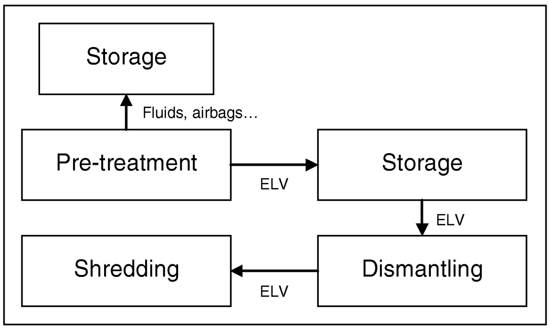


Fig 3. Example dismantling shop.

The detailed description of the work-cycles and accurate development of structured models are basic premise for choosing the plant and assessing the usefulness of robotic aids.

The state of the art is fast evolving, with example plants and prototypal facilities with different levels of sophistication. For instance, the Canadian AADCO line, http://www.aadco.ca/, is valuable reference, when productivity is special request, or the German ADEMA system, http://www.lsd-gmbh.com, is useful support, when leanness and flexibility are primary objectives. Basically, the on-progress trends look after modularity, with series of automatic units (fluids draining, wheels removal, rim/tyre separation, *ELV* lifting or overturn, gearbox/engine pulling out, etc.), and suitable handling and transport equipment, between the work-stands. This, quite commonly, leads to mixed-mode flows, with interposed human operators and robotic devices, as, typically, subsets of duties require careful decision-making (e.g., selective handling of hazardous materials, quality figures of reusable components, etc.) and automatic processing would result unreliable, at least, at the present state of the art. The functional modelling of the disassemble/dismantle line, thereafter, ought to deal with hybrid algorithmic/procedural knowledge frames, which include causal blocks, to duplicate the physical transformations, and heuristic blocks, to emulate the behavioural counterparts. The obtained models directly yield to *expert simulation*, providing powerful means to experiment on *virtual* plants, during facility design, or to compare *virtual* process-plans, for on-duty operation, Fig 3-4.

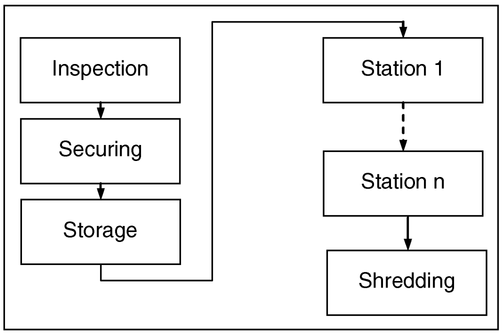


Fig 4. Example dismantling process.

For automotive systems, the exemplified investigation starts by the quantitative assessment of the recover/ reuse/recycle processes, to comply the compulsory targets of the directives. The market size is impressive, with 225 million vehicles (185 million cars) circulating, in 2001, in the original 15 EU Countries and, each year, 10 million cars, moved to landfills. The regulations affect national Governments, and *fair* law transposition is required, to not bias the common market. Thus, national, regional and local set-outs should be devised, properly tailored to current needs and entrepreneurship qualification. The Italian case shall face, for instance, some 2 million *ELV* per years, with, notably, around 10 000 *ELV* each year, to be processed in Liguria; the environmental bylaws are ruled on the regional scale, but, of course, other collection/processing settings could develop, should be achieved higher return on investment. Today, the car-wreckers represent quite distributed realities, with tiny processing abilities, scarcely complying the eco-protection requests. At the same time, overall inverse logistics networks are far from fully operational deployment, as for carmakers side, as for official monitoring. The law, DL 24.06.2003, suitably transposes the 2000/53/CE directive, but factual specifications still lack, with, e.g., *free-take-back* burdened by severe de-registration fees and loose carmakers interpretations. Last, but not least, the regional and local authorities, most of the time, seem to do not entirely realise the technical relevance of the overall incumbents and related business.

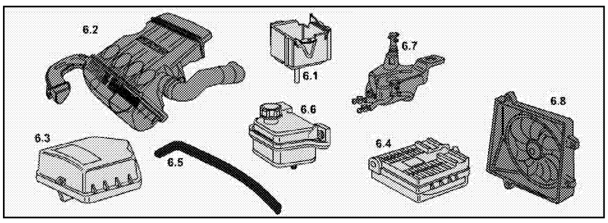


Fig 5. Example engine components.

EXAMPLE SIMULATIONS.

The recovery/reuse/recycle processes, rather than academic study or technology-driven issues, are urgent engagement, with fully specified legal (directive mandatory rules) and technical (available dismantle/ recycle equipment) constraints. Difficulties, due to lack of profitable markets for reverse logistics outcomes, are got over, by combining the pertinent economic instruments, namely, *free-take-back* and *visible fees* management on the artefacts lifecycle. This means: - careful effectiveness assessment between actual solutions, as *decision aid* for the organisational choice and setting; - complete transparency of forward and backward supply chain, with optimal allocation of the *visible fees*, not to penalise the national/regional/local competitiveness. Occurrence-driven expert-simulation helps reaching both the scopes: as off-process tool, at the design stage; as on-process support, at lifelong operation. On these premises, the PMAR*RLelv* code has been considered to simulate the *ELV* reverse logistics [1]. Basically, it addresses problems in automotive systems, for withdrawal, collection and dismantling end-of-life vehicles, and for shredding, sorting and dispatching the residuals; it is based on the WITNESS language, and it incorporates modules for intelligent data-vaulting, interactive remote inquiry and data reporting with text/graphic restitution, Fig. 5-6.

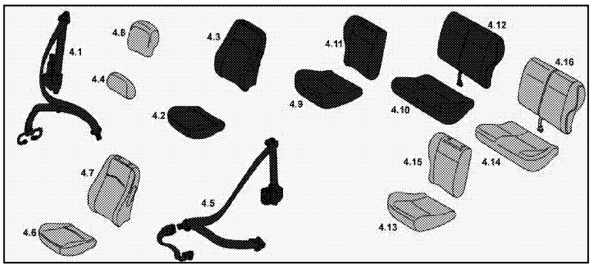
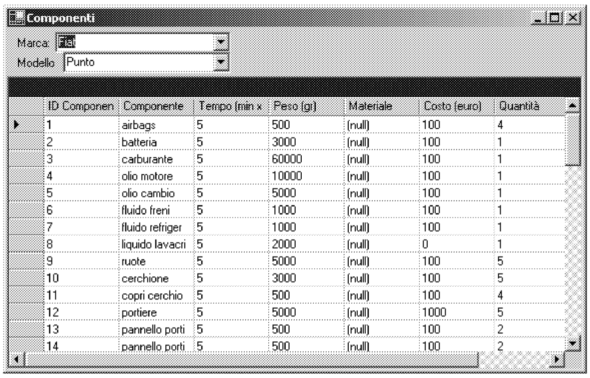


Fig 6. Example upholstering and stuffing.

For explanatory purpose, the example case, focused on *ELV* dismantling, is discussed. The Liguria situation is taken in consideration, singularly and in relation to broader national set-ups. Generally, wastes (and, with full evidence *ELV*) are bulky, potentially dangerous and very poor materials, thus, every added costs in handling and transportation will lower the process effectiveness. This suggests to localise the dismantling shops on the territory, possibly, with reusing/recycling flows linked to decentralised end-users. Modular lay-outs, based on the ADEMA concept, are arranged, distinguishing the separate stands (safety fitting; disassembly for reuse, dismantling), the transport and manipulation devices and the auxiliary storages (line-buffers and parts collectors). A reasonable setting deals with local facilities conceived to process 10 000 *ELV* per years, roughly leading to five lines in Liguria (200 in Italy), according to the expected market.

The simulation runs considered different feeding scenarios, referred to daily, weekly and monthly agendas, with due parameters setting and alternative end-users flows. The mixed mode automation is evaluated, with different labours (on-line personnel) or duty-oriented robots. On-progress schedules include discontinuities (failures, supply stops, etc.) by occurrence management. The example outputs show benefits and drawbacks of the decentralised lay-outs, and provide reference hints to acknowledge the main operation details of the investigated solutions; by the purposely implemented **WITNESS** modules Fig 7, appropriate performance figure and process statistics Fig 8-9, are provided as decision aid support.



Componenti

Marca:

Modello: Punto

ID Componen	Componente	Tempo (min x	Peso (gr)	Materiale	Costo (euro)	Quantità
1	airbags	5	500	(null)	100	4
2	batteria	5	3000	(null)	100	1
3	carburante	5	60000	(null)	100	1
4	olio motore	5	10000	(null)	100	1
5	olio cambio	5	5000	(null)	100	1
6	fluido freni	5	1000	(null)	100	1
7	fluido refriger	5	1000	(null)	100	1
8	liquido lavacri	5	2000	(null)	0	1
9	ruote	5	5000	(null)	100	5
10	cerchione	5	3000	(null)	100	5
11	copri cerchio	5	500	(null)	100	4
12	portiere	5	5000	(null)	1000	5
13	pannello porti	5	500	(null)	100	2
14	pannello porti	5	500	(null)	100	2

Fig 7. **WITNESS** modules structured input.

The concept, behind spreading on the territory local dismantling shops, leads to establish synergies in the reverse logistics of the different durables/consumables covered by the EU mandatory take-back policy. The waste electrical and electronic equipment, *WEEE*, case, already in force by mid 2005, is partially different, as it deals with non-registered goods. Without entering into details, producers/dealers, again, are required for items withdrawal, collection and dismantling; the recover and recycle targets apply, depending on appliances size and lighting rigs; a *visible fee* is collected when a new device is sold. The EU prescriptions are differently transferred into national acts, and some country is still considering transient alternatives for some legal (data registration and vaulting, etc.) and some technical (withdrawal, dismantling, recycling, dumping, etc.) incumbents.

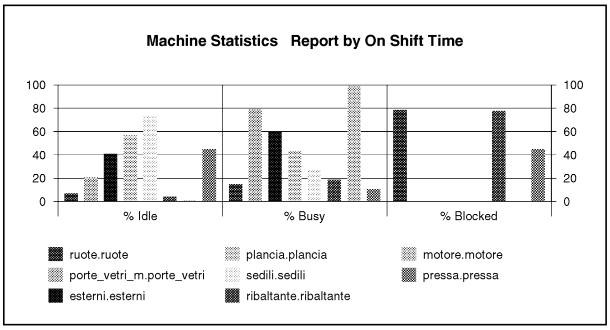


Fig 8. **WITNESS** modules example output.

The role of the manufacturers is fundamental to support reverse logistics by product-data management aids for enhanced dismantling and selective recycling, and turns to be critical, in the short time, to enable design-for-recycling paradigms (modularity and material segregation, disassembly pre-setting, etc.) and to profit of sustainable manufacture (high resort to recovered or recycled provisions, etc.) and maintenance (proactive up-keeping, re-integration, etc.) options. Besides, governmental authorities need to analyse the whole supply chains, which distinguish one producer from another, suitably defining pertinent taxing rule.

The costing will be ascribed to the consumers side, with tax collection duty, mainly, put in charge by the goods sellers, at the *point-of-sale*. The availability of local multi-task dismantling shops, capable of taking charge of the different backward chains and feeding the subsequent shredding, sorting and dispatching flows, becomes opportunity for eco-conservativeness.

The connection between backward chains has relevant falls-out. The complexity of problems needs not hinder the obligation that sustainability shall be put directly in charge to the parties having benefits, and not poured on unrelated peoples, if the impact on the environment ought to be kept under acceptable limits. The present EU policy is, certainly, at a very early stage; the enacted bylaws, however, already aim at:

- acting on the consumers side, to foster a repair society, instead of a throw-away economy;
- re-orienting the supplier side, to offer *functions*, instead of *artefacts*, for satisfying current needs;
- establishing positive actions (not series of vetoes, that simply give rise to hidden bargains), to empower, according to *fair* trade rules, the *service* economy, respectful of the ecological side of welfare;
- stimulating the jobs for the reuse of tangibles (*reverse logistic*: to collect, disassemble, recover, sort, sell, etc., *second-hand* materials);
- supporting the infrastructures of the knowledge-driven entrepreneurship, while discouraging the processes with intensive tangibles spoilage;
- fostering common practices all over the common market, to look after world-wide covering.

Shared motivations are, once more, effectiveness assessment and operation transparency: these allow checks on the return on investment and on the targets achievement. The simulation approaches are powerful means to study useful synergies, with assessment of the actually enabling options.

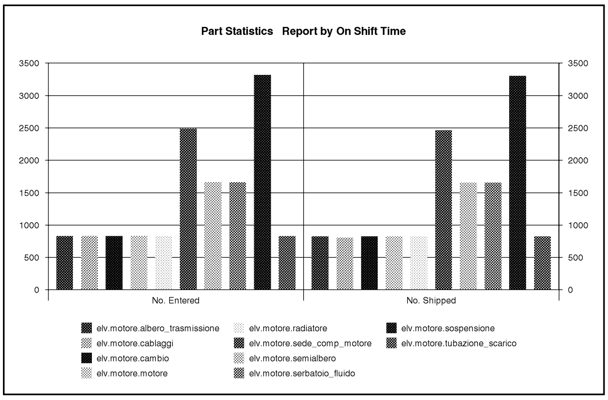


Fig 9. **WITNESS** modules example output.

CONCLUDING COMMENTS.

- Recovery and recycling, directly, address requirements along material supply chains, to lower the impact in pollution and consumption. Several analyses have been undertaken with contrasting issues. The enacted legal frames, today, are far from coherent regulations; there is, perhaps, too emphasis on details, such as waste safeguard, and attention on only limited classes of artefacts, without a comprehensive attack to the recover/reuse/recycle requirements, in view of sustainability. The critical nature of eco-consistency, thereafter, appears as consequence of mainly fragmented measures and odd precautions, which result from the EU policy, as sum of slightly persuaded Member States acts. This lack of consistency risks to undermine the credibility of series of rules, when the eco-advantages, weighed against the incumbents, do not clearly appear, and when the procedures do not give evidence that the onerousness is equally distributed among all the citizens and across EU partners. It should, also, be recalled that deployment of the affluent society paradigms is non uniform conquest and with uneven achievement times; then, a set of law provisions socially and politically correct in some States, could turn ineffective or even self-defeating in other contexts.

ACKNOWLEDGMENT.

The work is done within CIELI, *Centro Italiano di Eccellenza sulla Logistica Integrata*, established in Genova, with sponsorship and financial support of the MIUR, Ministry of Education, Research and University.

REFERENCES.

[1] G.M.Acaccia, R.C.Michelini, N.Qualich: "End-of-life vehicles collection and disassembly: modelling and simulation", *Joint ESM-MESM Conf.*, *EUROSIS 05*, Porto, Oct. 24-26, 2005, pp. 34-40, ISBN 90-77381-21-X.

[2] R.C.Michelini, R.P.Razzoli: "Product-service for environmental safeguard: a metric to sustainability" *Intl. J. Resources, Conservation and Recycling*, vol. 42, n° 1, Aug. 2004, pp. 83-98, ISSN: 0921-3449.

[3] R.C.Michelini, R.P.Razzoli: "Product-service eco-design: knowledge-based infrastructures" *Intl. J. Cleaner Production, Elsevier*, vol. 12, n° 4, May 2004, pp. 415-428, ISSN 0959-6526.

[4] R.C. Michelini, G.L. Kovàcs "Information infrastructures and sustainability", L. Camarinha Matos, Ed., Emerging Solutions for Future Manufacturing Systems, Springer, pp. 347-356, 2005.

[5] R.Dekker, M.Fleischmann, K.Inderfurth, L.N.van Wassenhove, Eds.: "Reverse logistics: quantitative models for closed loop supply chains", *Springer*, 2004, p. 436.

[6] H.Dyckhoff, R.Lackes, J.Reese, Eds.: "Supply chain management and reverse logistics", *Springer*, 2004, p. 426.

[7] D.F.Blumberg: "Introduction to management of reverse logistics and closed loop supply chain processes", *CRC Press,* Oct. 2004, p. 248.

[8] J.Wengel, P.Warnke, J.Lindbom: "Case study: the automotive industry personal cars", Fraunhofer Institute, Karlsruhe, The Future of Manufacturing in Europe 2015-2020, Challenge for Sustainability, JCR-IPTS, 2002, Sevilla.

[9] BMWGroup: "Innovation, efficiency, responsibility: sustainable value report 2003/2004", Brand: BMW, Rolls-Royce Motor Cars Limited, MINI.

[10] CEN/TC 273 "Logistics: WG4, Logistics performance measures requirements and measuring methods", Report BT N 5976, 19 November 1999.

[11] EU Directive: http://europa.eu.int/comm/ environment/ waste/elv_index.htm.

END-OF-LIFE VEHICLE MODELS WITH RECYCLING IN MIND.

G.M.ACACCIA, R.C.MICHELINI, N.QUALICH

PMAR*lab*- Laboratorio di Progettazione e Misure per l'Automazione e la Robotica
DIMEC- Dipartimento di Meccanica e Costruzione delle Macchine - Università di Genova
Via all'Opera Pia 15A - 16145 Genova - Italy
{acaccia|michelini|qualich}@dimec.unige.it

KEYWORDS.

Discrete, decision support, computer-aided analysis, dynamic modelling, event-oriented

ABSTRACT.

Growth sustainability requires to lower pollution and raw materials consumption. The paper deals with car end-of-life recovery (reuse, recycle) regulation as new business issue, and outlines modelling and simulation paths, to assess problems in reverse logistics for items collection and process data management.

INTRODUCTION.

Growth requires wealth creation with less eco-impact. For wellbeing expansion, the underlying organisations and technologies present shared features, with unifying opportunity in *service engineering*, issue in *extended* artefacts and support in *extended* enterprises. The economical instruments aims at: *voluntary agreements*, for on-duty conformance management; *compulsory regulations*, for recovery targets. The ICT aids play a main role in developing both. Moving to life-long concern, knowledge-driven frames are necessary, to make possible assessing items impact and natural resources decay. These frames establish on three facts:

- to market products-services, or *extended* artefacts, with collaborative aids for clients' support;
- to establish *extended* enterprises, or net-concerns, for point-of-service conformance guarantees;
- to monitor the tangibles yield per unit service with data recording by third-parties certifying bodies .

The frame develops with hierarchical topology:

- the inner cluster, to link the extended enterprise partners, for the product-service delivery;
- the specialised links, to support the point-of-service communication with individual buyers;
- the outer selective data channels, for the overseeing bodies, under proper security protocols.

The outlined framework is considered by the paper, especially addressing the EU environmental policy in the automotive domain, and suggesting *end-of-life vehicle* models with recycling in mind, properly adapted to the enacted sustainability targets.

The concluding comments address the links between ecological constraints and economic instruments, and stress on the fundamental role played by the ICT aids, both, as off-process and as on-process enabling tools.

SUSTAINABILTY AND RECOVERY RULES.

The wealth growth, by ceaselessly replacing tangible goods, built transforming natural resources into waste and pollution, grants benefits to the directly involved *consumers*, with penalty to third parties and future generations. The *consumers* side covers the all supply chain: manufacturer and user. The former establishes the product functional properties, choosing materials, specifying operation properties and eco-impact, and providing the construction files for maintenance and dismissal. The latter needs comply the technical and legal regulations, to be allowed to enjoy the purchased items. The industrial revolution has highly widened the manufacture market, by increased process productivity, but has, as well, speeded up the resource decay and the environment downgrading, so that the growth becomes critical. We are, apparently, approaching a bottleneck, and ***voluntary*** provisions and ***mandatory*** restrictions shall urgently apply, to alleviate or slowing down the world decay. The EU has set ambitious goals and sketched severe policies, based on the manufacturer's *responsibility*, considering three phases:

- within the production process: suitable antipollution regulations are issued at manufacturing, and design is promoted by series of advices and warnings (e.g., Eco-design of End-use Equipment, EEE-directive);
- along the supply chain: the eco-consistency figures, included in the construction file, need be followed, for conformance-to-use, and the service engineering comes out as valuable opportunity;
- at the product dismissal: mainly durables (in future, consumables) fall within suppliers responsibility, under the free-take-back scheme, aiming at reverse logistic flow for resource recovery and for dump avoidance.

Recovery is fostered as antidote, with inherent task of educating consumers, to conservative behaviours; thus, mass-produced goods are chosen, such as end-of-life vehicles, *ELV*, or waste electrical and electronic equipment, *WEEE*, burdening the replacement market, with high environmental impact. Eco-consistency is ruled through producers responsibility, by enforceable targets regulations. The *ELV* case provides explanatory hints to this trend. The 2000/53/EC Directive, enacted the 18.09.2000 (with Member States acknowledgement before April 2002), defines the rules to be followed by national acts for vehicles dismissal, and by carmakers for producing/selling cars respectful of legal requests. The Commission integrated the rules by notes, e.g., the 2003/138/EC, to explain standards for materials and parts. Basic aspects are summarised in Fig.1. The EU is concerned by *end targets*, not about *how* they are fixed. A Guidance Document collects harmonisation hints, to avoid misinterpretation.

- collecting systems for exhausted vehicles and parts need be established at authorised sites, for preliminary treatments, to grant safety and security fitting out, by removing potentially dangerous components;
- withdrawal needs to be performed without charge on the final owners (prescription to be fully enabled from 01.01.2007), but recovery visible-fees are included, as attribute of the product-service delivery;
- manufacturers (and dealers) shall bear technical responsibility for the artefact lifecycle, dismissal included, being liable of environment impact and resource consumption, according to the design choices;
- users co-responsibility is sanctioned, for voluntary non-conservative behaviours, when critical pieces are damaged, removed or modified, altering the original setting of the supply;
- dismantling and recovery incumbents ought to be acknowledged, with assessment of reused parts, recycled materials, thermo-exploitation and residuals dumping, to be notified to the European Commission.

Fig 1. Basic aspects of the EU Directives for ELV recovery.

The WEEE case, in force by mid 2005, is slightly different, to tackle with non-registered goods. Dealers assure items withdrawal; sets of recover and recycle targets apply, depending on appliances size or lighting rigs; a visible eco-fee can be collected when a new device is sold.
Once the changes of the EU environmental policy are on full effect, the manufacturers responsibility will lead to deeply different entrepreneurship organisations, due to *knowledge-driven* patterns and linked to computer-based instrumental aids. Hereafter, the PMAR***RL**elv* code, for reverse logistics simulation, is recalled and example results on the operation potentials of *ELV* case are discussed, to provide introductory explanations.

THE SIMULATION ENVIRONMENT.

The *reverse logistics* is the process to design, plan and control the recovery and reuse of worn-out products, to preserve natural resources and protect the eco-system. The considered simulation aids develop, exploiting occurrence-driven architectures, to provide real-time, description of the backward flows, by duplicating the effects of every forcing (and disturbing) inputs, as they take place in the reality. At the same time, alternatives of physical resources and/or logical plans are modelled, making un-expensive *virtual* tests, without actually building pilot plants. The basic options offered by this kind of simulation aids are well known, and we do not enter into details.
The PMAR**RL**elv code has the featuring property to be based on standard packages, [Fig. 2]:

- **WITNESS** tool, version *logistics*, for the collection-dispatching modules, and version *manufacture*, for the dismantling-shredding-sorting modules, with virtual reality aids, leading to self-sufficient and friendly modelling.
- **MS ACCESS** software, standard relational database, released as MS office suite.
- **VISUALBASIC.NET**, based on the **.NET**-platform, powerful and flexible, but reasonably friendly also for web applications, assuring management aids to create and to run interfaces with other languages, procedures and databases; to define interactions of blocks; to start/stop concurrent processes; to enable web applications; to establish masks and graphic displays.
- **CRYSTAL REPORT XL**, as standard to reporting.

This environment aims at a few scopes, such as:

- creation of digital mock-ups of actual physical resources involved in reverse logistics duties;
- provision of personalised access to web-clients, interested to investigate special opportunities;
- build-up of qualified knowledge, by systematic ranking set of architectures and schedules.

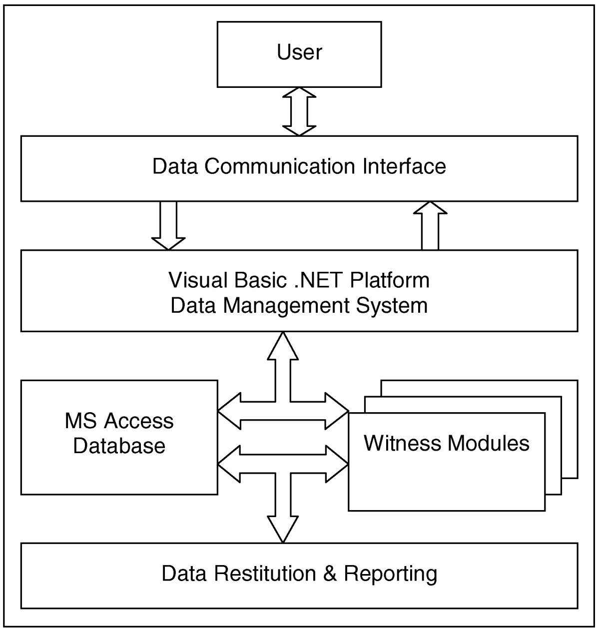


Fig 2. Block schema of the PMARRLelv code.

The involved physical resources distinguish different operation areas: transport and dispatching, dismantling and recovering, and shredding and sorting, with local buffers and materials handling rigs. An open network architecture is very important feature, due to the multiplicity of involved stakeholders: carmakers, end-users, service providers, certification bodies, etc., with quite different operation styles and interests. The knowledge build-up, thereafter, bears multiple fruits, directly, to acknowledge the behavioural properties of the piece or process under investigation, and indirectly, to re-engineer the whole supply chain.
The existing ICT aids make possible net concerns, with varying topology lay-outs and formal recall to *virtuality*, meaning by that each partner (a activity-node in the net, enabled when the role has to be played) is allowed full visibility, through twin-layer interfacing, of every pertinent details, as given by the enacted compulsory environmental rules. The resulting

virtual organisation does not correspond to a single industrial company: on the forward and backward chains there are lots of competencies requirements and duties fulfilments, with, possibly, contrasting interests that need be balanced, with transparent acknowledgement of obligations, achievements and profits. The simulation needs duplicate peculiarities and facets of the net concerns, with twofold issues:

- to help figure-out effective *virtual* organisations, when central, regional or local agencies look after operation legal frames guaranteeing the mandatory targets of the EU directives;
- to provide powerful decision-support to develop innovative *extended* enterprises, ensuring lifelong *extended* artefacts delivery, with full on-duty and withdrawal eco-servicing.

The PMAR***RL**elv* environment incorporates both features, including the relevant attribute of *virtuality*, in the above mentioned way of interfacing nodes, through the twin-layer diaphragm, separating the stakeholders, by competency and responsibility.

EXAMPLE KNOWLEDGE FEATURES.

The PMAR***RL**elv* code, as said, aims at assessing the ELV treatments, leading to a set of modules to grant:

- the take-back and gathering of end-of-life vehicles, for dismissal incumbents;
- the ELV dismantling, with law-consistent separation of reuse-parts and recycle-materials;
- the material recycling processes, after shredding and sorting, and safe landfill dumping;
- the backward-to-forward chain data-flow to enhance the conditioning economical instruments.

The development of these modules is on-progress, [1], in keeping with the EU enacted mandatory targets:

- *from 01.01.2006*: 90% by weigh of the vehicle ought to be recovered or recycled, and only 10% can be dumped (suitably neutralised) to landfill; in any case, materials recycling should be as high as 85%, since only 5% can be used as auxiliary fuel;
- *from 01.01.2015*: the figures are modified, allowing 10% for fuel use, but only 5% to landfills (aiming at the recovery figure of 95% by weigh).

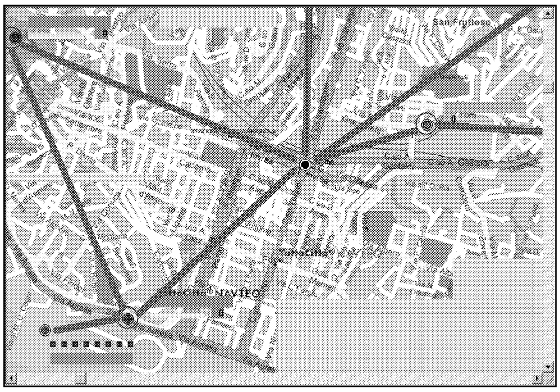


Fig 3. Schematic map for dropped-off ELV withdrawal.

The enforced figures could be object of revision. Today the ELV metallic content is consistent with the 85% recycling target, since, generally over than 15 years old car are withdrawn; the plastics content of recent cars is quite higher, and this target will become difficult. Moreover, the landfill limitation (10% today, and 5% from 2015) is quite restrictive, being mainly driven by anti-pollution summons, with no concern on economic return and energy consumption.

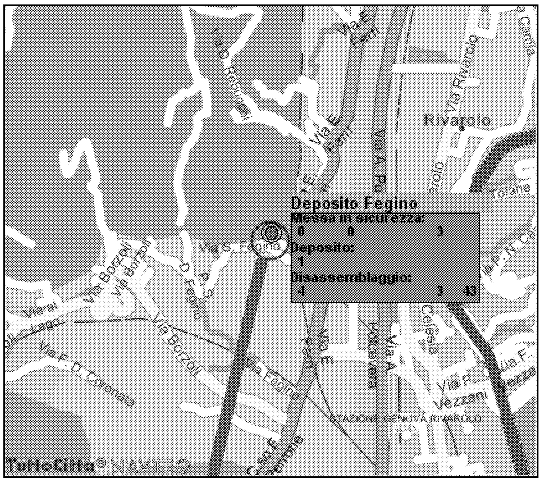


Fig 4. Map details and local operating facilities.

Knowledge flow transparency is *fairness* prerequisite. At the moment, the backward streams bear quite loose operation styles: ELV are fetched at random times and scattered locations and dropped to off-hand wrecking shops; disassembly and reuse duties are often done by extempore operators, as case arises; only, shredding and sorting tasks have resort to mechanised plants, but through human-intensive attendance. The economical end ecological efficiencies of the whole are, both, very low, with no interest of the car-makers to bring forth enhanced resource recovery, as this might pull down the sale of new items. The producers responsibility will foster more conservative behaviours, on condition to reach transparent economic instruments management, so that drawbacks and benefits are distributed with objective fairness. The PMAR***RL**elv* code is especially conceived to carry out analyses on the backward chain, acknowledging the critical steps, so that full visibility of the fulfilled processes is quantitatively achieved. The enabling role of the ICT ought to be emphasised, as it assures technology up-grading and systemic innovation for a class of policies based on:

- the resort to manufacturers' responsibility, linked with *economic instruments* drivers;
- the involvement of stakeholders, with dissimilar technological profile and capability, uneven market position and power, and different economic interest towards sustainability;
- the planning of the mandatory targets achievement, by business paradigm shifts, where the *extended* artefacts (or *products-services*) and the *extended* enterprises or *net-concerns*) play critical role.

These three policies apply when the recovery-reuse-recycle targets lead to fixed figures, and compulsory regulations are enacted, with sanctions and fines to infringers.

The knowledge build-up by simulation is powerful way to assess drawbacks and benefits, and their sharing by the involved stakeholders. Actually, the waste items (cars, domestic appliances, etc.) quickly reach *negative* value, in face of sound recovery (reuse/recycle) targets, due to dismantling and dumping costs higher than the sale prices of reused parts and recycled materials. Any pace-wise up-grading, based on voluntary agreements between private partners, is unable to economically achieve the EU mandatory targets with return on investment, thus inter-dependent value-chains shall be pursued. The induced off-springs shall address:

- the creation of *networks* of transporters, collectors, dismantlers, shredders, sorters, etc., with the related links, upwards, to manufacturers and dealers, and downwards, to re-users, recyclers and materials;
- the establishment of *networks* of certified bodies, registered by the national authorities and notified to the EU, to accomplish overseeing duties, to record achievements, to draw eco-fees, and, whether the case arises, to inflict the punishments;
- the deployment of *product-service* deliveries, with high recovery figures (no dangerous substances or parts, good dismantling properties, modular reused sub-items, etc.), supported by *extended* enterprises, with lifecycle responsibility.

Separately, the induced off-springs do not have the potential to reach economical profit and eco-sound targets. Important up-grading might exploit special adjustments (e.g., *cascade recycling*, say, the resort to secondary materials to increase the tangibles yield per unit service, etc.), but only the integration of the options leads to method *innovation*, say, the industrial setting based on knowledge-driven entrepreneurship.
As example application, the PMAR***RL****elv* environment is used as reference tool for the local authorities of Genova, to establish an effective withdrawal policy for dropped out end-of-life vehicles. The reclamation tasks are let out to a set of wreckers, required to fetch and bring the left-out vehicles to given store-shops in the town near outskirts. The problem [Fig. 3-4] deals with:

- the map covered by the withdrawal service;
- the travelled road lay-out, with traffic-periods weigh;
- the duty-service of the wreckers;
- the capacity of the (two) store-shops.

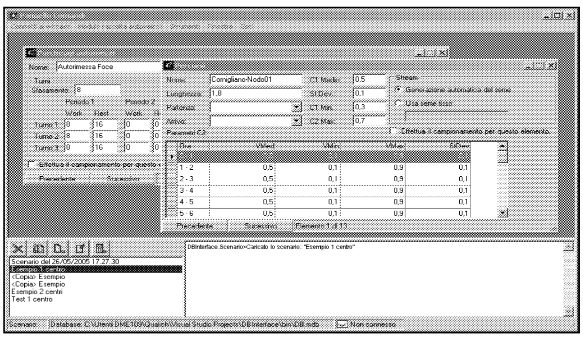


Fig 5. PMARRLelv environment structured input.

The analysis should consider several competing lay-outs, changing the wreckers allocation, the duty-cycles, the job priorities, etc.; then typical performance charts are generated, providing synthetic indices [Fig. 5-6] to show the wreckers utilisation ratio, the average figures of the considered tasks, with specification of the total amount of removed vehicles and nominal properties of each service. The case example is direct outcome of the burdensomeness of de-registration incumbents, so that the owners of old crocks take advantage by dropping off their ELV, to by-pass the bureaucracy. Then, local boroughs need face reclamation duties, with the further constraints to satisfy the enacted eco-rules.
The economic instruments, underlying the environment policy, attract increasing research attention. Industrial innovation is main answer; still, how specific measures affect the changes only address "induced off-springs" and "evolutionary up-grading". Many projects consider technological and organisational issues linked to given areas (automotive, electrical/electronic equipment, etc.) once regulations are enacted, and the enterprises' strategies follow the *black-box* approach, recognising the input-output flows and experiencing sample *on-the-path* adjustments. This is, typically, the way to exploit flexibility and leanness by pace-wise betterments, leaving aside *method* innovations, which only lead to competitive divide in business paradigms. To move a step onward, the *black-box* research projects ought to operate an integrated approach, joining *technical* and *legal* drivers, into balanced promotions, with properly wide-ranging actions. These paths shall encompass:

- ***process*** innovation: resource-recovery path, with higher value in part-reuse and material-recycling; thermo-enhancement path, by energy recovery from waste; full-substitution path, by modified material choice and items design for high recovering; etc.;
- ***method*** innovation: *extended* artefact path, through commodities and utilities joint delivery; *extended* enterprise path, through efficient supporting nets; information value-added path, to replace materials by intangible equivalent-functions; etc..

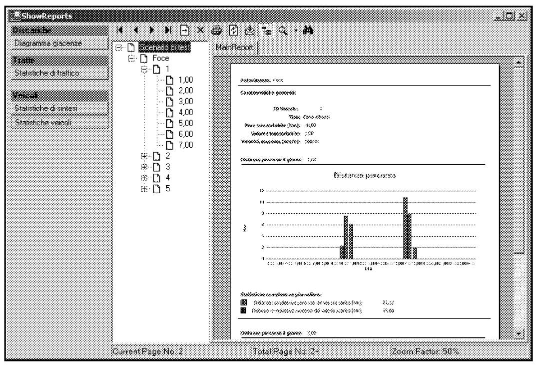


Fig 6. PMARRLelv environment example output.

The PMAR***RL****elv* environment helps providing visibility on the conditioning knowledge features, addressing noteworthy aspects. Example issues are:

- for ***process*** innovation: the logistics costs (handling and transport of very poor items) as compared with recovery (reuse, recycle) processing; this knowledge helps to localise the dismantling sites and to establish treatment chains, selectively addressing given goods (WEEE, ELV, etc.), either mixing material refuses, for scale economy;
- for ***method*** innovation: the value chain of *service engineering* supply, covering, out of the mandatory *take-back*, the whole eco-consistent operation life, to exploit synergic *value-added* options; this knowledge

will help organising the materials and information flows, setting transparency requisites, in view of balancing return on investment and (law-driven) fair-trade constraints (and costs).

The eco-innovation follows alternative/complementary paths with interrelated links, exploring *technical* and *legislation* drivers to achieve demanding targets and economical profit, with visible outcomes within the different EU partners, as shown by the noteworthy aspects. Other example issues could be mentioned, and the ability of progressing through modelling and simulation is winning opportunity to get over the simple *black-box* approach, and to aim at ***method*** innovation.

CONCLUDING COMMENTS.

Eco-conservativeness is imperative demand, if *quality of life* continuation is dealt with. This results in series of accomplishments, with, mainly, two purposes:

- to expand intangible additions in the value chain, exploiting ICT means for wealth creation;
- to lower tangibles consumption, lowering dumping and pollution, by mandatory recovery targets.

The EU environmental policy addresses both purposes, but the overall transparency of actual results does not assure clear understanding of worsening or betterment and acknowledgement of planned evolutionary up-grading or induced off-springs. The present paper discusses the role played by the ICT aids, in general, to support the conditioning information flow and, in particular (by the PMAR***RL****elv* code), to help assessing reverse logistics profitability achievements, once the EU compulsory ecological constraints are faced by the enacted economic instruments.

The discussion should conveniently link compulsory targets (producers/suppliers have the bylaw obligation to assure *free-take-back* of dismissed goods), with voluntary agreements (*extended* enterprises can expand their business by the contract engagement to support *fit-for-purpose* operation on the *extended*-artefacts life span). The incumbents are linked with series of eco-checks, yielding to the third-party certification of the environmental impact, pollutants remediation, resource consumption, etc., by process transparency and results recording, to provide the acknowledged assessment of actually used natural resources. Such facts are dramatic opportunity of the knowledge entrepreneurship, which will develop, once people become aware of the enacted regulations (for ELV, WEEE, etc.) and of the existing prospects (e.g., in *product-service* delivery). Now, immediacy or delay in enterprises birth depends on local politico-legal or socio-economical bents, as, on the scientific and technological side, the ICT, already, provides successful enabling supports, both, as off-process and as on-process instruments; the latter being typically related to the novel extensions in *service engineering* and *ambient intelligence*, for on-duty monitoring and lifecycle data vaulting.

To draw conclusions from discussing the topics in sustainable engineering management, the *compulsory targets* are recognised to establish characterising facts, driven by economic instruments:

- the producers responsibility principle, as it surfaces from the EU directives, with regards of the dismissal/ dismantling/recovery requirements;
- the law frame, given by the recover (reuse/recycle) compulsory targets, as economic incentive for the process effectiveness with return on investment;
- the example regulation enacted for ELV and WEEE, by free-take-back, grounded on recycling visible-fees and design-for-dismantling/recovering rules;
- the operation set-ups, leading to enhanced usability and reverse logistics achievements, for environment protection and eco-sustainability.

ACKNOWLEDGMENT.

The work is done within CIELI, *Centro Italiano di Eccellenza sulla Logistica Integrata*, established in Genova, with sponsorship and financial support of the MIUR, Ministry of Education, Research and University.

REFERENCES.

[1] G.M.Acaccia, R.C.Michelini, N.Qualich: “End-of-life vehicles collection and disassembly: modelling and simulation”, Joint ESM-MESM Conf., EUROSIS 05, Porto, Oct. 24-26, 2005, pp. 34-40.

[2] F.Aggeri: “Environmental policies and innovation: a knowledge based perspective on co-operative approaches”, J. Research Policy, vol. 28, 1999, pp. 699-717.

[3] D.F.Blumberg: “Introduction to management of reverse logistics and closed loop supply chain processes”, CRC Press, Oct. 2004, p. 248.

[4] S.Brunnermayer, M.Cohen: “The determinants of environmental innovation in US manufacturing industries”, J. Environmental Economics and Management, vol. 45, 2003, pp. 278-293.

[5] P.Calcott, M.Walls: “Can downstream waste disposal policies encourage upstream design-for-environment?”, American Economic Review, Paper & Proceedings, vol. 90, n° 2, 2000, pp. 233-237.

[6] R.C.Michelini, G.L.Kovacs: “Integrated design for sustainability: intelligence for eco-consistent products-and-services”, The Estonian Business School, EBS, Review, Tallin, Winter 2002-3 issue, n° 15, Dec. 2002, pp. 81-95.

[7] J.Stark: “Product lifecycle management”, Springer, London, p. XVII 441, 2005.

[8] H.K.Suhas: “From quality to virtual enterprise: an integrated approach”, CRC Press, Boca Raton, Aug. 2001.

SUPPORTING TECHNOLOGIES

TOTAL ENGINEERING THROUGH CONCURRENT ENGINEERING AND RAPID PROTOTYPE AS DESIGN: A POWERFUL COMBINATION

Olaf Diegel
Institute of Technology and Engineering
Massey University
Auckland, New Zealand
O.Diegel@massey.ac.nz

KEYWORDS

Rapid New Product Development, prototype as design, rapid prototyping, concurrent engineering

ABSTRACT

Concurrent engineering is a business strategy which replaces the traditional product development process with one in which tasks are done in parallel. Traditionally this has been a methodology applied more to the later stages of a project rather than the early 'fuzzy-front-end' stages. This paper introduces 'Rapid Prototype as Design' as an effective methodology, in conjunction with concurrent engineering, to help produce effective results from the very early conceptual stages of the project. This paper presents a case study on the GlucoFridge, a pocket-sized, battery powered, insulin refrigerator which was a technologically complex project involving a variety of technologies. The product was developed in three months through the tight integrated use of concurrent engineering and virtual and physical rapid prototyping technologies that allowed for an extremely fast reiterative design approach and a short development time.

INTRODUCTION

Most New Product Development projects (NPD) require the production of physical artefacts that demonstrate the working principals of the product. These projects often involve the development of a one-of-a-kind piece of equipment or machinery and often involve multiple areas of technology, such as electronics, mechanical, software, manufacturing and so on. They are, of course, also characterized by having to produce results within relatively tight time-frames, as dictated by company or market requirements.
Concurrent engineering is generally well suited to rapid product development as its parallel processes naturally tighten the product development cycle (Huang, 1999).
The traditional Prototype as Design technique, as used by the NASA's Ames Research Center, proves very useful in creating unique, one-of-a-kind research hardware for small, high-risk projects (Mulenburg, 2004). It is therefore a useful technique in NPD projects as it often helps them to produce better results faster.
With the relatively recent advent of newer and faster rapid prototyping technologies, both virtual and physical, a higher rate of design iteration can now achieved, which often results in a better project outcome. The incorporation of these technologies into the design process can be seen as a Rapid-prototype as Design process (Diegel et al, 2005).
The combination of concurrent engineering methods and 'rapid-prototyping as design' makes for a form of 'total engineering' that can allow better and faster development of new products.

PROTOTYPE AS DESIGN

When managing the design of unique and complex products, one must ask the question as to whether traditional project management is up to the task. According to Frame (Frame, 2002), "…traditional project management is broken." as it has failed to adapt to meet changing times and technologies. Whitney (Whitney, 1989) notes that: "In many large companies, design has become a bureaucratic tangle, a process confounded by fragmentation, overspecialization, power struggles, and delays."
Traditional project management tends to focus on the following areas: Cost, Time, and Quality (which is usually defined by the technical requirements of a project). With global commerce supported by technology and communication made possible by the internet, time is now 24/7 and cost and technical challenges are addressed on a global basis using team members in different countries, with different cultures and time zones, languages, and methods. In this changing world, New Product Development time is rapidly becoming the most critical factor to project success. High-tech products that come to market six months late but on budget will earn 33% less profit over 5 years. In contrast, coming out on time and 50% over budget cuts profit by only 4% (McKinsey and company, 1989). If companies develop products on budget, but in shorter times, they develop a commercial advantage and increased flexibility.
Design by Prototype is a technique in use by such organizations as the National Aeronautics and Space Administration's Ames Research Center (Mulenburg, 2004). It shows significant success in simplifying and speeding up the development of unique research hardware with large cost savings. Design by Prototype is a means of using the old artisan's technique of prototyping as a modern design tool. Prototyping is probably the oldest product development technique in the world and has been used by artisans for centuries. These artisans created prototypes of their ideas, to ensure that they worked, before making the primary artefact they were planning. Traditional Design by Prototype is useful in creating hardware for one-of-a-kind projects by eliminating much of the formal engineering design process.
It is often impossible to precisely specify requirements at the fuzzy front end of a project. Even when possible, it may be undesirable to do so (Frame, 2002). This often makes Design by Prototype critical to projects. Design by Prototype is a highly interactive, integrated process that allows multiple iterations of complex aspects of a desired R&D product to be quickly evaluated and adapted into a properly functioning whole (Mulenburg, 2004). This 'whole' almost always meets the users' needs, as they actively participate in the design as it evolves during development. It gets their buy-in with each further improved iteration of the prototype.
The need for using this new/old process in new product development companies is largely due to the proliferation of

highly functional and easy to use computer aided design (CAD) design tools to highly skilled and versatile engineers. One of the problems with CAD is that it does not always reflect reality accurately. In a review of 72 development projects in the computer industry (Eisenhardt and Tabrizi, 1995) it was found that the common perception that Computer Aided Design greatly enhanced product development time, was often not the reality. Further anecdotal experience also shows that the extensive use of computer design tools can result in both excessive time expended in design, and a lack of imbedded reality in the final product. A design may look pretty on the computer screen, but will it meet the users' needs and can it be efficiently made as designed? Often many design changes occur during the manufacture of these pretty designs that increase both schedule and cost to the project without a commensurate increase in product usability or quality. Beautiful three-dimensional computer models and detailed CAD drawings can result in difficult to manufacture hardware that requires expensive fabrication processes that add cost and/or increase schedule.

Prior to computers, designers who often were not engineers, converted engineering sketches into finished drawings for manufacture. While doing so, much design detail was added to not only meet manufacturing's needs, but also to ensure the end user's satisfaction. Computers have gradually eliminated the designer's role, leaving a gap that engineers are often not trained to fill: making the design manufacturable and optimizing its desired usefulness. For many high technology products, much design time can be saved and expensive rework eliminated during fabrication by using design as prototype.

Barkan & Insanti (Barkan and Insanti, 1992) advocate prototyping as a core development process for a way out of this dilemma. Mulenburg (2204) sees this is a major contributing factor in the 70~80% of projects that never make it through complete development, or fail in the marketplace because of compromises made during development that reduce content to save cost and schedule.

Mulenburg sees one of the major contributors to problems during the traditional linear design process as being an attempt to make every part as effective as possible (Mulenburg, 2004). Trained in design, many engineers try to optimize every portion of a product in trying to create an optimized whole, which is exactly the opposite of what is required for both speed and parsimony in design. The result is sub-optimization adding both time and cost to the design process without optimizing the final product. An old Zen proverb captures this problem in a few effective words; Perfection comes not when there is nothing more to add, but when there is nothing more to subtract.

The desired product must, of course, meet the basic needs of the intended user, and these needs must be agreed upon and defined as early and as clearly into the project as possible. Reality is that things are often optimized simply because they can be; not because they need to be. As an example, when only a few units of a product will be built, is anything achieved by a lengthy comparison of which fasteners to use in order to optimize the highest quality with the lowest cost when only a minimum order quantity will be purchased anyway? If the functional requirements can be adequately met by an early choice, it is much more important to make the selection and move on to more complex aspects of the design that may need extra time to ensure they meet the desired needs. In new product development, time truly is money.

RAPID PROTOTYPE AS DESIGN

The recent advent of the latest rapid prototyping, computer aided design (CAD), computer aided engineering (CAE) and computer aided manufacturing (CAM) technologies has added a new twist to the traditional prototype as design process. It is now transforming from a 'prototype as design' process into a 'Rapid Prototype as Design' process.

This new generation of tools now allows engineers to, for example, relatively easily perform complex finite element analysis (FEA) calculations on their products, to test for any thermal or structural problems, and even to simulate how plastic may flow through an injection molding tool during manufacturing.

Physical prototypes play a great role in product development as they are a means of demonstrating scale and realism in a way that paper drawings, and even computer 3D models, cannot. The translation from two dimensional to three dimensional representations is a key stage in product development (Vervis, 1994). The progression of prototypes can be seen as going from two dimensional to three dimensional on-screen, to three dimensional physical models. Only a three dimensional physical model can effectively achieve the real suitability of a physical product (Broek et al, 2000). There are invariably large differences in perception between a user seeing a traditional CAD model only and then seeing a real physical working model. The additional tactile, haptic and true three dimensional perception produce two completely different responses in the user (Emori, 1977).

The overall design process now looks somewhat as follows: Initial conceptual sketches are still often done in 2D, both on paper and on the computer. More advanced conceptual design and engineering design models are then produced using 3D CAD software. This produces a virtual model that can be rotated, zoomed in on, measured and manipulated on-screen. From this 3D computer model, a physical rapid prototype can be produced. Traditionally, the only way to produce a real, physical model was to either use a subtractive technology (such as CNC (Computer Numerically Controlled) machining or to produce expensive tooling into which the part could be injection molded. Both these methods were both time consuming and expensive.

The latest generation of rapid prototyping technologies, such as stereolithography (SLA), Selective Laser Sintering (SLS) and Fused Deposition Modelling (FDM) now allow physical prototypes to be produced within hours rather than days (Krar and Gill, 2003).

The rapid prototyping process begins by taking a 3D computer generated file and slicing it up into thin slices (generally ranging from 0.1mm to 0.25mm per slice depending on the technology and machine used). The rapid prototyping machine then builds the model one slice at a time, with each subsequent slice being built directly on the previous one. The technologies differ mainly in terms of the materials they use to build the part, and the process used for creating each slice of the model (Chua and Leong, 2003).

SLA uses a photosensitive epoxy resin cured by a UV laser. The laser traces each slice of the model in a scanning pattern, which cures the resin down to the desired slice thickness as it passes over it. As each slice is finished, the build platform is dropped one step into the vat of resin and another layer of resin is applied on top of the previous one so it can be traced for the next slice. This continues until the model is complete after which it is removed, cleaned up, and fully cured in a UV oven as a secondary operation.

SLS uses a similar process but uses a nylon based powder as a build material (though many other materials are now also available) and melts (sinters) it with a laser. In this case, a thin layer of powder is spread onto the build platform, and the laser traces the slice, bonding the plastic powder together. The slice is then lowered one step and another layer of powder is spread for the next slice. Each slice bonds to the previous slice.
FDM works by extruding a thin ribbon of plastic as the nozzle of the machine traces each slice. It is, in effect, not dissimilar to an inkjet printer but prints in plastic instead of ink.
There are many other technologies that use slightly different materials and process to the above. Most, however, work on the principle of building the model one slice at a time.
The rapid prototyping system used in the case study described in this paper was a Dimension BST 3D Printer, manufactured by Stratasys. The Dimension printer costs US$24,000 which is a relatively low cost in comparison to the more expensive SLA and SLS systems available. This system uses a FDM based system that extrudes a thin ribbon of ABS plastic to trace each slice. It allows for a slice thickness of 0.25mm per layer.

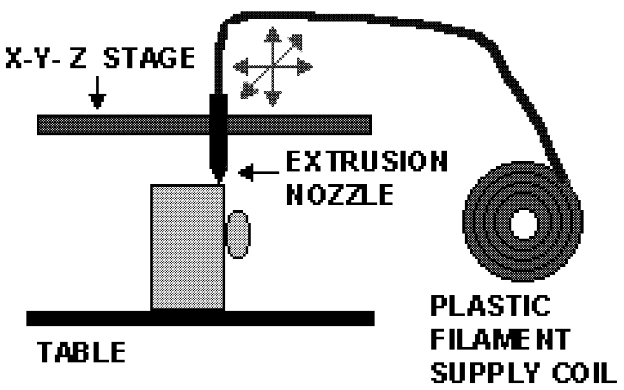


Figure 1: The FDM Process Used by the Dimension Printer

The parts produced are strong plastic components that are well suited to functional testing and can easily be sanded and painted to reproduce the aesthetics of the production product thus also making them useful for consumer testing.
The software included with the Dimension printer takes care of the model slicing automatically. All the user has to do is decide on the orientation in which the model is to be built, either to maximize strength, or to give it the best surface finish.
Some of the above rapid prototyping processes, which were previously only able to make plastic-like parts, are now becoming able to produce metal parts (Wohlers, 1999). Not only is the choice of materials and processes increasing, but the last few years have seen a significant reduction in the cost of these technologies. Systems are now also available for not only simulating the behavior and performance of electronic circuits, but also for rapid prototyping circuit boards.
These technologies mean that it is now possible to construct highly advanced virtual prototypes, and then working physical prototypes almost as fast as they are designed, thus allowing more iterations of a design within a shorter timeframe. This, in turn, potentially allows for products that are even better suited to their intended users in even shorter times (Krar and Gill, 2003).

CASE STUDY – THE GLUCOFRIDGE

The following case study, undertaken by product development staff and students at Massey University, in Auckland, New Zealand, demonstrates how using these virtual and physical rapid prototyping technologies can help in bringing a new product to market faster.

The Problem Statement

The latest WHO estimate for the number of people with diabetes, worldwide, in 2000 is 177 million. This will increase to at least 300 million by 2025. It has long been known that the number of deaths related to diabetes is considerably underestimated. A plausible figure is likely to be around 4 million deaths per year related to the presence of the disorder. Overall, direct health care costs of diabetes range from 2.5% to 15% annual health care budgets, depending on local diabetes prevalence and the sophistication of the treatment available (WHO, 1985).
When travelling, diabetics currently have to keep their insulin cool by putting it in a hotel refrigerator (if there is one available) and by carrying cooling devices such as icepacks when they travel away from their hotel. Insulin should be kept at temperatures below 25ºC. Many countries have temperatures ranging from 20ºC to 40ºC, which rapidly spoils insulin. The American Diabetes Association recommends that insulin be stored in a refrigerator (American Diabetes Association, 2005). Many hotels also do not have room refrigerators. These factors make it troublesome for diabetics to travel as they have a constant worry about whether there insulin is safe.
There are currently 3 commonly available cooling methods for insulin dependant patients as well as a number of derivative solutions. They are as following:
1) Ice packs: This still seems to be the most commonly used means of refrigeration. Ice packs are frozen and put in a cooler box in which the insulin is transported for short trips.
Cons: This is very bad for the insulin as it often comes in direct contact with the ice. Chilled insulin vials should be inspected for crystals and particulates before use. Freezing may cause crystallization, resulting in variable potency.
2) MEDIce Ice Packs: These are refreezable ice packs filled with a nontoxic gel that last 30% longer than ice cubes. Travel Organizers with MEDIce have a compartment in which the MEDIce Ice Packs are inserted so that there is no direct contact with the insulin.
Cons: Must be frozen for 46 hours before use. Does not last more than a few hours. Gel packs often leak with change of atmospheric pressure.
3) FRIO Wallets: This device came on the market in 1999 and it is designed to keep insulin cool and safe for 48 hours. The main advantages are that there are no bulky ice packs, you do not have to worry about finding a freezer to get supplies of ice and the wallet is light to carry.
It is activated by immersing it in cold water for 15 minutes. The panels of the wallet contain crystals that expand into gel with immersion in water. The wallet remains at a cool temperature for 48 hours. The system relies on the evaporation process for cooling.
Cons: The inner bag stays very damp despite drying it with a towel as recommended and so the labels on the insulin vials start to disintegrate. Even though the unit is damp, the instructions say that the cooler pouch should not be put in a plastic bag because it does not work so well. This makes it inconvenient to keep with other items that may be damaged by the dampness.

The GlucoFridge Portable Insulin refrigerator

The GlucoFridge is a portable, battery powered, pocket sized refrigerator. This "worlds smallest" refrigerator is designed for carrying insulin (or other medication), which needs to be kept at a constant and cool temperature. The applications for this technology are numerous, from transporting insulin, blood and sperm samples to anti-venom vaccines. It consists of a refrigeration unit that has drawers made for different size injection devices or medical samples. The unit comes equipped as standard with drawers for the most popular insulin injection devices.

The refrigerator is powered by rechargeable Lithium ion batteries and is of a size that allows it to fit in a jacket pocket. As its cooling technology it uses a Peltier device (heat pump). The novelty of this product lies in its portability, its uniquely small and compact size, and in its application of Peltier device technology. There is currently no other such product available.

The GlucoFridge is designed to be plugged into mains power at night, thus cooling the insulin and simultaneously charging the battery. Away from the hotel, it is powered by the batteries. As batteries it uses two lithium ion rechargeable batteries giving it a life 12 to 24 hours depending on how many times it is opened during the day.

Although many people enjoy travelling, patients with diabetes often fear or avoid it. This product gives diabetics both freedom of mind and freedom of movement when travelling.

Figure 2: GlucoFridge Final Pre-production Prototype

The GlucoFridge Design Process

The GlucoFridge went through three major, and several minor, design reiterations, the first two of which identified the major technical problems to overcome, before arriving at the final production design.

A form of concurrent engineering was used in which both the virtual and the physical prototyping for each functional area of the GlucoFridge, and the testing of each prototyped sub-system, was carried out in parallel. Almost all the sub-systems of the project were designed and virtually prototypes and several of them went through a few iterations of physical prototyping to ensure they worked the way they were supposed to. An implication of this highly parallel process was that a high level of control and good communication flow was required between the team members working on the various parts of the project. This was achieved through almost daily meetings, though not on a formal basis, and the use of a central database that graphically listed all the problems faced as well as possible solutions for the various problems.

The design process began with the inventor identifying the need for the product, and researching the market to the extent that a medical device manufacturing company was willing to take on the project, both from a manufacturing and distribution point of view, upon receiving a working prototype. From here, the following design process was used:

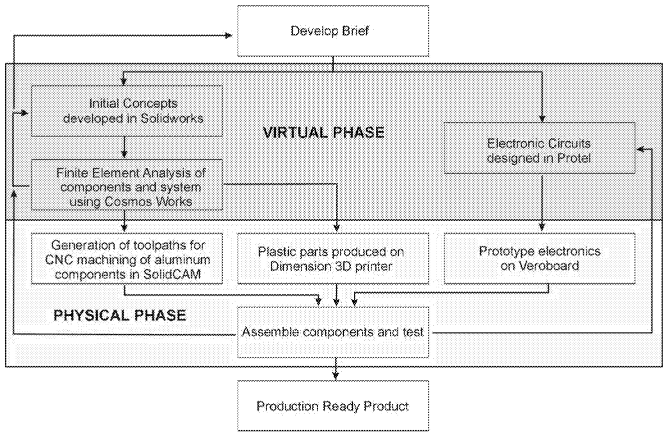


Figure 3: GlucoFridge Design Process

The first two weeks of the project were spent producing a conceptual CAD design of how the product might work from an engineering perspective. The software used for this was Solidworks, which has a relatively short learning curve and quickly allows the production of complex engineering models to simulate both the aesthetics and engineering functions of their projects. Initial concepts were designed within a time span of 4 to 5 days, which were then perfected and detailed over the following week as problems with the designs were identified.

At the same time, a large range of Peltier devices (thermoelectric coolers) was ordered so that a series of tests could be undertaken to test the ideas that were to be at the core of the product. Peltier devices are semiconductor devices that, when current is passed though them, get hot on one side and cold on the other. These devices are commonly used on small camping refrigerators, for example.

Though it may not always be considered a true use of technology, the Internet should not be ignored as a tool to speed up the product development process. It is a rich source of technical information and an extremely effective way of rapidly sourcing components. Most electronic components, for example, can often be sourced within a few days, and the extra cost this may involve is usually insignificant when compared to the benefits of the time saved. During the first 2 weeks of the project, the Internet was used extensively, both to order a wide range of Peltier devices from several manufacturers, and to obtain much information about thermoelectric cooler theory.

From the initial CAD design, a thermal finite element analysis (FEA) was performed to calculate the size of the cold plate that would be required to keep two vials and the insulin compartment of two NovoPen III (a commonly used insulin injection device) at a temperature of approximately 10ºC. The initial calculations were first done on paper to select the most likely Peltier candidates, and these selections were then run through a computer simulation. The FEA was done using the CosmosWorks add-in to Solidworks, which allows both stress and temperature analysis of systems. CosmosWorks is setup such that designers can very easily use the stress and deformation analysis sections of the software. The thermal analysis section of the software did however require some more extensive knowledge in

thermodynamics in order to better understand all the variables the software was asking for. In addition, FlowWorks was also used to simulate how the air was likely to flow through the system and allowed the creation of extra turbulence around the heat-sink area to help cool the system.
The initial design was based on using a Sony Handycam style 7.6V 3800mAh rechargeable battery running a 7.8V, 5.5W Peltier device sandwiched between a large hot plate heat-sink and a cylindrical cold plate extrusion.
From within the CAD software, tool-paths for all the aluminium components were generated within a very short time, and the aluminium components were then rapidly produced on a CNC milling center. This effectively demonstrated the tight integration now available between CAD and CAM systems. The CAM system used for this particular project was SolidCAM which is a system that is tightly integrated into Solidworks.
A rapid prototype of all the plastic components was printed on a Dimension 3D printer which produced ABS plastic parts of the design which was assembled and tested. The time taken for the first set of parts was 54 hours.
A simple version of the temperature control circuit was prototyped on Vero-board to allow for testing, while at the same time the more complex real temperature control circuit was still being designed.

Figure 4: First Iteration of GlucoFridge Design

The initial tests revealed an almost immediate problem in reaching the target temperature. Though it had initially been hoped that it would be possible to reach the target temperature without the use of a fan, it was almost immediately found that, even with a large and efficient heat-sink for the hot side of the Peltier device, the system was not quite efficient enough to dissipate enough heat for the temperature differential created by the Peltier to allow the desired refrigerator temperature to be reached.
The design was immediately modified to allow for the use of a 40mm x 40mm x 6mm fan, and a new plastic top cover component was printed and the unit was once again tested.
In this second batch of tests on this minor revision of the first design, the cold plate easily reached the desired temperature, but a second major problem soon became apparent: The design included an electronic temperature control circuit which, when the cold plate reached the lower end of the target temperature range of 5ºC, would cause the power to the Peltier to be switched off. When the temperature, after gradually increasing, then reached the upper end of the temperature range of 15ºC, the Peltier would switch back on. The intention of this feature was to extend the battery life by making do as little work as possible.
What became almost immediately apparent was that, once the power to the Peltier was switched off, it only took minutes for the cold plate to climb back up to the upper end of the temperature range. Even with the polystyrene insulation that was used to insulate the system from the outside ambient temperature, the temperature still climbed too rapidly, which meant that the battery had to switch on more often in order to keep the temperature in the desired range. This meant that the refrigerator only had a battery life of approximately six hours, which was much less than the minimum requirement of 12 hours of battery life.
A solution was soon found to this problem: To build the entire cooling assembly into the inside of a vacuum flask type cooling container. The vacuum would act as an effective insulation that would ensure that, once the cold plate reached the low end of the temperature range, the Peltier would switch off, and the cold plate would then, because of the effective insulation, take a long time to warm back to the upper end of the temperature range.

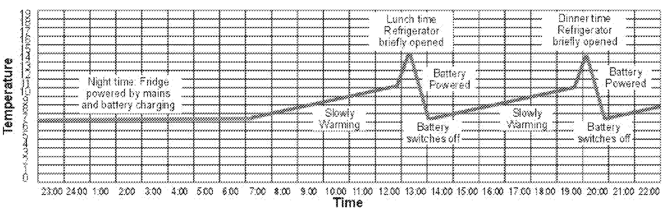


Figure 5: Desired Effect of Temperature Control Circuit

This major change forced a complete redesign, as the vacuum flask meant that the battery could no longer be on the rear of the unit. This, in turn, meant that the size and shape of the Handycam type battery made it less than ideal for the product, so a new battery needed to be found. Some Internet research and fast tests were carried out with a variety of batteries, and it was found that four 2500mAh, 1.2V AA size batteries, in combination with a 3.75V 5.5W Peltier device could achieve the desired temperature levels. This appeared to be an ideal solution as the AA batteries were substantially smaller and would allow for a smaller product.
New CAD models were generated to see if the AA batteries could be configured in such a way as to fit within the vacuum flask together with the insulin vials and pens. This proved possible by completely changing the configuration of the hot plate, Peltier, and cold plate, and by changing to a pair of 25mm fans. Another prototype was made of the plastic and aluminium components. The lead-time to get a sample of the vacuum flask made was about 3 weeks, so test were initially carried out with an off-the-shelf vacuum flask of the same volume as that in the design
This second prototype proved to work, and was taken to potential users for comment. From these focus groups a few minor and a potentially major problem were identified.
One of the minor problems was one of battery status and temperature status indication, which was easily solved through the use of multiple coloured LED indicators. Other minor problems identified included the need for the addition of a physical on/off switch so that the unit could be physically switched off when not in use, and the effectiveness of the locking clip that locked the medication

tray into the vacuum container and provided a tight seal between the two.

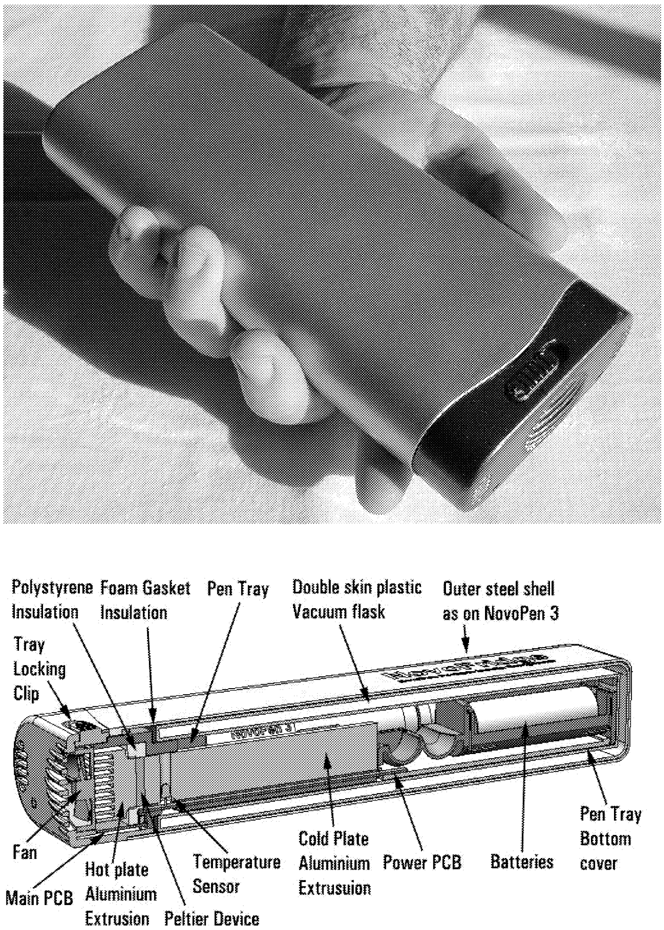


Figure 6: Second Design Iteration of GlucoFridge

The potentially major problem was to do with the batteries. Almost all the users immediately asked about the possibility of putting in regular, non-chargeable AA batteries. Though this was, in theory possible, the batteries had to be good quality, high mAh rated batteries capable of providing a high continuous current output (such as the Energizer Titanium e2 batteries, for example). This was perceived as a major problem by the users because of the perceived high chance of them using the product with unsuitable AA batteries. If the wrong type of batteries were used, they would not only tend to get very warm, thus cancelling the effect of the cooling, but also ran the risk of being damaged and leaking inside the product. It was therefore indicated by the users that it would be better if the batteries were not changeable by the user, and they even indicated a preference not to use standard AA batteries in order to eliminate the risk of the wrong type being used. The potential for condensation forming around the battery compartment was also identified as a worry by the users. It was therefore decided that it would be preferable to have the batteries outside of the vacuum flask.

From this user feedback, the third major design iteration was entered into, and a new type of batteries was selected. The new batteries were to be two 3.7V, 2200 mAh lithium ion batteries in parallel. These batteries, at 18.3mm diameter and 65.2mm in length, would not be easy to fit inside the vacuum flask unless it was made too big to fit in a pocket. The new design iteration therefore changed to a removable battery pack on the outside of the vacuum flask, which at the same time alleviated any risk of any heat produced by the batteries affecting the temperature of the insulin or creating condensation. The larger batteries also meant a small increase in the height of the product to allow for the batteries, but this in turn allowed for the use of a larger, and more effective, 35mm fan which in turn made the system more efficient.

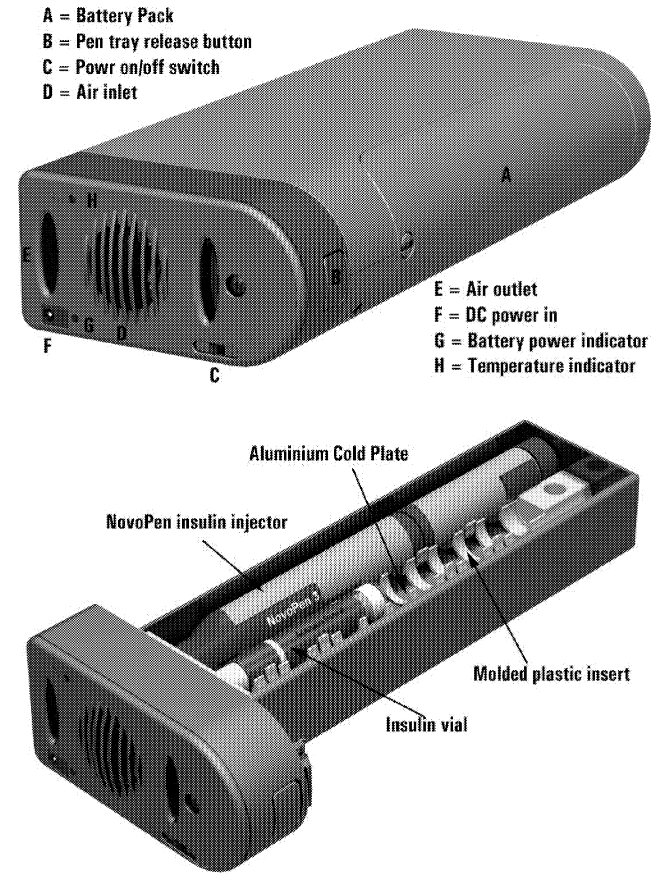


Figure 7: Final Design Iteration of GlucoFridge

From this final rapid prototype, a few minor changes were made, such as moving the power adaptor connector to the back of the unit in order to make assembly easier, and the addition of extra air-vents to the front of the medication tray in order to make the warm air exhaust off the hot plate more effective.

At this stage, less than three months after the start of the project, it was deemed that the product was at a level where it would both meet customer expectations and be manufacturable at a cost acceptable to the manufacturing company.

The entire project, including all CAD files and data, as well as the physical working prototypes was then handed over to the medical product company for manufacturing and production.

Discussion on the Design Process

The hands-on experience with Rapid Prototype as Design as presented in the GlucoFridge project was of great benefit to the design team in applying a design management technique on a real project. It allowed them to not only experience the possibility of going through several iterations of a design, but also to see the incremental changes that each iteration brought to the performance of the end product.

Another side-benefit of working with a Rapid Prototype as Design process was that it brought to the design team a true realization that, although most management and design methodologies and process are taught in a relatively linear manner, they must, in reality, often be applied in a much more parallel fashion. The process of prototyping any ideas as they come along means that you often have to look at parts of the manufacturing, and even marketing processes, even when you are still at the fuzzy front-end concept development stage. This clearly demonstrated how closely Rapid Prototype as design was related to concurrent engineering processes.

Item #	Component	Process used	Design time	Make Time	Cost	Traditional process	Design time	Make time	Cost
5,6	Body Plastic moldings	FDM rapid prototype	< 2 weeks	65 hours	$750	Machine from plastic block	1 week	1 week	$1000
38, 39	Battery pack components	FDM rapid prototype				Machine from plastic block	1 week	1 week	$1000
22	Pen tray body	FDM rapid prototype				Machine from plastic block	1 Week	1 week	$1000
24	Pen tray cover	FDM rapid prototype				Machine from plastic block	1 Week	1 week	$1000
25	Pen tray inner plate	FDM rapid prototype				Machine from plastic block	1 Week	1 week	$1000
23, 29	Buttons	FDM rapid prototype				Machine from plastic block	1 week	1 week	$500
3, 4	Vacuum flask	Welded stainless		3 week	$500	Welded stainless	2 days	3 week	$500
26	Cold plate	Machined aluminium		2 hrs	$100	Machined aluminium	1 week	8 hrs	$400
16	Hot plate	Machined aluminium		2 hrs	$100	Machined aluminium	1 week	8 hrs	$400
7, 8, 37	Springs and contacts	Handmade/catalogue		5 hrs	$200	Handmade or catalogue	1 week	5 hrs	$200
	Total		< 4 weeks		$1650		12~14 weeks		$7000

Table 1: Final Design Iteration: Comparison of Time and Costs Between Rapid and Traditional Methods

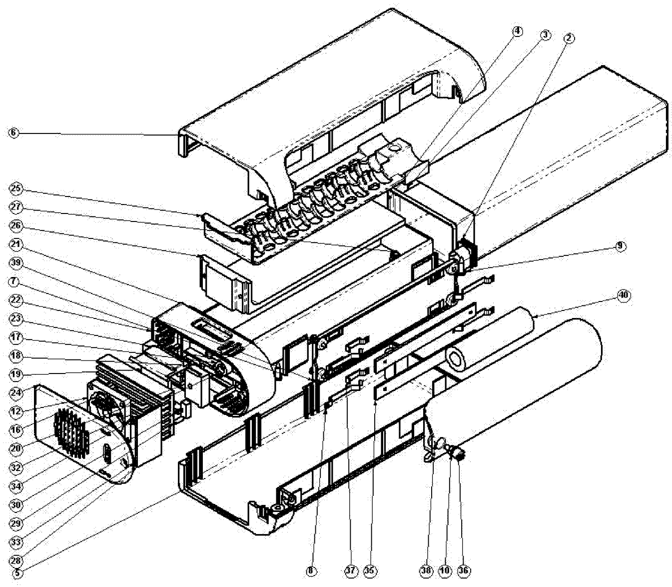


Figure 8: Final GlucoFridge Components

The components used in the final design iteration of the GlucoFridge are summarized in table 1 in an attempt to show the time differences gained by using rapid prototyping, both virtual and physical in comparison to more traditional processes such as 2D drawing and manual machining.

Some things to note about the above table are that all rapid prototypes were made in a single build that lasted 65 hours and used $750 of material. The machining costs represent the time and material costs as estimated by experienced University workshop staff, and do not therefore represent commercial rates. The Design times include all testing of components, including stress and thermal testing. The design times estimated in the traditional processes column are based on 2D drafting of the designs and physical testing of the components. These could only be roughly estimated by workshop staff based on their past experience with components of similar complexity and sizes.

The total times can also not be arrived at by adding all the individual times in series as many of the tasks are, in fact, done in parallel. The long lead time for the stainless steel vacuum flask, for example, was alleviated by sending it to the sub-contractor as soon as it was finalized in the design, and it was therefore given obvious priority in the design process. The total time for the traditional processes methods was therefore arrived at through estimates from experienced workshop staff.

CONCLUSIONS

The use of concurrent engineering methodologies and the ability to effectively integrate the many existing and emerging virtual and physical rapid prototyping technologies into the new product development process has the potential of giving engineers the ability to produce new high technology products at an increasing rate.

It is important to remember that, though any single technology or methodology adds the potential for increased product development speed, it is the paralleled use of all the methodologies and technologies available that truly gives the design team an exponential advantage over those working in a serial manner or those who do not exploit available technologies.

Design teams are often expected, as part of their projects, to produce physical artefacts that demonstrate the working principles of the products they are designing These projects often involve the development of a one-of-a-kind piece of equipment or machinery and often involve the use of multiple areas of technology, such as electronics, mechanical engineering, software, manufacturing and even marketing. They are, of course, also characterized by having to produce results within a relatively tight time-frame, as dictated by the company strategy or market requirements.

Rapid Prototype as Design can be seen as a natural evolution of the more traditional Prototype as Design process facilitated by the advent of improved CAD, CAE and CAM technologies as well as the proliferation of low cost rapid prototyping technologies. Rapid prototyping as design, particularly when many sub-systems are designed in parallel, has the potential for a vast improvement in product development speed.

The GlucoFridge portable medication refrigerator case study described in this paper was designed in a concurrent engineering fashion and employed a rapid prototype as design process for all its sub systems. It successfully went through three major design iterations, each to working prototype level within a time span of less than three months.

REFERENCES

American Diabetes Association, Resource Guide 2005, USA, http://www.diabetes.org/rg2005/insulin.jsp, (2005)

Barkan, P. and Insanti, M., Prototyping as a core development process. Design for ManufacturingCourse Handout, Stanford University (1992)

Broek, J.J., Sleijffers, W., Horvath, I., Lennings, A.F., Using physical models in design, CAID & CACD 2000, Hong Kong, 2000

Chua, C.K., Leong, K.F., Rapid Prototyping: Principles and Applications. (2nd ed). World Scientific Publishing Co, Singapore, 2003

Diegel, O., Xu, W.L., Potgieter, J., The GlucoFridge: A case study in Rapid Prototype as Design, VRAP 2005, 2nd International Conference on Advanced Research in Virtual and Rapid Prototyping, Portugal, pp 601~606

Eisenhardt, K.M., Tabrizi, B.N. Accelerating adaptive processes: product innovation in the global computer industry, Administrative Science Quarterly, Volume 40, 84-110, (1995)

Emori, R.I., Scale models in engineering, Pergamon Press, Oxford, 1977

Frame, J.D. The New Project Management. Jossey Bass, New York (2002)

Huang, G.Q., Design for X - Concurrent Engineering Imperatives, Springer, 1999

Krar S. & Gill A. Exploring Advance Manufacturing Technology New York: Industrial Press Inc. 2003

McKinsey and Company, Fortune Magazine, Feb 13, (1989)

Mulenburg, G., IAMOT 2004, International Association for Management of Technology, Design by Prototype: Examples from the National Aeronautics and Space Administration, Jan 23, (2004)

Mulenburg, G., From NASA: Don't overlook the value of Prototype as Design in developing your new products, Vision Magazine, Product Development and Management Association, Vol 28, Issue 4, October (2004)

Vervis, M., The importance of the use of physical engineering models in design, IDATER94, Loughborough University, 1994

Whitney, D.E. Manufacturing by design, in, Managing Projects and Programs, Harvard Business School Press, Boston MA. (1989)

Wohlers, T. Rapid Prototyping and Tooling State of the Industry: 1999 Worldwide Progress Report. Wohlers Associates, 1999

World Health Organisation, Diabetes Mellitus: report of a WHO study group, World Health Organisation Technical Report Series 727, WHO, Geneva, (1985)

PDM-based Mechatronic DMU

Dr. Marcus Krastel
Christian Donges
:em engineering methods AG
Lise-Meitner-Str. 10
D-64293 Darmstadt, Germany
E-mail: marcus.krastel@em.ag

ABSTRACT

Product innovations are today only possible if the mechanical and electrical components are closely integrated. Due to the fact that the IT-support of the interdisciplinary cooperation between the mechanical and electrical departments is inadequately, big human efforts are necessary in order to integrate the data of the two disciplines.

This paper shows a solution to integrate the MCAD and ECAD data by introducing new Digital-Mock-Up (DMU) tools which are capable to visualize the exact electrical components geometry as well as the mechanical geometry within one 3D-representation. It is crucial for a successful implementation of such a Mechanic-Electric-DMU-system (ME-DMU) that the functions are based on a common PDM system.

INTRODUCTION

Virtuality requires Mechatronic Simulation. This statement represents the experiences which enterprises make today as a result of two mutually affecting developments. On the one hand one observes the very strong striving for a "Virtual Product Development" in order to reduce the number of physical prototypes and replace them by their "virtual twins". On the other hand the integration of electrical, mechanical, and software components into one mechatronic product increases.

Based on this background, Digital-Mock-Up investigations accompanying the product development take a leading role in the entire product development process. In the automotive and supplier industry the technique of mechanical DMU simulations is used today on a broad basis. In order to perform a DMU analysis with mechatronic products, a consistent representation of the multidisciplinary information in an integrated product model is needed. Different authors call this representation the "Virtual Product Definition", "Digital Reference" (Kellner et al. 1999), or "Enterprise Data Model" (Scheder 1997).

The system's complexity of mechatronic products increases dramatically due to the fact of integrating more and more functions and placing the components in extremely small design spaces. Innovations are only possible if the mechanical and electronical components are very closely integrated. Today the IT-support of the interdisciplinary cooperation between the mechanical and electrical departments is inadequate. Big human efforts are needed in order to harmonize the two parallel product developments which often results in time delays.

In order to minimize errors very early in the product development process, new DMU tools need to be introduced, giving the functionality to visualize the exact electrical components geometry as well as the mechanical geometry within one 3D-representation. It is crucial for a successful implementation of such a Mechanic-Electric-DMU-system (ME-DMU) that the functions are based on a common PDM system. Therefore it is necessary that a 3D-component library with the exact geometry of the electrical components is established. An implementation of the presented concept was done using

- the PDM system SMARTEAM for the management of mechatronic data and
- the 3D-CAD system CATIA V5 for the ME-DMU functionality.

GOAL

The goal of the solution described below is the support of the EDA process (Electronic Design Automation) by an integrated data management. Therefore a dedicated customizing of the PDM system needs to be realized in order to represent the mechanical data generated by 3D-CAD systems and to handle the electrical data generated by 2D-ECAD systems. The following topics need to be addressed within an ME-DMU concept:

- Support of the EDA process by an integrated data management
- Linkage between the MCAD and ECAD structure through a neutral product structure (MCAD/ECAD-BOM)
- 3D-ME-DMU visualization of the printed circuit board layout (PCB layout) within the context of the 3D-MCAD assembly
- Management of electrical documents within the mechatronic database such as netlists or 2D-ECAD data
- Conception and implementation of a 3D-ECAD component library
- Tracking of changes incl. version and release management for electrical information

It will be shown that the presented ME-DMU implementation shortens the iteration cycles between mechanical and electrical development. The product quality increases and the time-to-market will be shortened.

FUNCTIONALITY OF TODAY'S PDM SYSTEMS

Product Data Management (PDM) systems evolved from document management systems for CAD data. Today, they assist mechanical engineers in every phase of the product life cycle. Many PDM systems still have a CAD centric architecture. Their fundamental function is to handle all type of digital documents (Svensson 2003, McIntosh 1995).

PDM systems provide methods to control the document "check in", "versioning", and "release" management. Beside documents, PDM systems manage the metadata of a product, such as serial numbers, ownerships, etc.

Depending on the PDM system, the maintainable items spectrum includes physical parts as well as abstract objects. These items are organized hierarchically within the product structure. Additionally, they may have references to one or more documents. A PDM system is able to make changes and control the related structure tree by providing a graphical, browser-based user interface. It is possible to define views to the tree and to filter the product information according to specified attributes (Sielaff 2003).

Historically PDM systems were introduced - at first for the management of design data. Until today, mechatronic data such as information from the electric development process are considered inadequately. The data generated in the EDA process is stored often in file systems. This may result in problems related to version management, release processes and data redundancy. For that reason electrical models have to be managed within the PDM system with a close relationship to the mechanical data to ensure a consistent data management (Krastel 2002).

In order to solve these problems and to support a mechatronic development process with consistent product data, a new, multidisciplinary product structure needs to be defined which is able to store data of both domains (Anderl and Krastel 2001). Figure 1 shows a concept for such a mechatronic product structure. The structure will be described in the next chapters.

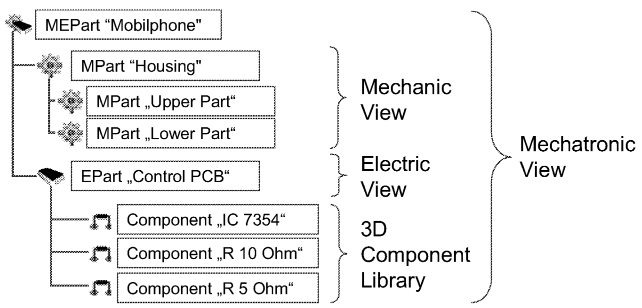


Figure 1: Mechatronic product structure

MCAD/ECAD INTEGRATION

Within the multidisciplinary mechatronic product development process many requirements concerning the interface between the domains exist. The focus in this paper is to present the requirements and solutions for a better integration of the mechanical and electrical departments.

An example of a typical today's problem: The mechanical department places big electrical components (e.g. capacitors, coils etc.) and the interface components (e.g. switches, LEDs, connectors etc.) considering design space issues. If a capacitor is placed by the mechanical department and the electrical department re-places (e.g. rotates) it during the development process because of electrical issues, the mechanical department does not notice this re-placement necessarily. This situation happens due to the fact that on product structure level no linkage between the two domains exists.

For that reason it is not possible to establish a version management which covers both domains. As a consequence an exchange of delta information (changes between an older and a newer version of the product) is not possible. The need of the engineers of both domains could be expressed by one sentence: "They want to discuss planned changes together with their colleagues and to view the results directly in the 3D-CAD model."

An efficient support of the mechatronic collaboration between the mechanical and electrical domain needs also to cope with the topic of how the 2D-information -generated within the ECAD systems- can be transferred to the 3D-space (Krebs and Franke, 2005).

ROLES WITHIN THE EDA PROCESS

The challenge of an integration solution is to support the different roles of the mechatronic product development process. Figure 5 gives -later in this paper- a role dependent overview on the use cases for the developed solution. Especially the MCAD/ECAD integration needs to support the following roles within the enterprise.

Project Manager: The project manager is responsible for the overall product development project. He defines the mechatronic product structure, the requirements, and the project tasks. Furthermore he controls the results gained by the project team. For the project manager it is important to know for which component each project team member is responsible.

Mechanical Designer: The mechanical designer designs the surrounding geometry in the 3D-CAD system. He has to fulfill the requirements defined by the project manager. He defines requirements for the PCB development such as the design space, the electrical board outline or important connection conditions.

Electrical Designer: The electrical designer is responsible for the definition of the schematics of the board. His output is a netlist with the logical electrical components and their interconnections. The input for the electrical designer is the requirement specification of the planned product function for the electrical PCB and components.

Electrical Layouter: The electrical layouter uses the information of the netlist generated by the electrical designer in order to layout the PCB in the 2D-ECAD system. He places the physical components and their interconnections. He needs to keep the design space and the connection conditions which were defined by the mechanical designer.

Library Manager: The library manager is responsible for updating and maintaining the electrical library components of the enterprise. He has to inform the electrical designer and electrical layouter if a component is obsolete or a new version exists in the library.

REQUIREMENTS AND RESULTING CONCEPTS

In the following chapter the requirements of the different roles involved in the MCAD/ECAD development process will be described.

Shortening the interdisciplinary iteration time

The main requirement of a project manager is to shorten the development iteration cycles between mechanical and electrical departments. This is only possible if the data model of the underlying PDM system is able to represent information of both domains. Furthermore it is crucial that the information is not stored separately for each domain. Only with the representation of the interconnections and dependencies, the expected benefits could be achieved (Kleiner 2003).

Figure 1 shows an example of a mechatronic product structure. The mechanical information is stored in the MPart structure. The electrical component information is stored on the same level in the EPart structure. Both structures are linked together within a new item called MEPart. This item represents the mechatronic view of the product.

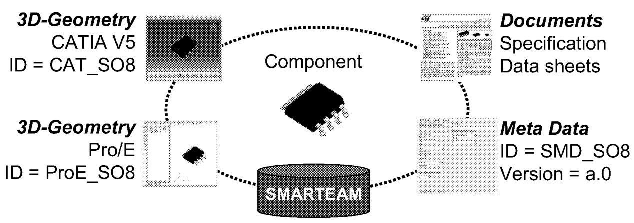


Figure 2: 3D-component library

3D-component Library

In order to support the mechatronic product development process with the ME-DMU functionality, it is necessary to build up a 3D library of the electric components in the PDM system. During the process of generating the electrical structure, a linkage of the EPart to the component library needs to be realized (see figure 1). This link should store the instance specific information such as instance name and transformation information.

With such a centralized component library a further requirement of the library manager could be addressed: the centralized handling of the electrical data sheets, accessible for all the people involved in the design and evaluation process, including a fast and efficient visualization of these documents. Figure 2 shows an example of information and documents which could be linked to one component in the library.

Mechatronic "Where used" list

The knowledge about the usage of a specific component within an assembly is very important for the mechanical designer, the electrical layouter as well as for the library manager. Today this information is often stored in different systems. The mechanical designer will find this information in his PDM system, the electrical layouter usually will find this information in his ECAD system – often the generation of a PCB spanning component used list is not possible. In order to fulfil this requirement in the electrical domain, the integrated component library is necessary.

3D-ME-DMU

A further step to support the development process is the introduction of a mechanical/electrical DMU, which shows how the electrical components fit into the mechanical surrounding. It is important that the DMU is controlled by the PDM system, because only the PDM system is able to provide the versioning information. It ensures that the DMU is build up with the latest version of product development data. Figure 3 shows an example of a real-time generated ME-DMU in the 3D-CAD system CATIA V5.

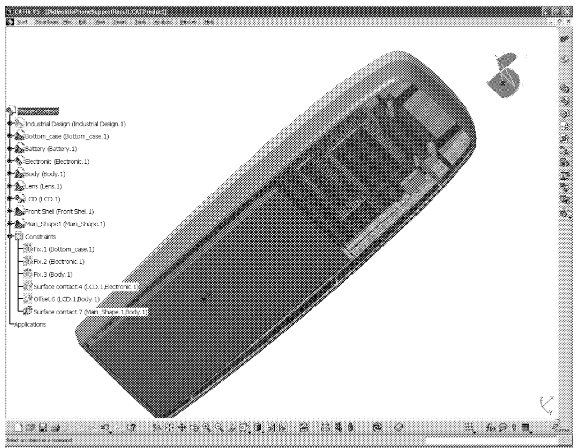


Figure 3: 3D-ME-DMU

A further requirement for the ME-DMU is to select the electrical components directly in the 3D-CAD model and to identify automatically the library component in the PDM system. An enhancement of this requirement is to not only find the component in the PDM system, but also to update the PDM system's meta data information based on examinations done within the 3D-CAD system. An example for this will be described in the chapter "Use Cases".

CONCEPT

In order to fulfil the requirements mentioned above, the PDM system needs to be enhanced by new functionalities and an extended data model able to cope with the mechatronic issues. Figure 4 gives a brief overview on the necessary top level data model elements.

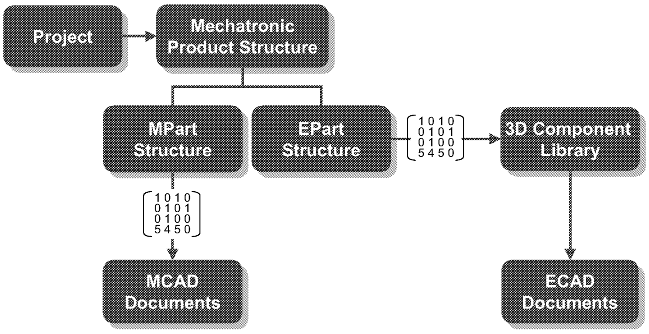


Figure 4: Data model

Starting with the definition of a project, the described mechatronic product structure needs to be defined. The mechatronic product structure is divided into the two domains mechanics (MPart structure) and electrics (EPart structure). For the mechanics domain the necessary documents such as 3D-CAD models, specifications or documentations are directly assigned to the product structure. If a 3D-CAD model is attached to the MPart item, additionally the transformation matrix is represented.

In the electrics domain the product structure is only build up with links to the electrical library. This concept ensures that every electrical product uses the same library elements with the same ID, nomenclature and revision. Additional ECAD-related documents, e.g. specifications, simulation results etc., can be linked to the library elements.

USE CASES

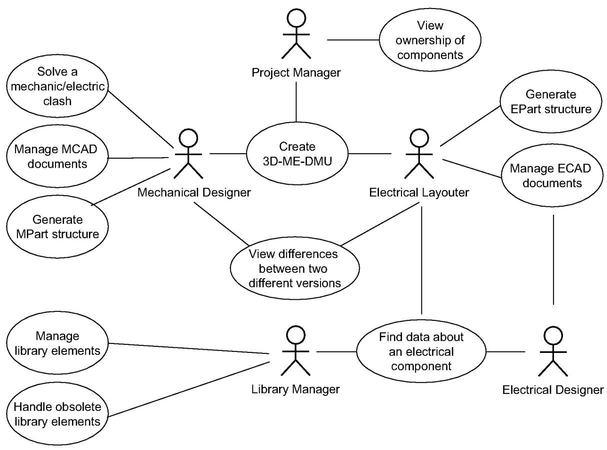


Figure 5: Use cases diagrams

In the following section different use cases will be described. The overview is shown in Figure 5 as an UML (Unified Modelling Language) use case diagram.

Creating the mechatronic product structure: This use case is the first step which is necessary in order to create the ME-DMU. The mechatronic product structure is created within the PDM system. The mechanic part of the structure is filled using the standard MCAD interface of the PDM system. The electric structure is created based on an additional import functionality which assigns the library elements to the EParts.

Create 3D-ME-DMU: In order to create a 3D-ME-DMU the following prerequisites are necessary in the PDM system:

- A 3D-representation of the electrical components is stored in the 3D-component library
- The 3D-models of the MCAD system are checked in and assigned to the MPart structure
- The electrical 2D-Layout information was imported, interpreted and linked to the 3D-component library and the EPart structure.

Starting from the MEPart it is now possible to assemble a mechanical/electrical structure in 3D based on the data stored in the PDM system, see Figure 6.

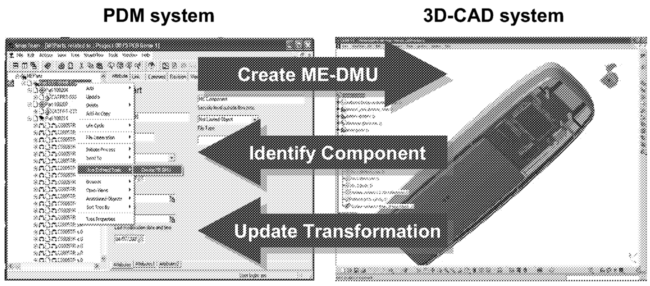


Figure 6: Interaction between 3D-CAD and PDM system

Solving a mechanic/electric clash: Based on the generated 3D-ME-DMU a person can detect potential clashes, using the standard 3D-MCAD functionality and solve them. In the next step he could decide if the changed, clash-free position of the components should be stored back to the PDM system. The developed function "*Update transformation*" stores the transformation matrix between the MPart/EPart and the linked documents back to the PDM system (see figure 4 for the data model).

View differences between two versions (delta visualization): In order to support the iterations between the mechanical and electrical department with an appropriate software system it is necessary to support the visualization of delta information. The user needs information about differences between two versions of the current product development to answer questions such as:

- Which components are new?
- Which components have been deleted?
- Which component position was changed?

Starting from the MEPart a function compares the elements of two different versions stored in the PDM system. The visualization of the differences is done in 3D with the following rules:

- Unchanged components are set to gray
- New components are set to green
- Deleted components are set to red
- Modified components are set to yellow

An example for the delta visualization is shown in figure 7.

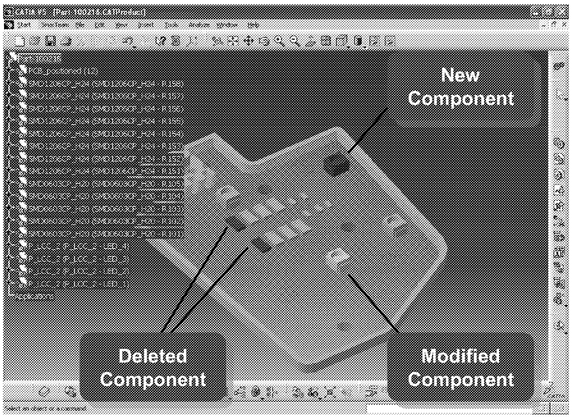


Figure 7: Delta visualization

Find data related to an electrical component: This use case happens if an electrical component supplier stops the delivery of a specific electrical component. The library manager could start an ME-DMU in order to identify the obsolete electrical component in 3D. After he has identified the component, the function "*Identify component*" shows the selected component instance in the 3D component library within the PDM system. In the next step he could use the PDM system's link management in order to find all assemblies (EParts) where the specific library element is used. With this information it is now possible to inform all EPart owners that a new version of the library component is available and should be used in the further product development.

BENEFITS

With the presented concept and software solution the following benefits could be achieved.

- Faster mechatronic product development due to a centralized access of up-to-date information
- Higher product quality due to a reduction of errors caused by wrong version information or component collisions
- Early 3D validation of the mechatronic product design and therefore cost reduction for error correction
- Reduction of redundant development work due to up-to-date information access
- Increased usage of standard components based on a common library
- Enhanced functionality for project management and project controlling due to a centralized multidisciplinary project database

CONCLUSION

In this contribution a current deficit of product data management arising from the insufficient processing of mechatronic data was pointed out. A data model concept was presented in order to handle the MCAD/ECAD data within the PDM system. By implementing the data model, the PDM system is capable of supporting the interdisciplinary product development process with new ME-DMU functionalities. It was shown that the requirements and use cases of the different roles within this process could be adequately supported. In summary, the MCAD/ECAD integration takes a crucial role in advanced virtual product development. Only by establishing an MCAD/ECAD data integration based on a PDM system the number of time-consuming and cost-intensive prototypes can be reduced significantly.

REFERENCES

Anderl, R.; Krastel, M.: Managing Mechatronic Simulation Models of Technical Products with PDM-Systems. In: Proceedings of the International Conference on Engineering Design, ICED 01Glasgow, 2001.

Kellner, P.; Meißner, M.; Müller, A.: Integration von CAx – Produktstrukturen – PDM-Systeme als Vorbereitung für virtuelle Produktentwicklung. In: VDI Bericht Nr 1487, Düsseldorf: VDI Verlag, 1999.

Kleiner, S.: Föderatives Informationsmodell zur Systemintegration für die Entwicklung mechatronischer Produkte. Dissertation TU Darmstadt, Shaker Verlag, 2003.

Krastel, M.: Integration multidisziplinärere Simulations- und Berechnungsmodelle in PDM-Systeme. Dissertation TU Darmstadt. Shaker Verlag, 2002.

Krebs, T.; Franke, J.: Konstruktionswerkzeuge für elektronisch/mechanisch integrierte Produkte. In: Elektronik 18/2005

McIntosh, K. G.: Engineering Data Management – A Guide to Successful Implementation. Berkshire: McGraw-Hill Book Company Europe, 1995.

Scheder, H.: Requirements of car manufacturers for Product Data Management in an extended enterprise. In: Proceedings of STEP Forum '97, München, 17.4.1997.

Sielaff, F.: Integrating Software Design in the Mechanical Product Development Process. In: Proceedings of the ProSTEP iViP Science Days 2003, 8th-9th October 2003, Dresden

Svensson, D.: Towards Product Structure Management in Heterogeneous Environments, Chalmers, Göteborg, Sweden, 2003

BIOGRAPHY

MARCUS KRASTEL studied Mechanical Engineering at the Darmstadt University of Technology. After his Diploma in 1997 he worked as a research assistant at the Department of Computer Integrated Design. His main focus was the development of a STEP based data model for the representation of mechatronic systems (MechaSTEP). In 2001 he founded together with three colleagues the :em engineering methods AG located in Darmstadt. Within the :em AG he is responsible for the business unit "business processes" with focus on supplier integration with CATIA V5 and SMARTEAM. Since two years he works as well in projects in the area of ECAD/MCAD integration.

FORMAL METHODS AND TECHNIQUES

THE EFQM SELF-ASSESSMENT MODEL IN PERFORMANCE MANAGEMENT

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

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

KEYWORDS

Performance management, EFQM, self-assessment

ABSTRACT

Organizations are seeking new, integrated systems that enable rapid changes through early identification of opportunities and problems, tracking of progress against plans, flexible allocation of resources to achieve goals, and consistent operations.

In performance management a new category of systems and processes have been setup linked with the need to integrate strategy and key indicators of performance into management processes, and to exploit technology to improve monitoring, management reporting, and decisionmaking.

Starting from the EFQM model we identified critical performance indicators that will result in improving business performance on the path to excellence. We developed a tool SCAN consisting of questions about 42 aspects. We developed a system SCANA reporting over the answers of the management team.

1. BUSINESS PERFORMANCE MANAGEMENT

Performance management should be a main concern of every company. Organizations are seeking new, integrated systems that enable rapid changes through early identification of opportunities and problems, tracking of progress against plans, flexible allocation of resources to achieve goals, and consistent operations. Traditional systems, including ERP and business intelligence software, are not sufficient by themselves to meet this demand. Out of these requirements has emerged a new category of enterprise automation technology: Performance Management.

The term Business Performance Management (BPM) is used to represent a new category of systems and processes and it pinpoints the need to integrate strategy and key indicators of performance into management processes, and to exploit technology to improve monitoring, management reporting, and decision making.

In research the following factors were identified as being good predictors of performance attainment: Strategy towards corporate and social responsibility, employee loyalty, staff responsiveness, customer satisfaction as well as market share, cash flow and costs. Predictors for good practice adoption include benchmarking, skill and job training, customer orientation, problem solving, and waste elimination.

2. THE NEED FOR A MODEL AND THE EFQM EXCELLENCE MODEL.

Regardless of sector, size, structure or maturity, to be successful, organizations need to establish an appropriate management framework. The EFQM Excellence Model is a practical tool to help organizations do this by measuring where they are on the path to Excellence, helping them understand the gaps and then stimulating solutions.

The EFQM Excellence Model was introduced at the beginning of 1992 as the framework for assessing organizations for the European Quality Award. It is now the most widely used organizational framework in Europe and it has become the basis for the majority of national and regional Quality Awards.

This model is a non-prescriptive framework which recognizes that there are many approaches to achieving sustainable excellence. It can be used as a self-evaluation tool for organizations, large and small, public and private sector.

2.1. Characteristics of excellence

The Model, which recognizes there are many approaches to achieving sustainable excellence in all aspects of performance, is based on the premise that:

Excellent results with respect to Performance, Customers, People and Society are achieved through Leadership driving Policy and Strategy, that is delivered through People, Partnerships and Resources, and Processes.

The fundamental concepts or characteristics of excellence are:
• results orientation
• customer focus
• leadership and constancy of purpose
• management by processes and facts
• people development and involvement
• continuous learning, innovation and improvement
• partnership development
• public responsibility.

The EFQM model is based on those fundamental concepts or characteristics of excellence

The characteristics of excellence have links to different evaluation areas and also to each other. Partnership development, for example, requires identification of partnerships, prioritization and setting objectives for partnerships such that they generate added value for customers.

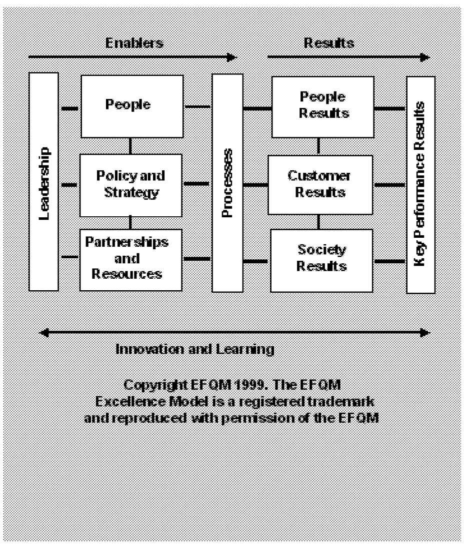
In addition, customer focus in vocational education and training requires identification of the needs of customers, such as students and the world of work, development of products and services based on these, and monitoring and analysis of customer results achieved. Results should be used as a basis to improve operations and set new objectives.

2.2. The EFQM excellence model

The EFQM Excellence Model is a non-prescriptive framework based on 9 criteria. Five of these are 'Enablers' and four are 'Results'. The 'Enabler' criteria cover what an organization does. The 'Results' criteria cover what an organization achieves. 'Results' are caused by 'Enablers' and 'Enablers' are improved using feedback from 'Results'.

The EFQM Model is presented in diagram form below. The arrows emphasize the dynamic nature of the Model. They show innovation and learning helping to improve enablers that in turn lead to improved results

The EFQM Excellence Model



2.3. Criteria and sub-criteria used in the model.

The Model's 9 boxes represent the criteria against which to assess an organization's progress towards Excellence. Each of the nine criteria has a definition, which explains the high level meaning of that criterion.

To develop the high level meaning further each criterion is supported by a number of criterion parts. Criterion parts pose a number of questions that should be considered in the course of an assessment.

1.Leadership
Leadership is even as important as products and processes are. Management can motivate and stimulate in the way to continuous improvement.

- How is management engaged in creating a culture of continuous improvement
- How is management supporting the improvement activities
- How is management evaluating and motivating the staff?

2.Policy and strategy
The EFQM is concerned not just with product and service quality but is concerning itself with organizational policy and strategy. Policy deployment to ensure that the strategy is formulated and is known to management, is important.

- The use of relevant information supporting the formulation of the strategy.
- The formulation of the strategy
- The implementation of the strategy
- The communication about the strategy
- The evaluation and the improvement of the strategy

3.people
EFQM covers aspects of training and service quality, but is goes further requiring effective human resource development, teamwork, empowerment, rewards and career planning.

- the organization of personnel management
- deployment of expertise
- participation of staff in the organization

4.Partnership and resources
Suppliers are becoming partners with emphasis on mutual beneficial relationships. Development and use of knowledge is point for attention. On point of resources facilities need to be maintained for capability.
- The financial resources to realize continuous improvement.
- How effective is the delivery of information
- Relation with suppliers and the procurement function
- The role of technology and knowledge management

5.Processes
The focus of EFQM is on the key processes necessary to deliver the organization's strategy. Quality processes are important too.

- identification of the processes
- control and management of processes
- evaluation and improvement
- incentives to innovation and to renovate the processes
- implementation of process re-engineering

6.customer appreciation
The major box requires evaluation of customer satisfaction through surveys and interviews

- customer satisfaction
- Loyalty
- customer focus

7.functioning of people in the organization
People are supposed to be surveyed with ideas such as team briefings and suggestion schemes to know their appreciation of the organization

- satisfaction survey
- functioning in the organization
- personnel administration

8.Position in the society
EFQM asks the company to establish its impact on wider society, for example involvement in community activities.

- role and link with society

9.company results
EFQM requires measuring the results of the company in a BSC way

- financial measures
- operational measures

2.3. Use of the EFQM model

- The EFQM model is a tool that organizations may use for the following purposes, among others:
  as a framework for **self-evaluation** that enables an organization to identify its strengths and areas for improvement and the extent to which its operations and results are in line with the characteristics of an excellent organization;
- As a way to **Benchmark** with other organizations
- As a guide to identify areas for **Improvement**

3. SELF-ASSESSMENT

EFQM believes that the process of self-assessment is a catalyst for driving business improvement. The EFQM definition of self-assessment is as follows:
Self-assessment is a comprehensive, systematic and regular review by an organization of its activities and results referenced against the EFQM Excellence Model. The self-assessment process allows the organization to discern clearly its strengths and areas in which improvements can be made and culminates in planned improvement actions that are then monitored for progress.

3.1. Benefits

Organizations have enjoyed various benefits as a result of undertaking self-assessment using the EFQM excellence model. Some of these included:

Providing a highly structured, fact-based technique to identifying and assessing your organization's strengths and areas for improvement and measuring its progress periodically
Improving the development of your strategy and business plan
Creating a common language and conceptual framework for the way you manage and improve your organization
Educating people in your organization on the Fundamental Concepts of Excellence and how they relate to their responsibilities
Integrating the various improvement initiatives into your normal operations

It is imperative when starting down the Self-Assessment path to be clear on what the desired outcomes are to reduce the risk of failure due to misplaced expectations.

3.2. How to do self-assessment?

There is no definitive answer to the question “which technique is the right one for my organization? There is no single “right” way to perform self-assessment.
We adopted the questionnaire approach.
This technique can be one of the least resources intensive and can be completed very quickly. It is an excellent method for gathering information on the perceptions of people within an organization. Some organizations use simple yes/no questionnaires, others use slightly more sophisticated versions that use a rating scale.
Self-assessment using standard questions designed to get the organization started thinking in terms of process improvement. Questionnaires can also be used to facilitate group discussions about improvement opportunities and to inform management workshops.

3.3. Our EFQM self-assessment tool SCAN.

The EFQM Excellence Model is a non-prescriptive framework based on 9 criteria. Five of these are 'Enablers' and four are 'Results'. The 'Enabler' criteria cover what an organization does. The 'Results' criteria cover what an organization achieves. 'Results' are caused by 'Enablers' and 'Enablers' are improved using feedback from 'Results'.
Till to now, the focus of our self-assessment tool is on the enabler or the organization criteria.
In fact we are assessing how the organization has been organized and how it is managed.
The results criteria of the EFQM model will deliver the structure for the performance measures using measurable performance indicators about the results.
Another point is that we structured our self-assessment questionnaire following the main business functions of the organization, and not the theoretical criteria of the EFQM model.
On the same way the performance measures or results measures will be structured following the same structure.

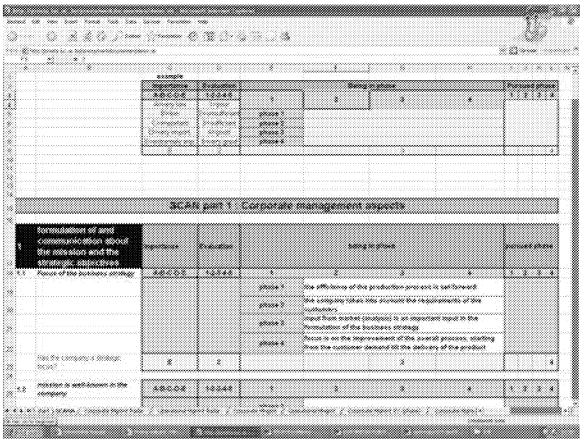
Structure of organization criteria:
I. Corporate management
1.mission, strategic objectives
2.planning and organization built upon the strategy.

3.Information and knowledge in the organization.
4.to strain after improvements
5.communication
6.personal leadership of management

II. Operational management aspects
1.employees and staff members
2.procurement and relation with supplier
3.customer and market share
4.finance
5.environment (society and environment)
6.business process and product/service

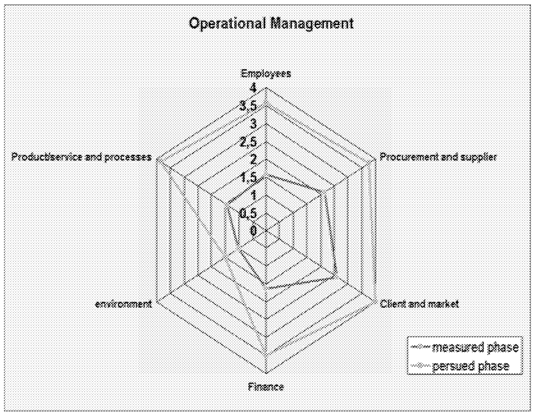
The management team, all of them being responsible for one or more business functions, will assess a set of selected management aspects. Future excellence will be dependent on them.

Our SCAN is a self-assessment instrument for the organization criteria. A questionnaire has been developed. A set of sub-criteria has been defined for each main criterion. A question has been formulated for each of the 42 aspects/sub-criteria. In each case we formulated 4 answers, presenting 4 fazes of growth to excellence for the aspect. The management team has to decide on:
1.What is the interest or relevance for the organization?
2.In which faze can the organization be positioned?
3.Which faze do they have in view?



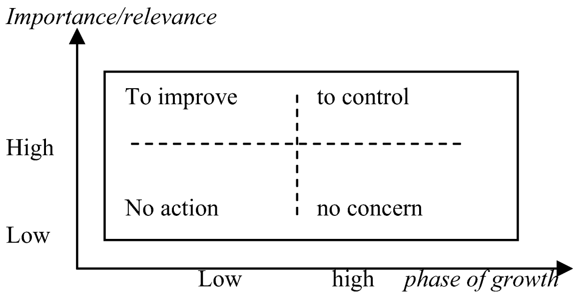
3.4. Our system SCANA

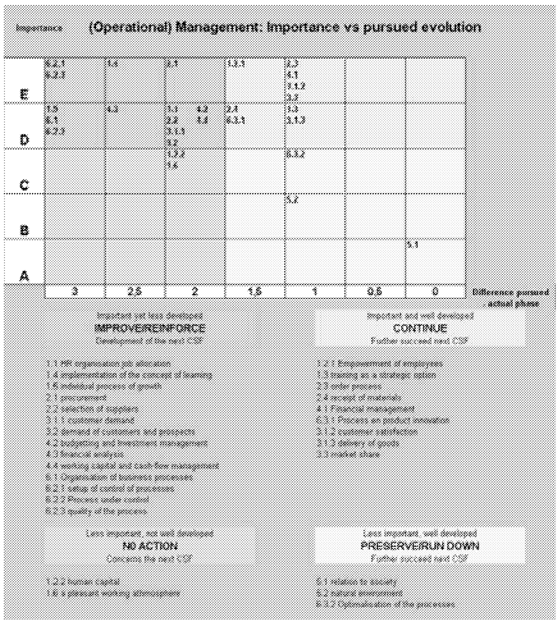
A system SCANA has been developed.
The self-assessment data from the management team can put into this system. A self-assessment management report will be produced at the end of the assessment session.
At the end of the session some control and analysis reports were produced by the system.
Two of them are shown in the following figures.
In a first report we can see the mean value of all main criteria being the mean value of the results of their sub-criteria aspects. The results can be compared with the faze they have in view too.



In the following figure we show this report, but we have taken the difference between the real measured phase and the faze in view for the future.

All aspects have been plotted on this chart.





The critical aspects are those marked as being in a low phase and being important at the same time. Those have to be improved.

The conclusions of this analysis form one of the main inputs in the discussion in the management team on how to continue.

4. THE LIMITATIONS OF THE EFQM SELF-ASSESSMENT.

Besides the EFQM excellence model, we can introduce the balanced scorecard (BSC) model.
Both tools are using measures of an organizations performance to drive organizational improvement, by highlighting current shortfalls in performance to management teams.
The BSC forms a much better basis for the development of a tool for the strategic management of an organization. Starting from the strategic goals the critical success factors (CSF) were identified and by the way the performance indicators linked with them.
Our suggestion is to mix the two models into one model. The EFQM model will have the primary role of self-assessment. The identification of the indicators and the reporting about the measures will be done in the BSC model.
The decisions about the priorities of adjusting actions can be based on a strategy map. This strategy map is the result of the built-in structure of the BSC model and of the tree structure of the CSF's as defined by the management team.

5. CONCLUSIONS

It is an opportunity for management to control the company (better) on its performance. Performance management should be on the agenda of every company.
Performance management is based on the identification of a set of indicators or business aspects and on their measures, delivering a good view of the performance of the organization
The EFQM model is a tool that uses measures of an organizations performance to drive organizational improvement.
We identified criteria or indicators to be measured. We developed a tool SCAN, being a self-assessment questionnaire following the main business functions of the organization. We developed a system SCANA that produces at the end of the self-assessment session some control and analysis reports.
Resulting from our experience in practice we decided to improve our performance management model by uniting the EFQM and the BSC model into one performance management model.

REFERENCES

H.Andersen, G.Lawrie, M.Shuiver:"The balanced scorecard vs the EFQM Business model", 2GC Working paper.

R.Kaplan and D.Norton "The Strategy-Focused Organization"; 2001

K..Monahan "Balanced Measures for Strategic Planning" A Public Sector Handbook; 2001

P.Niven "Balanced Scorecard Step by Step: Maximizing Performance and Maintaining Results"; 2002

A.Robson and V.Prabhu, "What can we learn from leading service practitioners", *Managing Service Quality,* Vol.11, No.4, 2001

J. Sandt, U. Schaeffer, J. Weber, "Balanced performance measurement systems and manager satisfaction", Otto Beisheim Graduate School of Management, 2001

J.Schreurs, M.Voets, R.Moreau: "Resultaat gericht managen van een KMO" in " Staten-generaal van de Limburgse economie en werkgelegenheid"; ed.J.Heeren & K.Milis p. 81-110; ISBN 9020956817

S.Ten Have: "Key management models"; Financial Times Prentice Hall 2002.

E. Wiersma, "De balanced scorecard: wetenschappelijke onderzoekresultaten", Handboek Management Accounting, 2001

M.Wijn, W.Hofenk, R.Hoekstra en M.Hengeveld " Kritische succesfactoren aan een nadere analyse ontworpen." 1994

www.efqm.org

Developing Concurrent Engineering Software: A Socio-technical Approach

Yacine Rezgui
Email: y.rezgui@salford.ac.uk

Abstract: The characteristics of international funded Research and Technology Development (RTD) projects, and more generally any initiative that pursues research and software development at the same time, require the use of an adapted methodology that factors the multi-disciplinary and multi-cultural nature of these projects while addressing their inherent high-risk dimension. The methodology presented in the paper has been developed and refined over a decade in the light of experiences made of successes and failures acquired through several international software developments research projects in the field of concurrent engineering. It combines the use of a high-level process modelling method, namely IDEF0, elements of Checkland's Soft System Methodology (SSM), and UML. The methodology is iterative and incremental and is based on Action Research principles.

1. INTRODUCTION

Research and Technology Development (RTD) projects tend to involve a diverse number of partners ranging from industry end-users to research and academic institutions with varying levels of software development maturity and capability. In this context, the challenge has always been to use the right Information Systems Development (ISD) methodology that addresses the high-risk element of the project while ensuring a successful completion of the research and achievement of the project objectives. As noted elsewhere [1], ineffective ISD methodologies are an important risk category that is of major concern to software development managers.

The paper is based on the author' personal experience and involvement over a decade in multi-national Information and Communication Technologies (ICT) based concurrent engineering projects both in the capacity of project coordinator as well as RTD partner. The paper describes an ISD methodology, which was developed and refined over several funded RTD projects. The paper, first, provides an overview of the issues that arise in the selection of an appropriate ISD methodology. This is followed by a brief description of the RTD projects used as case studies for the purpose of this research. Lessons learnt from these projects are presented, followed by a description of the proposed methodology. Reflective conclusions are given at the end of the paper.

2. ISSUES IN THE SELECTION OF AN ISD METHODOLOGY

A large and increasing number of ISD methodologies are available for use on any given project [2]. The variety and scope of these methodologies present a difficult choice as to which is most appropriate for use in a particular project. Commonly, a methodology will be chosen according to a range of factors including, but not limited to, implementation languages, availability of suitable tools, previous experience and needs of the current project.

A range of factors may influence the successful use of a particular methodology on a given project. These can be viewed as factors relating to the suitability of the methodology in capturing and manipulating information particular to the project, and those, which relate to the soft or human aspects. These factors can be viewed in terms of distinction between expressiveness and effectiveness of an information representation. Applied to the context of ISD methodologies, we can define expressiveness as being satisfied if the methodology can be used to capture and manipulate the desired information. Effectiveness is achieved only if the methodology fully exploits the capabilities of all of the stakeholders in achieving their tasks. It is more usual for the information specific needs - expressiveness - to be focused on when selecting and evaluating a methodology but we argue that this overlooks the effectiveness aspects which may prove more critical to its success.

Human cognition is one factor that may influence the effectiveness of a particular methodology. Successful completion of a task is influenced by the use of different notations for creation of a representation [3], and it follows that the structure and presentation of information within a particular methodology may influence its effectiveness. A number of authors recognise that the choice of languages influence the success of software system development [4,5]. Studies have shown that user performance with UML (Unified Modelling Language) diagrams may be influenced by individual preference [6,7,8]. Additional human (or 'soft') factors, which may influence the successful use of a methodology, include the experience of analysts, modellers, users and other stakeholders. Also the ability of those involved to articulate relevant information within the constraints of the methodology may be another factor of influence. The methodology described in this paper is not examined from a cognitive viewpoint. However, it is clear from the analysis of the refinement of the methodology, that human factors influence its successful use.

In the context of the paper, the chosen methodology must satisfy the following requirements:

- Enable collaborative, multi-user development of the software solution and handle group dynamics.
- Take into account the multi-cultural (including multi-lingual) and multi-disciplinary dimension of research project consortia. A typical RTD project involves at least five partners from three different countries.
- The methodology should be flexible and comprehensive enough to accommodate cultural and work differences between industry, research and academic partners.
- The solution must be developed incrementally with high participation, involving the end-users.
- It must support research and software development at the same time.
- The developed solution is only a prototype. Proper steps should be allowed for turning this into a usable product.
- There is no unique prototype solution. In
- fact, generic tools are developed which are then deployed in the form of prototype demonstrators within the context of each participating end-user organization.
- Legacy and current systems in use within each organization are highly likely to be part of the final demonstrator solution.

These requirements have acted as the primary constraint on the choice of methodology in each project, and it can be seen that the first four of these requirements have a strong human influence. Careful consideration has also been given to the influence of human factors as the methodology has been refined, and these are discussed within the context of the relative success of the methodology within each project. In the following section we outline the projects and detail the methodology used in each.

3. DESCRIPTION OF THE CASE STUDIES AND ASSOCIATED METHODOLOGIES

Four case studies that span a period of over ten years form the focus of this study. While these exhibit a common similarity in that they target advances in ICT applied to the construction sector, they present some interesting differences in terms of consortium partners' profile (Figure 1) as well as the level of ISD experience of the team members (Figure 2). A brief description of each RTD project chosen project is given below:

- The ATLAS Project (1992-1995): ATLAS aimed to develop a common open platform to support information sharing and exchange between diverse applications used by the wide range of large and small businesses involved in large-scale engineering, with a targeted application in the Construction domain.
- The CONDOR Project (1997 – 1998): The aim of the project was to develop means of bridging the gap between the traditional document-centred approach to electronic document management (whereby documents tend to be poorly structured and handled as black-boxes) and a proposed model-based approach promoting highly structured documents. Robust models supporting this transition have been developed, together with a prototype implementation, which demonstrated a pro-active use of document

management techniques in a collaborative environment using end-users' native electronic document management systems.

- The OSMOS project (2000-2002): The overall aim of the OSMOS project was to enhance the capabilities of construction enterprises to act and collaborate effectively on projects by setting up and promoting value-added Internet-based flexible services that support teamwork in the context of a Virtual Enterprise (VE).
- The eCognos Project (2001-2003): The general aim of eCognos was to specify, implement and evaluate an open ontology-based infrastructure and a set of services that promote knowledge management in the Construction sector.

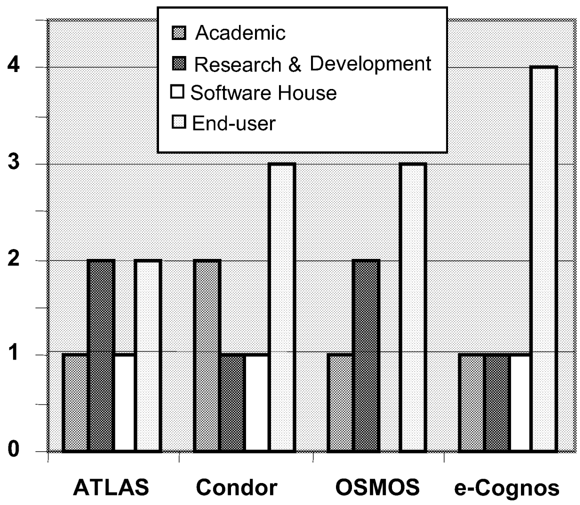


Figure 1. Team profile for each case study

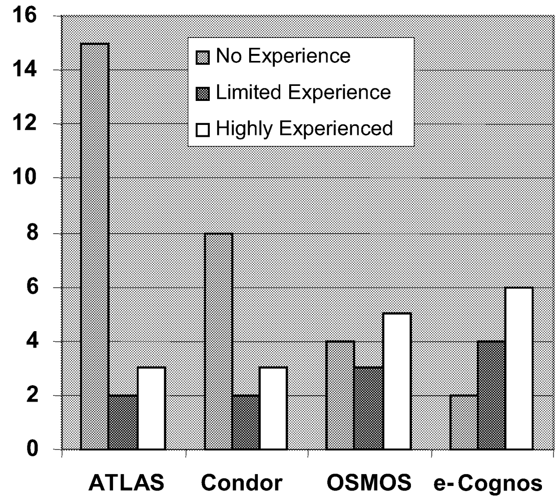


Figure 2. Team level of experience with using ISD.

4. CRITICAL EVALUATION OF THE METHODOLOGIES EMPLOYED WITHIN THE FOUR CASE STUDIES

As rightly pointed out in [9,2], evaluating an ISD methodology is not an obvious task. This is today more an area of research [10,11,12] than an industry best practice. This section provides a snapshot of the methodologies used in the four previously described projects. Table 1 provides an overview of the scope of the methodologies used within each project case study.

	ATLAS 1992-1995	CONDOR 1997-1999	OSMOS 2000-2001	eCognos 2001-2003
Strategy	Ad-hoc based on the STEP Standard	Used initially the Waterfall model then moved to using the Boehm's spiral model.	Rational Objectory Process using three iterations.	Unified Process using two iterations.
Feasibility	No feasibility	No feasibility	No feasibility	Carried out at the inception stage using some aspects of SSM
Analysis	IDEF0	IDEF0	IDEF0/ UML Use Cases	Elements of SSM + IDEF0/ UML Use Cases
Logical Design	NIAM / EXPRESS-G	Class Diagrams	UML Interaction Diagrams / Class Diagrams	UML Interaction Diagrams / Class Diagrams
Physical Design	EXPRESS	IDL	IDL	IDL
Programming	C++	C++	Java	Java / Soap/ RMI
Testing	Conducted at the end of the project	Conducted at the end of the project	Conducted at the end of each iteration	Conducted at the end of each iteration
Implementation	Client-Server Environment	CORBA-Component-based implementation	Proprietary version of the Web Services Model as the latter was still being specified at the time	Web Services Model
Evaluation	Carried out at the End of the Project	Carried out at the End of the Project	Carried out at the end of each iteration	Carried out at the end of each iteration
			Ongoing within project	Ongoing within project

Maintenance	N/A	N/A	partners currently exploiting the solution	partners currently exploiting the solution

Table 1. Overview of the project methodologies.

5. PROPOSED METHODOLOGY

A methodology has been formulated that best utilised the valuable 'real-world' contributions that exist amongst the project participants. To be successful, ISD methodologies must be well designed, robust, and more importantly, both supportive for adaptation, and effective in the context of human factors.

Action Research principles within the context of the Unified Process Methodology are employed to conduct the research. This is by essence iterative and incremental. A minimum of two iterations should be planned. The first should be short (6 to 9 months), and should be dedicated to securing consortium overall cohesion, sharing of common goals, objectives, and vision, raising awareness about the project in the end-users' organizations, and securing management support. Each iteration should include: Feasibility, Analysis, Specification, Implementation, Deployment, Evaluation and training, and Maintenance.

The choice of Action Research is rationalised by the pragmatic, 'real world' nature of the purpose of the research. It integrates theory and practice through change and reflection. This approach provides the following attributes: Responsiveness (the approach is flexible and adaptable to change), Iterative (the early stages help to plan and decide how to conduct the later iterations. In the later iterations, the interpretations previously developed have been tested, challenged and refined), Participative (the end-users are heavily involved in the research process), Qualitative (the use of qualitative information during the research increases responsiveness from the end-users), Reflective (critical reflection upon the process and outcomes are important parts of each iteration).

Figure 3 illustrates the key steps involved in the proposed methodology. These are further detailed in the following subsections.

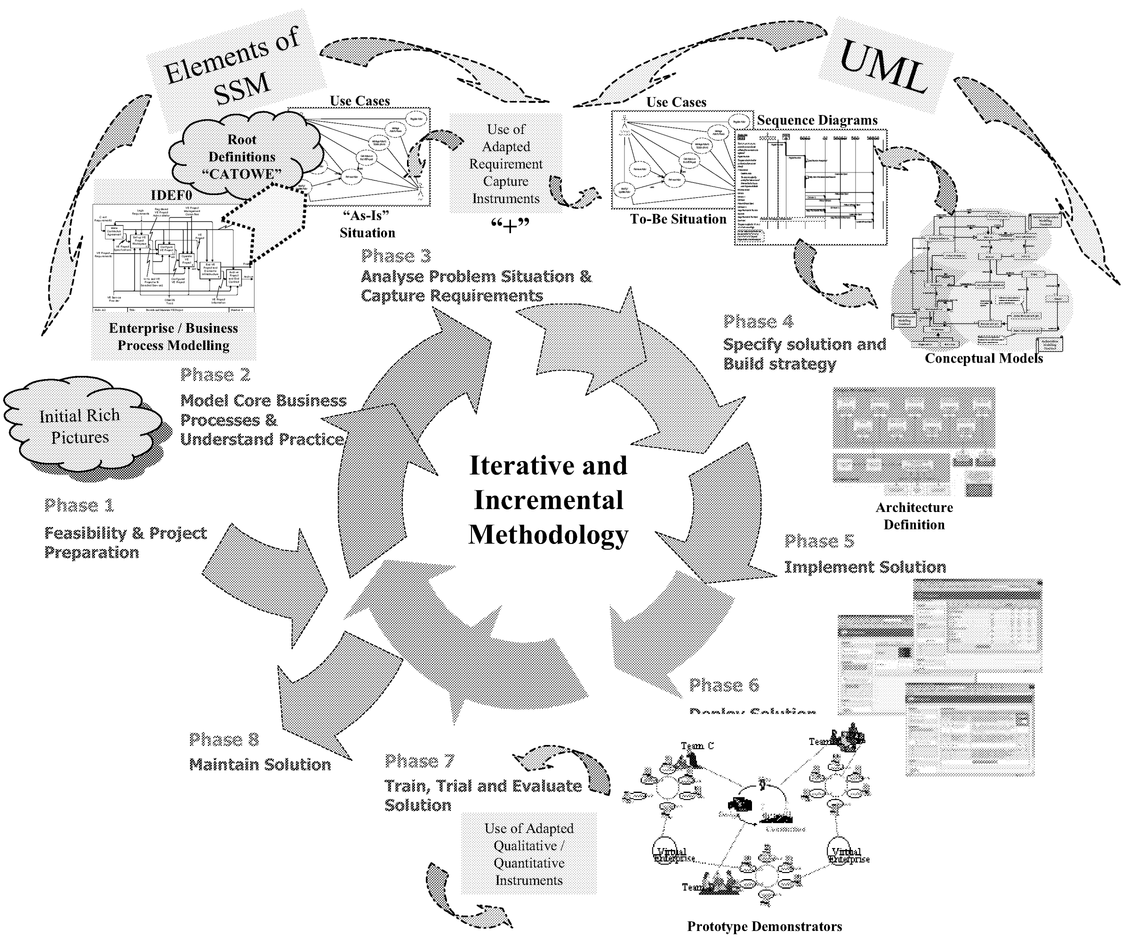


Figure 3. The proposed iterative and incremental methodology.

6. Feasibility and Project Preparation

An initial feasibility study should be carried out where the project objectives and vision are revised, and risks reassessed. Also, the fit and match of the project partners to the revisited vision and objectives should be reassessed. Potential partners that represent a high risk should be identified and their role reviewed. Adapted plans should be devised to ensure successful analysis and specification of the solution that best utilizes the resources and skills available within each organization. The latter may have had poor experiences with previous similar initiatives, which can lead to management and employee resistance to both the implementation and use of any new system. Therefore, this stage is crucial to the potential success of the project, as it aims to build an initial understanding and commitment to the research initiative. The key aims of this stage may be summarised as follows:

- Build management understanding and commitment to undertake and pursue the research.
- Introduce and explain the project to company staff.
- Assess risks related to the project, and find out about company past initiatives (successful and unsuccessful): KM, BPR, etc.
- Capture success and / or failure factors.

The end-user organizations need to be educated to ensure a successful environment for the research. Ideally, representatives of the development team should spend some time within the end-user organizations ensuring that they are fully in line with the objectives of the research.

Analysis

Initially, this stage should attempt to achieve the following objectives:

- Understand current company strategy: In what direction the company is going – what are their likely current / future strategic requirements?
- Understand the structure of the organization: Division of the work, the tasks, and the responsibilities both horizontally and vertically.
- Understand the culture of the organization: Values, norms and views shared by employees.
- Understand existing Systems: Rules, Procedures, Guidelines, Software / Hardware systems in use.

This can be facilitated by the use of holistic techniques such as, on the one hand, Rich Pictures [14] for understanding the social, organizational and political aspects that characterize each end-user organization involved in the research complemented by Root Definitions of the identified systems using CATOWE checklist techniques; and, on the other, enterprise modelling techniques such as IDEF0 to provide a process mapping and organizational context in which the solution will be developed. The IDEF0 functional modelling technique is used as its non-temporal representation of decisions, actions and activities, and the relation between them, strongly supports the detailing of activities in terms of process overview, workflow management, interaction analysis, and resource allocation. Dedicated instruments for capturing requirements might be devised and adapted based on the culture in place within each organization as analysed in the previous stage.

The next phase involves extending and analysing the developed Rich Pictures and Root Definitions and attempt to better comprehend the problem situation. This provides a stronger basis for UML Use Case development which are employed to detail ways in which the solution to be specified can be used at a business level, and to derive the required functionality of the system. In fact, SSM (need to expand SSM) root definitions help identify use cases' actors, while IDEF0 activities complemented with SSM interaction models (derived from conceptual models) help identify and describe use cases. Therefore, UML Use Cases describing the "as is" (to be done by the project end-users) and "to be" situation (to be led by the project analysts with support from the end-users) are developed.

Specification

The Use Case descriptions are used as the bridging link between the requirement capture and the system specification. With the Use Cases defined, sequence diagrams and class diagrams are employed to specify the internal system design, and provide the foundation (including models and APIs) to develop the software solutions. A detailed architecture of the system must be devised, with the different constituent components clearly identified and defined. The specification stage should be done in line with UML recommendations.

Implementation

The implementation stage consists of (a) writing the code that would provide an implementation of the API (need to expand) , and thus of the services of the software application; (b) develop the front-end that would be used as an access-point to all or a subset of the services; and (c) extend and augment key corporate EIS (need to expand) /Legacy applications that are expected to be part of the final demonstrator solution by adding functionality from the developed services.

Obviously, testing takes an important dimension through all this process, and should be done at the right level of granularity. Services are discrete software components and should form the basis for carrying out testing. This should be done incrementally across the planned iterations.

Deployment

The purpose of this stage is to produce the demonstrators that will form the solutions to each end-user organization. This is done by assembling the generic software tools developed within the project with a selected number of relevant end-user applications augmented with additional functionality developed in the previous stage.

Training and evaluation

The principal means of evaluating the resulting demonstrators and approach is provided through field trials simulating work in a real business context. Criteria need to be determined for technical, social and economical evaluation of the demonstrators together with other considerations, including legal, contractual and organisational factors. An end-user team should be selected to conduct the evaluation of the system. The end-users should be trained collectively as a group for the generic aspects of the demonstrator, and individually focussing on the specific services to be tested. Scenarios should be planned, detailed and handed to each evaluator describing work to be carried out on the chosen services that should reflect real functions carried out on projects. An evaluation form should be completed by the evaluators, and fed back to the design team.

Maintenance

The resulting demonstrators at the end of the research project should be treated as an advanced prototype system. Additional steps should be carried out at the organizational level to turn this into a robust system that can be deployed in a real business environment. The end-user organization might take over the exploitation of the system and work closely with a software house to produce this final system and agree on procedures to ensure maintenance and extension.

7. CONCLUSIONS

ISD methodologies tend to be perceived as offering generic solutions to class of practical problems [12]. However, the organizational context in which the ISD methodology is applied might require some form of re-adaptation for it to be effective. Furthermore, as highlighted in the OSMOS case study, the ISD methodology had to be re-considered and adapted, and the end-users from the participating organizations educated, for the methodology to be successful. Therefore, those who will apply the proposed methodology must be sensitive about the possible difficulties they may face when confronted with the contextual conditions of the participating organizations. Hence the usefulness and relevance of action research in that the researchers / software analysts are immersed within the work environment and practice where the solution is expected to be found, and reciprocal effects do take place between them and the practitioners. Furthermore, the course of action research cannot be predicted at the outset and its strength lies in its ability to deal with the emergent nature of human systems [12].

The authors hereby argue that the methodology presented here satisfies both expressiveness and effectiveness criteria. While variants of the proposed methodology have been, and are still being, used extensively in national and EC funded projects, mainly through the involvement of the partners in several companion ICT projects, this is the first ever attempt to formalise the methodology building on lessons learnt through its use and deployment in a variety of RTD projects, including the ones described in the paper. It is hoped that this will provide the ICT funded project research community with a helpful method and recommendations on the conduct of similar initiatives that pursue software development and research at the same time.

8. REFERENCES

1. Lyytinen, K., Mathiassen, L. & Ropponen, J. (1998) Attention shaping and software risk: a categorical analysis of four classical approaches. Information Systems Research, **9** (3), 233–255.
2. Avison, D. & Fitzgerald, G. (2002) *Information Systems Development: Methodologies, Techniques and Tools*, 3rd edn. McGraw-Hill, Maidenhead.
3. Scaife, M, & Rogers, Y. (1996) External Cognition: How do Graphical Representations Work? International Journal of Human-Computer Studies, 45, 185–213.
4. Roast, C. (1997) Formally Comparing and Informing Notation Design. Proceedings of HCI'97: People and Computers XII. In: Thimblely, H, O'Conaill, B, & Thomas, P (ed.). Springer.
5. Britton, C, & Jones, S. (1999) The Untrained Eye: How Languages for Software Specification Support Understanding in Untrained Users. Human Computer Interaction, 14, 191–244.
6. Purchase, H.C., Allder, J.A. and Carrington, D. (2000). User Prefernce of Graph Layout Aesthetics: A UML Study. Proceedings of the Graph Drawing Symposium, J. Marks (Ed). LNCS, 5-18, Springer Verlag.
7. Britton, C., Kutar, M., Anthony, S., Barker, T., Beecham, S. and Wilkinson, V. (2002) An Empirical Study of User Preference and Performance With UML Diagrams. Proceedings of IEEE 2002 Symposia: Human Centric Computing Languages and Environments (HCC02), Arlington, Virginia.
8. Swan, J., Kutar, M., Barker, T. and Britton, C. (2004) User Preference and Performance With UML Interaction Diagrams. Proceedings of the IEEE Symposium on Visual Languages and Human Centric Computing (HCC04), Rome.
9. Brinkkemper, S., Lyytinen, K. & Welke, R. (1996) Method Engineering: Principles of Method Construction and Tool Support, Chapman & Hall, New York.
10. Tolvanen, J.-P., Rossi, M. & Liu, H. (1996) Method engineering: current research directions and implications for future research. In: Method Engineering, Principles of Method Construction and Support. Brinkkemper S., Lyytinen K. & Welke R. (eds), pp. 296–317. Chapman & Hall Publications, New York.
11. Kelly, S. & Rossi, M. (1998) Evaluating method engineer performance: an error classification and preliminary empirical study. Australian Journal of Information Systems, 6 (1), 39-47.
12. Grant, D. and Ngwenyama, O. (2003) A report on the use of action research to evaluate a manufacturing information systems development methodology in a company. Info Systems Journal, 13, 21–35.
13. Martin, J. (1989) Information Engineering, Prentice Hall, NJ.
14. Checkland, P. and Scholes, J. (1990) Soft Systems Methodology in Action, John Wiley & Sons, Chichester.

ENGINEERING PROCESS MANAGEMENT

Agents and Their Behavior's Reconfiguration

Michal Radecký
Ivo Vondrák
Faculty of Electrical Engineering and Computer Science
VSB-TU Ostrava
17.listopadu 15, Ostrava, Czech Republic
e-mail: michal.radecky@vsb.cz, ivo.vondrak@vsb.cz

KEYWORDS
Agent, MAS, Behavior, Process, UML, Activity Diagram

ABSTRACT

The demands for modern, powerful and advanced systems or applications are increased, nowadays. Several current solutions and technologies can be used to comply with these enhanced requirements. The *Multi-Agent System (MAS)* technology is one of these solutions.
A recent point of view on software development process offers many tools and methods for specification, analyze, design, implementation and maintain applications. These facilities can be used for standard applications development, as well as for the systems based on MAS technology. In the case of the multi-agent systems, some of these standard approaches have to be adjusted or extended. This paper describes the ideas and methods to model and develop MAS, based on the internal agent behavior and intelligence of particular agents. This paper is concerned with analyze of the MAS from the agent internal behavior point of view.

Introduction

The MAS technology is quite youthful field of computer science that occurs on the borders where "Software Engineering", "Artificial Intelligence" and other computer, natural and social science areas meet. The MAS technology is based on the concepts of the Complex Systems (e.g. macromolecules, ants colony, economical systems), and on the facilities and capabilities of software information systems (Kubik 2004).

The essential properties of MAS are

- autonomy and intelligence of elements
- communication among elements
- mobility of elements
- decentralization of the control
- adaptability, robustness, etc.

The MAS can be formed as an general information system that is composed from a number of autonomous elements (called *Agents*). In this context, the Multi-Agent System is a framework for agents, their lives, their communication and mobility and it is an environment where the goals of the particular agents should be obtained too. From the perspective of the standard information systems, the *Agent* are components of such systems. These components have a special features like autonomy or intelligence, which is consistent with the essential properties of MAS.

Thus, the *Agents* are the elements of MAS. It is a software entity created in order to meet its design objectives. These objectives are subordinated autonomously with respect to the environment, sensorial perceptions, internal behavior and to the cooperation with other agents.

The typical applications of this multi-agent technology are the large systems with a big volume of decentralization and autonomy of their elements. The examples of such systems could be a transportation and logistical system (vehicles, drivers, dispatcher, etc. are the autonomous and world-wide distributed elements of MAS), system for manufacturing line (robots, transport bands, etc. are the Agents) or some kinds of monitoring system (network elements, computers, etc. are the watched or watch elements). The overall behavior of whole system, that is based on the MAS technology, is formed by internal behaviors of several elements (Agents) and also by the communication between all of these elements.

Classes of the Agents

The essential element of MAS is an Agent. The several types of agents could be found in one Multi-Agent System as well as in the "real world" which is realized by this MAS. These types of agents are able to denote as *Classes of the Agents*, according to the goals, internal architecture and behavior of particular agents. These classes are shown in the figure 1. Each Agent can belong to a given class for its whole life, but the classification of the Agent can be also changed during its life process, as a reaction to the situations, intelligence, capabilities and behaviors. We speak about "behavior classification" in the case of changing during the agents life (Radecky and Vondrak 2005).

	proactive	reactive
deliberative (route-finding)	**10%** (production agents)	**90%** (classical agent models)
processional (route-discovery)	**90%** (classical agent models)	**10%** (reconfiguration models)

Figure 1: Agents classification

The classes that describe a processional approach in the combination with reactive or proactive perspectives are the main goal of this research. These classes cover the concept of "Intelligent Agents" that is formed on the principles of an internal agent behavior processes. It means that the "classical agent models" and "reconfiguration models" are the subjects of our research.

Modeling of the Agent Behavior

Each Agent is determined by its own objectives and the way to meet these objectives is founded on the internal behavior of a given Agent. Internal behavior of such Agent is specified by the algorithm. The Agent lives, behaves and reacts to stimulus and environment, according to the requirements of the algorithm. Any Agent in the MAS has the main internal life process consequent on the agent's classification. Several templates for this *Primary Process* of agent behavior can be defined for each processional class of the agent ordination. This primary process can be decomposed into a number of sub-processes that refine the internal behavior on. The changing of agent classification during its life is based on sufficient template selection for each process, sub-process as well as primary process. This template selection must take the algorithm structure into account. Nevertheless, the template for primary process is given by default agent classification.

The algorithm structures respective to the particular templates could be follows:

- *Simple proactive agent or behavior* – sequential algorithm structure (just one thread)
- *Simple reactive agent or behavior* – algorithm structure with loop (just one thread processes the incoming requests in the loop)
- *Hybrid reactive/proactive agent or behavior* – parallel algorithm structure (more than one thread, where each thread can be realized as a proactive, reactive or hybrid behavior independently)

It is necessary to take into account the fact that each Agent is an absolutely autonomous element of MAS and thus the internal behavior have to be based only on the processes, activities, knowledge and facilities that belong to a given Agent. Then, the result behavior of whole MAS is formed by communication of separated agents and by the interconnection of several internal agent behaviors. This interaction is realized through the use of message passing adapted to the demands of MAS.

In the context of MAS modeling and agents behavior, the term *Agent* expresses only the "type of agents". The real separate agents are the instances of this type. It is analogous with the terms Class and Object from the object oriented approaches. The real agents (instances of Agent Classes) are not issues of MAS design and modeling phase but they will appear in the implementation phase, simulation and operation of a given MAS.

UML Activity Diagrams Extension

The UML (Unified Modeling Language) is an essential tool for process modeling, both on the business level and analytic level of description (Vondrak 2004; Pilone 2005). It can be applicable for modeling of the internal behavior of agents as well. The UML Activity Diagrams is a standard diagrammatic technique which describes the series of activities, processes and other control elements that together express the algorithm from the dynamical point of view. They are able to capture internal behavior of each Agent, but they are not able to describe the interactions and communication between them. The UML Activity Diagrams are especially suitable for modeling of the agent behavior, though some modifications and extensions are required, in terms of interaction support.

The *Agent Behavior Diagrams* could contain all the elements of the standard UML Activity Diagrams, and moreover some new elements that can be used in the agent modeling process. These new elements are concerned with message passing between agents or with other specific attributes of MAS. In early phases of development, these extensions are supported by the implementation of special "send/receive activities" which include an additional information about message content and message receiver/sender identification, see figure 2. The decision elements from standard activity diagrams are improved too. The modified "decision elements" and their output edges can hold the extra information. This information is usable for message-based determination of following control flow of the behavior led to the agent's objectives.

The simple, clear and formal definition of the internal agent behavior is a precondition for moving to the next phases of the multi-agent software process. Thanks to the new stereotypes established in the *Agent Behavior Diagram* and other additional information, it should be

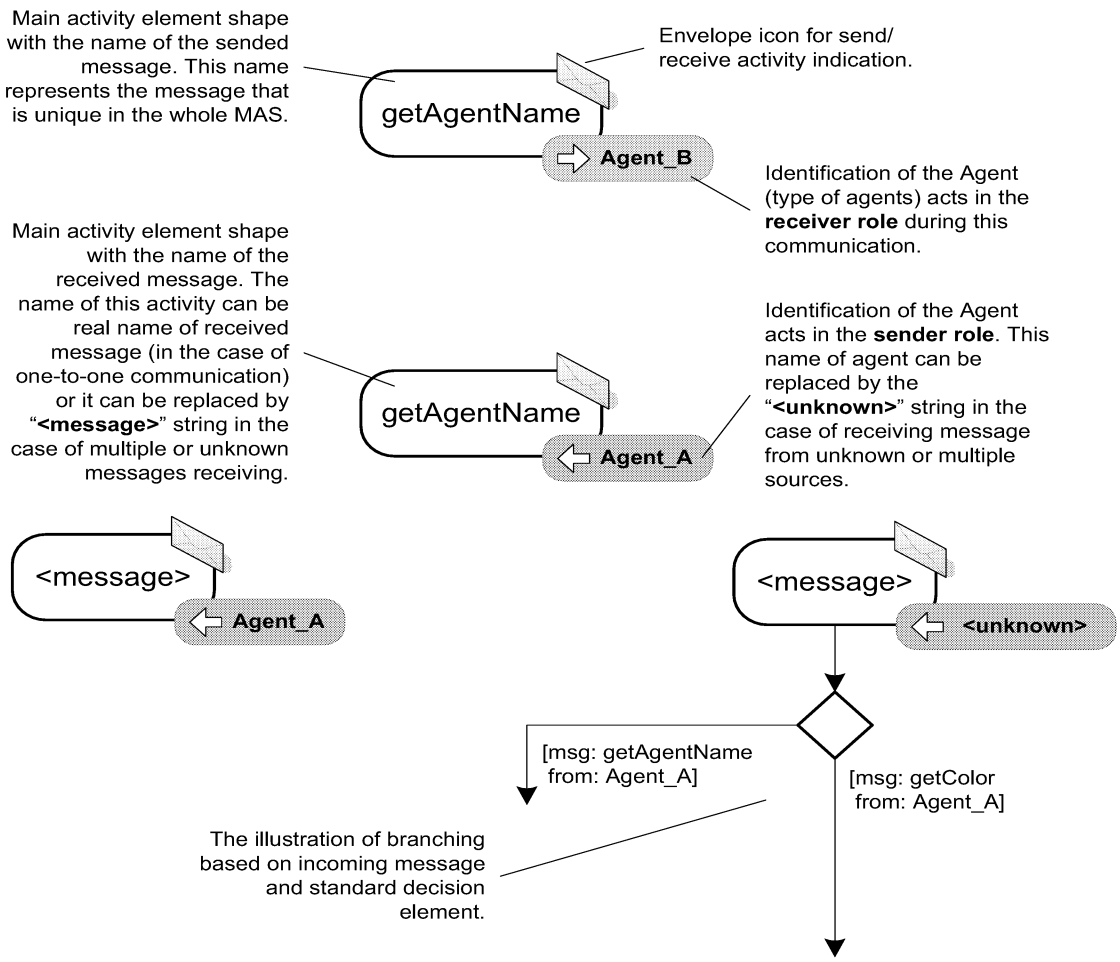


Figure 2: The stereotypes of the new communication activities and illustration of decision point. These send/receive activity nodes are able to use beside the other standard activity and process elements. Internal implementation of these new activities is concerned only with preparation, en/decapsulation and receiving/sending of the message.

able to generate other types of diagrams (e.g. sequential diagrams, maps of the agents communication) as well as the source code templates of the agents and MAS automatically (e.g. interfaces of agents, class and method templates).

As it is mentioned above, the Agent internal behavior is specified by the algorithms expressed in the processes. Each process is modeled as just one *Agent Behavior Diagram* based on UML Activity Diagram. The process can be specified by more than one diagram in the case of multiple realizations belong to a given process. The realizations are the topic of the following text. A couple of rules are bound together with creation process of these diagrams. At first, each process, as well as diagram, have to have just one "initial node" and just one "final node". This prerequisite is necessary for the further connection of the processes together to model overall agent behavior. The demand of the "well-formed" diagrams is also required. There is a set of general structural rules (e.g. the level of split/join nodes preservation, no crossing of the levels of control flow) that should be kept in mind during the drawing of diagrams (Aalst 1997). Thanks to these rules, the well applicable expressions of algorithms are obtained. It is able to verify them, to transform them to other forms or to process them by several formal tools.

Internal Structure of the Agents

The figure 3 illustrates an example of the processional structures of two agents. Each Agent must have one *Primary Process* that covers whole life of a given Agent. This primary process and its realization are unique in the context of a given Agent. Each process, as well as this primary process, can contain a control structures, activities (atomic elements of algorithm) and references to other processes. The usage of "send/receive activities" is the only one way to connect all behaviors of separate agents together. The activities and processes,

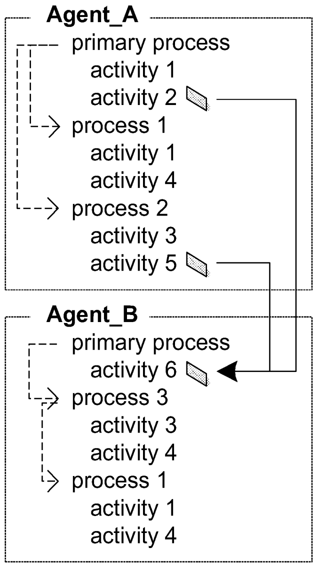


Figure 3: Illustration of the internal agent's structures

except primary processes, can be specified globally in the whole MAS. These globally specified elements are possible to subsume into the behaviors of a set of Agents (types of agents).

Demonstration of afore mentioned situation
For instance, the "moveToPlace" process (`Process 1`) or the "startEngine" activity (`Activity 1`) could be an examples of this approach. In this case, `Agent_A` could be a "Car" and `Agent_B` could be a "Motorbike". So, both vehicles have the same process of moving to a specified place and the same activity of engine starting. These identical elements are subsumed into a number of behaviors; therefore these elements can be specified only once and they can be used multiply. The activities are able to use in a number of processes within the frame of one Agent (`Activity 4` within `Agent_B`) as well as within the whole MAS. The correspondence between activities and processes is based only on definitions, purposes, goals and input/output objects of them but not on the real implementation behind them. This approach is utilizable only for the specification level of MAS modeling and it brings the reusing possibilities of the model elements.

Intelligence within the Agent Behavior

Only the static and structural modeling of agents is mentioned above, though, we want to speak about *Intelligent Agent* too. The *Intelligent Agent* is a standard agent that disposes of certain kind of "brainpower" during its life. These capabilities are hidden inside the agent behavior and they can be found in various points and situations of such behavior. The intelligence is ensured by application of several tools based on logic, artificial intelligence, etc.

The intelligence within the agent behavior can be concerned with three tasks:

- *Intelligence contained in the activities* – this usage is a question of activity implementation phase. The function of the logic is concerned with decision making and derivation within the activity, e.g. weather forecast. This application of intelligence is hidden from the modeling perspective. Only the result of activity execution is relevant for the next parts of process or whole agent behavior.
- *Intelligence of the control flows routing* – this application of several logic or intelligent tools is covered in the decision points. The decision making and derivation are activated whenever the "intelligent decision point" is reached during the process execution. The intelligent routing can be used for all branching that request more complex and knowledge based decision making, e.g. suitable car to a given cargo assignment. The following direction of process control flow depends on the outgoing result of the decision point. The branching is realized by the output edges of decision point and by the edges conditions which represent the possible outputs of the logic tools execution. It is able to substitute this approach with one special activity (with intelligence within it) succeeded by one decision point that uses the results from the previous activity. But the usage of only one "intelligent decision point" could be more clear and straight approach.
- *Intelligent selection of Process Realizations* – The third task of intelligence subsuming into an Agent life is concerned with the real-time running of MAS. The potential of this application is mentioned in the next chapter.

Intelligent Selection of Process Realization

Each Agent must try to realize its tasks and to solve the upcoming situations in order to meet its design objectives during its life. From this point of view, the standard Agent is consisted in definite and constricted description of its behavior already defined during the modeling phase of the Agent. Therefore, there is no way to change the behavior during the agent running. It is able to do this, in the case of *Intelligent Agent.* Such type of agents is based on the main life process model that specifies just the framework of its behavior. The Agent can dynamically change the pieces of its own behavior according to the situations. This principle is denoted as *behavior reconfiguration approach.* The process of reconfiguration is founded on the replacement of a given part of the whole agent's process by another one that is the most suitable for current situation and conditions. A set of possible applicable behaviors, called realizations (defined by Agent Behavior Diagram), of each "reconfiguration" point (generally the process node of activity diagrams) is defined for this purpose. These definitions of the behavior can be distributed on the whole MAS as *definitions of process realizations*. They can be saved in some global repository or within the particular agents; actually they can be deliberated in the real-time too.

The important and expected situation will appear whenever one process corresponds to the two or more realizations. The logic tools are responsible for solving this situation. But before this state, it is necessary to find the realizations which are applicable to the reconfiguration of a given process. There are two ways to find them – selection based on the process or on the logic grounds.

Pursuant to these ideas, it is able to define such procedure as an algorithm of reconfiguration process.

1. finding a set of suitable realizations for a given process
2. selection of one, the most suitable realization from a defined set
3. realization executing

The first part of this procedure can be solved by two approaches which are mentioned below. A set of process realizations is required for the next step of such algorithm. This second part is concerned with the selection of just one process realization and this selection could be realized by some logic tools. This selection of the most suitable realization is based on its input objects and also on the objects that are occurring within the MAS in the moment of reconfiguration process starting.

Process vs. Logic Approach for Suitable Selection
The implementation of the first algorithm event is a complex problem. It can be implemented by a process or logic approach (fig. 4), where each of them brings different advantages and disadvantages.

The *Process Approach* (figure 4 A) is founded on the idea of assignments of the realizations to the parent process already in the MAS modeling phase. The creation of a needful set is only about surveying of the realizations that are connected to a given process. In this algorithm, it is able to subsume the realizations defined in the local repositories of particular agents as well as in the global repository of whole MAS, it is also possible to apply some restrictions to the scope of validity of separate processes, it means that the process, let us say the realization, can be defined as private, public and so on. This well-founded set of process realizations will be used in the next step of the reconfiguration algorithm – selection of the most suitable realization.

The main advantage of this approach is in the clear selection of realizations, which are usable for the execution of a given process, from which can be chosen just the most suitable one. On the other hand, this approach allows only choosing of the most suitable realization from a set of realizations that is formed in the system modeling phase already. The creation of such set is affected by the skills and knowledge of the model author – the Human Factor acts an important role in this process. This process approach is a less flexible at the expense of almost zero error rate during the creation of a set of realizations, in comparison with a logic approach. In other words, the more effective realization leading to the objectives of a given process which is not connected with such process during the modeling phase may exists. Nevertheless, it is impossible to choose the realization that indeed leads to the process objectives but the execution of this realization is mismatched, unusable or illegal.

In the case of the *Logic Approach* (figure 4 B), the first step of reconfiguration algorithm is changed. The finding of a set of suitable realizations arises from the precondition that no connection between the process and realizations is there and each process and realization has defined its output objects and its objectives in accurate and clear form. The challenge of logic tool is to find (local or global repository can be taken into account with a respect to some restrictions or rules) a set of realizations for the second step of the reconfiguration algorithm. The description, properties, objects and objectives of these realizations have to be in accordance with objectives and objects of a given process. This finding process is automatically executed within the Agent in the moment of necessity during the life of this Agent.

The logic approach when compared with process one is more flexible thanks to the fact that all relations between process and realizations are constructed automatically in the real-time in the moment of request. Unfortunately, there is relatively hi-risk of mismatched, unusable or illegal realizations selection. This problem occurs when the descriptions of processes or realizations are incorrect or incomplete. These descriptions of objects, objectives or properties are made by a human, thus the quality of descriptions and efficiency of selection process are highly affected by the author's skills, knowledge and fantasy. In this case, the Human Factor is much more crucial to success than in the process approach. Generally, thanks to this approach it is able to find some secret and effective realizations leading to

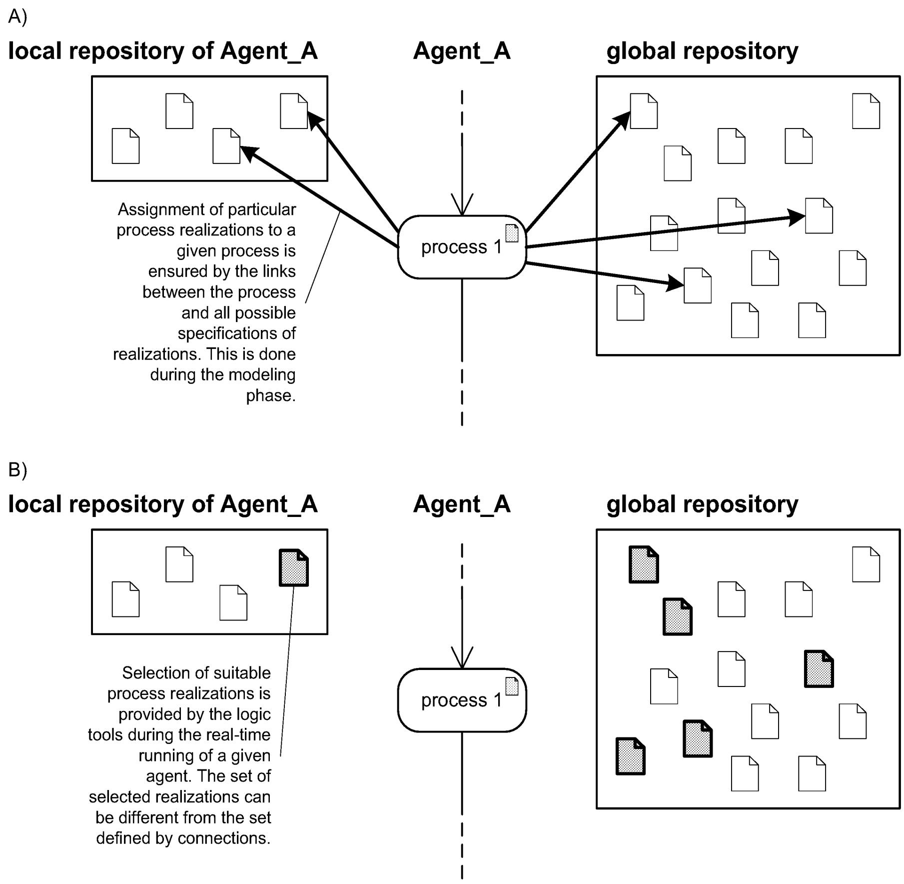


Figure 4: Illustration of two approaches for suitable selection of process realizations

the objectives of a given process. Due to the high demands on the quality and complexity of elements descriptions, the realizations that fit into the descriptions and that will be associated as members of searched set do not need to be usable for the execution of a given process. So, the descriptions of elements in the MAS have to agree with the real world that is expressed by such MAS – and this is a big challenge.

Application Domain

The method that is mentioned in this paper, and that is based on the internal Agent's behaviors, is a headstone for semi-automatic development of Multi-Agent Systems. Thanks to this technique and some developed tool, it will be possible to specify particular components of the system, their properties, skills, features as well as their behaviors. This specification covers the initial phases of software process. The graphical language (extended UML Activity Diagrams) with some formalization and also with some additional textual information (internal activities specifications, objects specifications, etc.) are used for this purpose. This graphical solution could be a good way to move the development process of Multi-Agent System closer to the customers and analysts, it means that the emphasis on the initial phases of software process is there. The next phases of development process should be simplify and automatized, in accordance to the quality, correctness and sophistication of the model that is specified by this technique. Of course, it will be still necessary to implement some parts of agents and system, but the amount and severity of these parts should be minimalized.

The other direction of progress and usage of this technique is focused on the real-time dynamic of Agents' behaviors during the working of a given MAS. Each process has a set of realizations which describe the possible algorithms of the process execution. Because, the number of realizations that belong to a given process can be greater than one, it is necessary to choose just one realization in the time of the process firing – it is a reconfiguration approach mentioned afore. The selection of the applied realization is based on many factors (input and output objects of process and its realizations, costs of the realization execution, etc.) and it is executed just in the moment of necessity during the Agent's life. These Agents, that can change their behaviors in according to the instantaneous situation

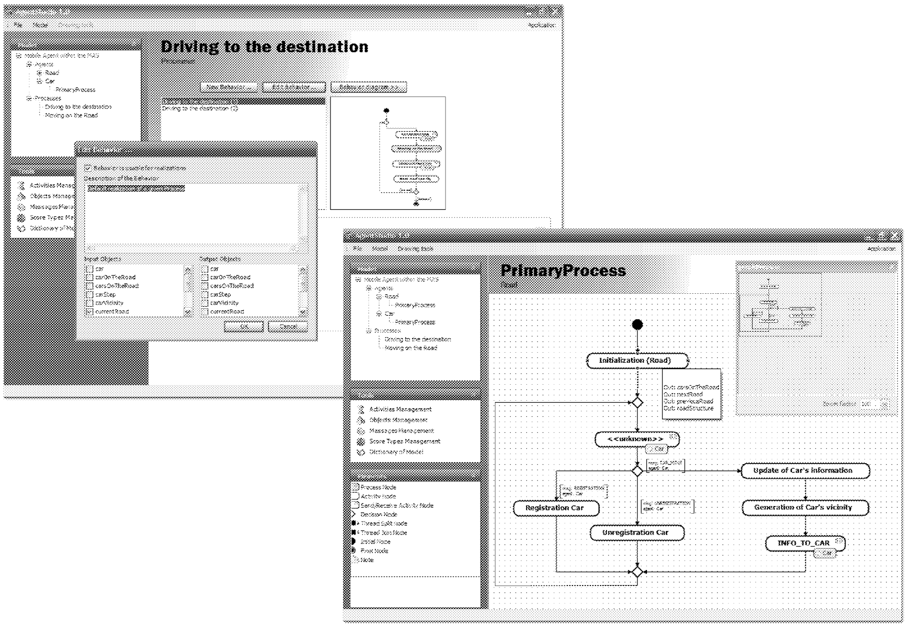


Figure 5: Some screenshots of the AgentStudio software

and conditions, are formed on this reconfiguration idea, they are called "Intelligent Agents" in the sence of this research. It extends the application domain by an ability to build the information system with the Agents which are able to solve the problems separately (based on the available realizations of processes which form the behavior of a given Agent) and in the moment when they occur.

Each software application whose internal structure corresponds to the multi-agent technology is a subject of the application domain of the mentioned approach or research. The application domain is focused on the whole MAS software process and brings the automatization into the design or implementation phases of it. The software applications based on this approach can be also "intelligent", thanks to the usage of Intelligent Agents and the reconfiguration of their internal behaviors. The knowledge redistribution, its sharing and learning of the Agents are possible to realize within these systems too.

Future Work

Although the research of the mentioned technique is in the initial phases, the basic ideas and some concrete conclusions and results concerned with the problematic of Intelligent Agents and the behaviors modeling are done and published at present. The scientific and detailed specifications of all parts of MAS modeling area mentioned in this paper are necessary to do now. It is necessary to define subtle rules, properties and facilities of all elements as well as the relationships between them and whole software process. For instance, many questions concerned with this research are still open (how to specify input and output objects, if the specification of atomic activity should be defined globally or not, which logic tool is suitable for the purposes of this problem domain, which approaches should be use, how to move the model to other phases of software process, etc.) The development of software application that will provide methodology and application framework for the modeling, controlling and operating of MAS system that is based on the mentioned approaches, as well as its components, is an important task. The work on this software application is now in progress. The figure 5 gives a preliminary preview of this "**Agent**Studio" application.

Finally, the goal of this research is to specify complex techniques for modeling of the agents within the MAS based on process modeling. Thanks to this future work, it will be also possible to verify and simulate agents as well as map the descriptions of agents onto the source codes and onto the real implementations of them. The verification, simulation and other models manipulation can be realized by existing algorithms and tools, thanks to the planned support of the process description based on the standard techniques (Petri Nets, BPML, BPEL, XMI, etc.)(Aalst 1997; Matjaz 2006) Above mentioned approaches and techniques also offer the possibilities of distribution of knowledge and learning of the agents.

Conclusion

This paper is engaged with the MAS technology that combines many computer, natural and social science areas and poses specific claims on the standard software process. The future and power of this technology are, among others, dependent on the facilities and quality of the tools for development process of such systems. These tools must be designed with a respect to the skills of normal users that will be a “modelers” of the MAS and that will determine the objectives, requirements and behavior of whole MAS from the real-world point of view. The fundamental ideas and methods to model and develop MAS, based on the internal agent behavior and intelligence of particular agents are described in this paper. The mentioned problems and solutions are the results of preliminary research which will be elaborated with respect to the future work. The modeling and intelligence approaches, graphical nodes, some other solutions, etc. were discussed along with a preview of future work. The application domain, future possibilities and advantages of MAS development based on these ideas was described too.

This research has been supported by the program "Information Society" of the Czech Academy of Sciences, project No. 1ET101940420 "Logic and Artificial Intelligence for multi-agent systems"

References

Aalst, W.M.P. van der. 1997. “Verification of Workflow Nets”. In *Application and Theory of Petri Nets 1997*, P. Azema and G. Balbo (Eds.), Springer-Verlag, Berlin, 407-426

Kubik, A. 2004. *“Intelligent Agents”*, Computer Press, Prague, (publication in czech language).

Radecky, M. and Vondrak, I. 2005. “Formalization of Business Process Modeling”. *ISIM 2005*, Hradec nad Moravici.

Radecky, M. and Vondrak, I. 2005. “Modeling of Agents Behavior within MAS”. *Information systems in practice 2005*, MARQ, Ostrava, (paper in Czech).

Radecky, M. and Vondrak, I. 2005. “Modeling of Processes”. *WOFEX 2005*, Ostrava.

Vondrak, I. 2004 *“Methods of Business Modeling”*, VSB-TUO, Ostrava, (lecture notes in Czech).

Matjaz, B. J. 2006 *“Business Process Execution Language for Web Services BPEL and BPEL4WS 2nd Edition ”*, Packt Publishing, Birmingham.

Pilone, D. and Pitman N. 2004. *“UML 2.0 in a Nutshell”*, O'Reilly Media, Sebastopol.

SPATIAL EXTENSION IN BUSINESS PROCESS ENACTMENT

Dušan Fedorčák
Jan Kožusznik
Ivo Vondrák
Department of Computer Science,
VSB-Technical University of Ostrava
17. listopadu 15, 708 33 Ostrava-Poruba
Czech Republic
E-mail: dusan.fedorcak@vsb.cz
jan.kozusznik@vsb.cz
ivo.vondrak@vsb.cz

KEYWORDS

BPM, Petri Nets, CPN, Business Process, Spatial View, Visualization, 3D

ABSTRACT

Approach to workflow should be used to coordinate external autonomous agents as they perform various tasks; a task formalism incorporating graphical (but mathematically rigorous) process models is being used. BPM (based on Colored Petri Nets) provides the meth-odology for business process modeling, simulation and enactment. Problem is how to describe where are agents expected before and after execution of specific task in virtual space. This problem is solved by our approach of spatial view based on CPN spatial extension.

INTRODUCTION

The purpose of this research is to develop extension of BPM that adds spatial perspective to the process enactment. We will define spatial preconditions for execution of activities. These preconditions are represented by areas where the tokens have to be before activity execution.

MODELING METHODOLOGY, THE BPM

The BPM, methodology for modeling autonomous agent behaviors and tasks, uses the Colored Petri Nets (CPN) as the underlying Task Formalism (TF). CPNs are implemented as first class objects and are directly executable using their class methods. This CPN-based methodology was defined during the late 1990s [1]. This methodology is based on three views describing different aspect of model - **functional, object** and **coordination.**

The **Functional View** is used to compose process architectures and to enumerate process owners, users, and products. Two types of relationships between processes are available - containment and collaboration. Containment identifies subprocesses while collaboration indicates concurrent operation of two or more processes If possible.

The **Object View** characterizes all objects that are essential for processes both active and passive objects as well as their attributes and relationships. Active objects provide services that may be needed during process activity performance. Passive objects represent things such as materials, products or documents which are consumed, produced or manipulated by process activities. All these objects have their own attributes which are associated with them.

The **Coordination view** incorporates object type definitions from the object view and hierarchical process relationships from the functional view into sets of state transition graphs that can be executed without additional interpretation or argumentation. Coordination diagrams show the consumption and production of tokens (occupying active and/or passive object places) by transition (activities or behaviors) in accordance to rules governing CPN execution.

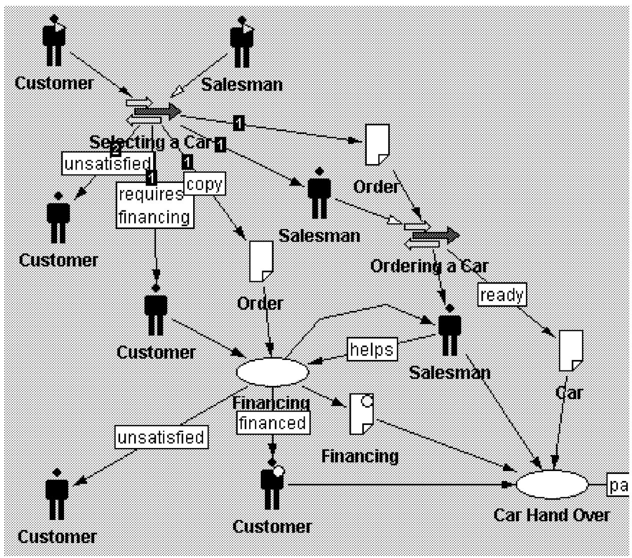


Figure 1: Coordination View Represented by CPN

Coordination diagrams allow modelers to specify an execution order for contained activities and processes, including preconditions and post-conditions that affect concurrency in operations. The expected execution order is defined, as well as sharing of used resources. Each activity can have multiple execution "cases". Case choices determine an execution path through process models representing composite behaviors, tactical tasks and missions. They also allow enable execution to adapt/respond to exogenous events, yet enforce the coordination needed to successfully complete tasks or missions.

SPATIAL VIEW

Commercial usage of Business Process Studio (lead BPM implementation) shows one minor disadvantage of the BPM methodology. The methodology doesn't provide suitable information about spatial distribution of the process. It is possible to include some special activities to describe e.g. transporting resources, however the model quickly become too complex and unusable.

Possible solution of this issue is to create a new view into the process, the Spatial View. This view should provide missing information through a new diagram and also cooperate with other views to give user more accurate sense of the process. To develop well-formed Spatial View, underlying task formalism (CPN) has to be extended by two new entities shown below:

CPN Extension

Position
Spatial View shows the process layout in real word e.g. on a map. It's obvious that we have to use some kind of positioning. It's effective to set position for many parts of the process model elements e.g. tokens.

Location
Locations are defined as polygonal areas with assigned position. These areas can be organized in hierarchical structure and are used as intermediates between CPN and the map itself.

Formal Definitions

In more formal language, there are set of locations L, set of tokens C participating on the process (CPN) and set of coordinates X. We define new mapping:

$$Pos: C \cup L \rightarrow X$$

Our spatial extension of BPM is based on assigning places to locations. This approach corresponds with the model behavior. (Tokens are "real" objects which can be positioned and places can be understood as "waypoints"). Place to location mapping M_L is defined as follows:

$$M_L: P \rightarrow L$$

With position mapping *Pos* and place to location mapping M_L we are able to define specific classes of CPN elements:

Location-specific place
A *place p* is location-specific if it is assigned to *location*.

$$\exists l \in L \ : \langle p, l \rangle \in M_L$$

Location-specific activity
An *activity t* is location-specific if exists at least one location-specific input or output *place p* for that *activity*.

$$\exists p \in ({}^{\bullet}t \cup t^{\bullet}) \, , \exists l \in L \ : \langle p, l \rangle \in M_L$$

Hard location-specific activity
An *activity t* is hard location-specific if exists at least one location-specific input *place p* for that activity.

$$\exists p \in {}^{\bullet}t \, , \exists l \in L \ : \langle p, l \rangle \in M_L$$

The approach was also found very useful for describing preconditions of activity execution (activity execution depends on its resources, or in CPN language, firing transition depends on its enablement). We can define this behavior as follows:

Hard location-specific activity enablement
Hard location-specific *activity t* can be fired only when needed *tokens c* occupy all input places for that *activity.* In the Spatial View, occupy means that the token is "inside" *location* to which was input *place* assigned.

$$M: P \rightarrow C^{\infty}$$

$$\forall p \in {}^{\bullet}t, \ \forall c \in M(p) : Pos(c) = Pos(M_L(p))$$

EXAMPLE PROCESS

Spatial Extension can by explain on simple produce & transport process. A worker produce a product from shipped resources and the produced product is transported. Fig 2 shows the process elements e.g. activities (produce, transport, return) and active and passive objects participating on the process (resource, worker, truck, product).

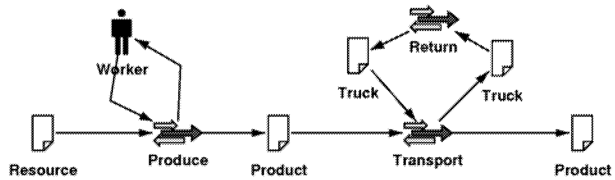


Figure 2: Example Produce & Transport Process

Fig 3 shows the basics of the Spatial Extension, place to location mapping (red arrows). The mapping adds new information to the process. The product is produced in factory and then transported to store. This new information is desired output in our research and has to be visualized properly.

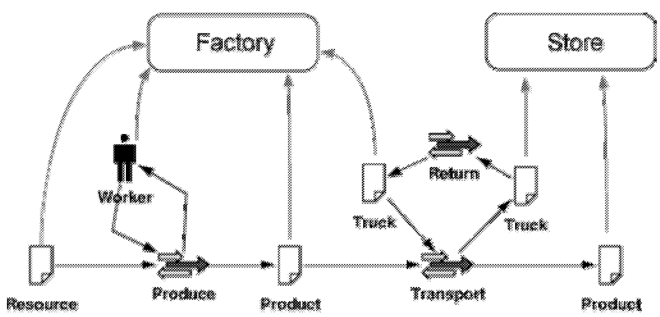


Figure 3: Example Process, Place to Location Mapping

SPATIAL VIEW VISUALIZATION

Visualization of presented spatial extension is essential. The problem is, how to fit so much information into the map and do not lose lucidity. At present we have two prototypes of visualization tool, classic 2D and experimental 3D approach.

Classic 2D Visualization

Spatial View (fig 4) shows the map, locations (blue polygons) and passive or active objects (CPN tokens, yellow flash icon). Spatial View also provide information about place to location mapping (a tool tip for every location with list of mapped places).

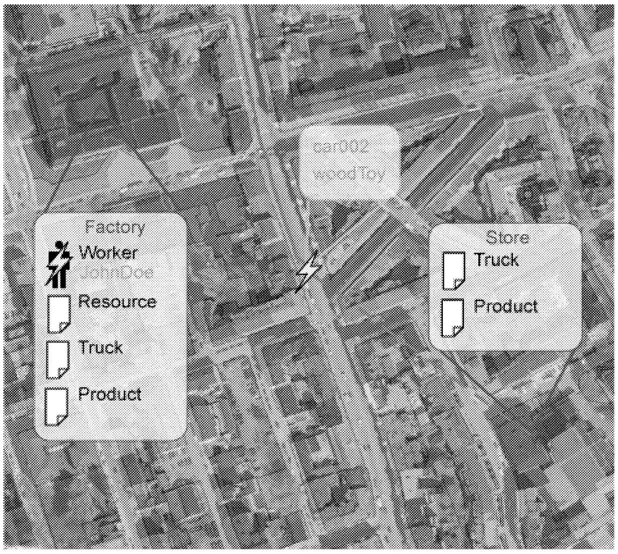


Figure 4: 2D Spatial View, Process Enactment

Fig 5 shows Coordination View of the same process in the same phase (process enactment). Tokens (car002, woodenToy) are transported within transporting activity from factory to store. Token tracing is also provided (red lines).

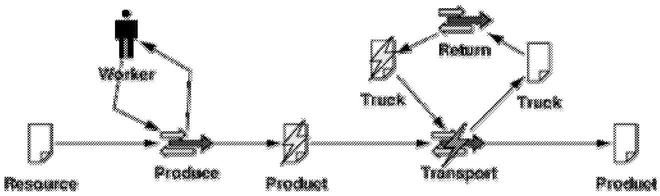


Figure 5: Coordination View, Process Enactment

Experimental 3D Visualization

However, 2D visualization as presented has one major issue. When the process becomes more complex, place names get duplicated and Spatial View (especially tool tips) becomes unusable. We suggest that 3D visualization approach can fix this issue.

3D approach is based on joining Coordination View and 2D Spatial View (fig 6). This hybrid diagram can shows both diagrams in one view. Also there is "lot of space" to visualize location mapping, token position, token tracing etc.

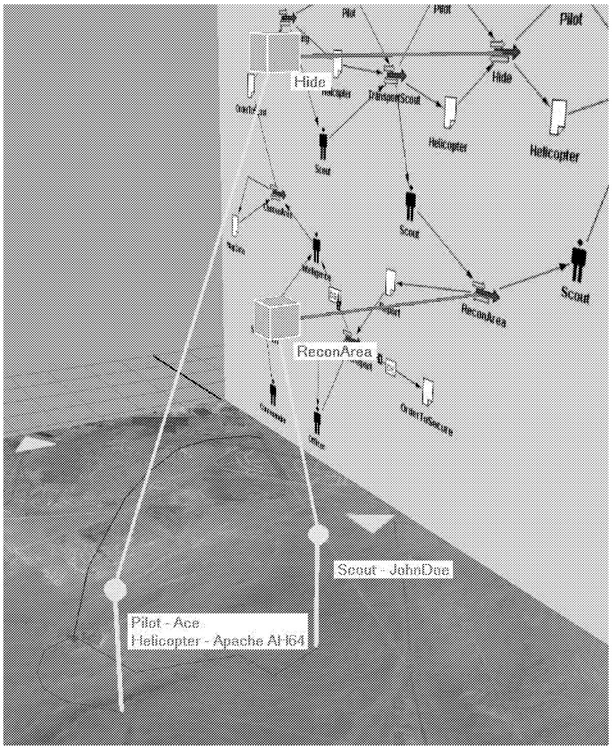


Figure 6: 3D Spatial View, Process enactment

CONCLUSIONS

In this text, we have presented spatial extension of BPM methodology. We have defined new preconditions for transition enablement and demonstrated their purpose according with the model behavior. We have also presented visualization of this extension. Further studies have to be concentrated to unique 3D visualization tool to extend its possibilities.

This research has been supported by the program "Information Society" of the Czech Academy of Sciences, project No. 1ET101940420 "Logic and Artificial Inteligence for multi-agent systems"

REFERENCES

Vondrak I., Szturc R., Kruzel M. 1999. "Company Driven by Process Models" European Con-current Engineering Conference ECEC 1999 (Erlanger-Nuremberg, Germany), SCS, Ghent, Belgium, pp. 188-193.

Vondrák, I. 1997. " Methods for Business Process Modeling" Research Project, UNDP

COLLABORATIVE CE-ENVIRONMENTS FOR VIRTUAL TEAMS

WORKFLOW AUTOMATION AND OPTIMIZATION IN AUTOMOTIVE DESIGN

Guy K. Kloss
Simulation and Software Technology – German Aerospace Center (DLR)
51170 Cologne, Germany
E-Mail: Guy.Kloss@dlr.de

KEYWORDS

Problem Solving Environments, Optimization, Simulation Integration Framework, Distributed Environments, Grid Computing, Data Management.

ABSTRACT

In the automotive design process individual complex computer aided engineering (CAE) applications are needed to interact in a coordinated way. The engineer goes through many iterations for various model modifications in order to obtain improving qualities of the design. This process often involves time consuming manual interactions with the involved applications. This paper presents a solution to the problem of task automation. The engineering problem described highlights several aspects of process automation and collaboration commonly found in CAE environments.

Conducting this optimization task should be managed for the engineer as clearly and easily as possible. As an example, a complex iterative optimization of a car's body part has been implemented. Handling of data and execution of applications have been automated. An additional abstraction of the execution process gives the freedom to migrate the computation to different hardware platforms and computational resources.

INTRODUCTION

Optimizational computations in automotive design processes require a coordinated interaction of possibly complex single applications. In this chain of processes the applications interchange large and complex sets of files.

In this design phase the parts worked on are still within the virtual world of model simulations. The engineer goes through many iterations for various model modifications in order to obtain improving qualities of the design. This implies to cycle through a repeated loop of performing similar actions within the various CAE applications, including the exchange and conversion of large volume data sets.

The engineers engaged in this process go through the chain of process steps to advance from fractions of design models, through compound models including material properties, to the simulational analysis of those models. This is a very time consuming process. Some engineers automate certain steps using (shell) scripts. This will proabably increase productivity, but the approach has definite restrictions in everyday's use due to difficulties in maintenance. The users of these CAE applications are engineers, not software developers.

The following sections outline the chain of operations in more detail, and then procede through an integration of an optimizer to increasing the efficiency of the automated process.

Process Automation

The method of creating variant designs of parts within the virtual design phase lets the engineer evaluate the approaches without the costly real-life manufacturing of prototypes. This virtual design and evaluation process usually involves the following:

1. Design/modification of models
2. Conversion of the models for further applications
3. Merging the modified model with unchanging parts and adding information (material properties)
4. Analysis of the obtained virtual design
5. Possibly visualizing the results and/or extracting comparable values describing the design "quality"

Depending on the desired process, some of these basic steps may occur multiple times at various steps in the process chain.

An automation of this process has two major benefits: The most obvious it saving time. The time for conducting manual processing can be used to give the engineer the chance to concentrate on creating more competing designs in a given time, and thus increasing the engineering efficiency. The second benefit is reducing (unnoticed) error prone interactions in the process, induced by manual management.

Optimization

The approach of automating a process chain can undergo further improvement by introducing a looped optimizing workflow. An optimizer within the workflow can generate sets of parameters defining design variations (e. g. thicknesses of sheet metal, dimensions of structural beams, layout of reinforcing design elements). These parameters are translated into the variable components of the design model (step 1), and automatically evaluated through the process described in the previous section. The process chain acts as a solver to the optimizer in the workflow.

The quality of a particular design generated in an optimization loop is determined by extracting some key values from the results obtained in the process chain (see step 5). These values will be fed back into the optimizer for the (multi disciplinary) optimization. The optimizer generates a new set of parameters for the next iteration until reaching a defined stop criterion.

Paralellization

In multi disciplinary optimization environments very often different analyses are carried out. These can be

- either using different computational components (e. g. weight, structural strength, eigen frequencies),
- or using one analysis method parameterized differently for multiple runs (e. g. different load cases, differently conducted crash tests).

These tasks are usually independent from each other and set up as individual computational jobs. So they can be distributed easily to different computational hosts for execution. Compared to a sequential processing, approximately the same resource in CPU time is consumed, but the execution time to the engineer may be multiplied by a factor up to the degree of paralellization. Complex optimizations with many loops for example then can be conducted overnight, whereas sequential processing would yield a multi day job.

INTEGRATION SYSTEM

Traditionally coupled simulations composed of individual and independent applications, as found in optimization processes, are integrated by (shell) scripts. They copy data using file I/O on shared file systems (NFS), applications are started – and maybe controlled – using file system "flags" (creating and removing files to trigger actions within the application). More sophisticated coupled integrations usually require the modification of the source code by the original developers to create a higher, and integrated, single application.

"Integration technology" is a term describing developments of the last few years for flexibly controlling complex workflows. The approach is to keep the need for modifying the simulation's source code minimal (at the most). The main aspect of the approach is an object oriented design concept, using component technologies and the communication and interaction technology CORBA/IDL. Problem solving environments integrate applications, and manage the data in an automated workflow. The goal is to prevent distractions from the engineering job. So the simulation and optimization tasks need to be presented in a simple and clear way, and the distribution of computations should be handled transparently. This also implies the distribution of data between the various components within a workflow.

The problem solving environment developed by the German Aerospace Center is a system called TENT. TENT is an engineering framework to set up an environment for distributed application integration and workflow management, including the necessary features for data handling and management. It has been implemented to provide a very high level of flexibility of use, with the least degree of restrictions for the target applications. To ensure platform independence the target platform only requires a Java Development Kit (JDK).

Application Integration

Integration environments need to wrap the individual computational code with software to interact with the application in a suitable way. By doing this the integration system interacts with each application wrapper in an identical manner, whereas the wrapper's purpose is the specialized interfacing of the application itself.

This may involve various approaches for interfacing. The CAE tool may provide an application programming interface (API) the integration environment's wrapper can connect to in a suitable programming language, a remote control interface (e. g. through a socket connection, Web Services based), or – unfortunately very common – the simple communication through the file system by creating or deleting file system entries is used. If none of these methods is available, the application will need to run as a "black box application", that will just be started without the possibility of controlling it (besides terminating the complete process).

The wrapper, together with the application, then yields a component for the integration environment [Forkert et al. 2004]. This component can interact with the system's governing and controlling components on the one hand, and with other computational components on the other hand.

In TENT the application integration is more and more performed using Python scripting at run time [Fischer et al. 2000]. This relieves the developer of a time consuming edit–compile–deploy–run cycle. The implementation is performed in a language that is highly structured, readable, and powerful to the system designer. Possibly slightly slower execution times of embedded Python code in comparison to the application's native code or the system's Java code are usually neglectable. Only high level managing code with a minimum fraction of the total execution time for the binding logic (the "glue code") is handled with it.

As mentioned above, code integration can be driven to various levels. For example customized control panels for the graphical user interface (GUI) can be developed. This could be valuable for the need of user controllability for certain workflow components, depending on the degree of possible interactions (between the wrapper/TENT and the integrated application). These panels will be embedded into the GUI at run time for the specific configuration and control purposes.

Distributed Computing

The performance for obtaining solutions can be increased through paralellization as mentioned previously.

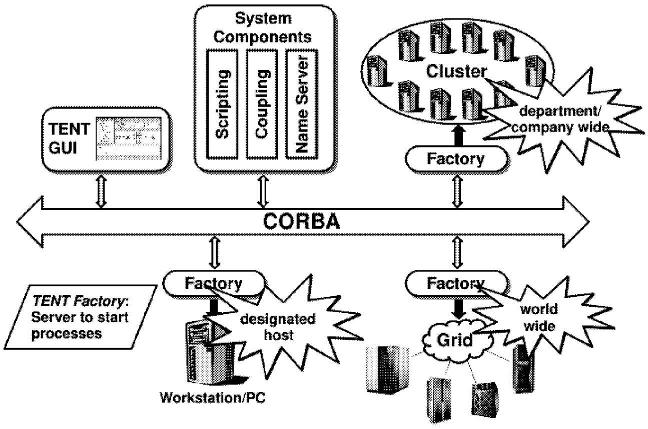


Figure 1: Component distribution.

This can be achieved in two ways: by distributing a single task of a workflow component onto several computational nodes, or by distributing several independent tasks onto different computational hosts. The first approach is well outside the scope of integration systems, it has to be managed by the application itself. The latter can only be conducted from within the integration environment.

The efficiency gain of this type of parallelization is highly dependent on the freedom of migrating the single computational job to a different host. If the simulation integration environment itself is independent of the execution location, several factors determine the "mobility" of the process:

- Availability of the application on different hosts
- Application's platform and system dependencies
- License restrictions
  (for hosts and number of processes/CPUs)

Depending on these three factors, the total execution time may be reduced. Some common means of distribution of data processing are execution on a

- remote host (e. g. execution using SSH/RSH or on super computers),
- scheduling system (e. g. LSF, PBS),
- or computational Grid (e. g. Globus start).

Relocation of computations to a host fittest for the individual task, and identification of parallelization potential, are the key factors for increasing workflow efficiency. The capability to migrate either the application's execution, or certain system components, is given by a transparent abstraction of the components from each other (see figure 1). This way the interacting components can be distributed individually. Localizing of one or more resources together of course would be a special case already covered by the design. TENT is capable of utilizing all three means of computational distribution mentioned in the list above.

Data Management

Simple workflow management by (shell) scripting (and many other integration environments) assume a common shared file system (e. g. NFS). Information on file system accessibility is not available prior to execution time. So workflow components executing individual applications on independent hosts flexibly, result in a demand for data management. At simulation time this data management's purpose is to (A) provide means of data transfer between individual components, and (B) facilitate the access to data storage instances accessible from within the distributed environment. Depending on the disciplined behavior of engineers, dealing with (B) in collaborative environments may lead to data chaos sooner or later. So another problem to solve is (C) the collaborative and orderly management of data on a larger scale.

Workflow Communication

Distributed components communicate by sending and receiving CORBA events between them (task A above). A set of events for simulation workflow management have been implemented in TENT. Some are usable for data transfer by various means: Simple referencing of commonly accessible files (e. g. on a file system, or server), or using direct transfers (e. g. Globus GASS transfer, GridFTP). The TENT framework's software development kit (SDK) includes means for using various methods through the API in the application wrappers.

Data Storage Server

For data storage (task B above) a data server is deployed for use within the problem solving environment. For various reasons we use a WebDAV [WebDAV 1999] server. It implements a data server concept building on HTTP protocol extensions for communication. The advantage of this approach is the ease of access even in distributed and protected environments in comparison to many other remote file system access methods. The TENT GUI and all workflow components can access this data server to share (store and retrieve) files and other data.

Collaborative Data Management

In collaborative environments with many simulations and optimizations conducted, the need for controlled data management arises quickly. Simulations or workflows need to be stored in an orderly way, they need to be annotated by meta information for closer identification on the server by the team, and the corresponding data for the simulation (input data and computed results) needs to be related to it.

In the implementation this management has been split up. A data management system, based on open and future proof standards (XML and WebDAV), was employed. An infrastructure for technical and scientific data management purposes was used (DataFinder [Kloss 2004]). With the DataFinder repositories for collaborative data management can be set up. Within a repository structured data objects are defined. The structure between the data objects is defined by relations (data hierarchy, to prevent uncontrolled placement of data objects) and annotated by custom defined attributes (meta data). The system provides means for access control (access control lists), server side searching and locating of content, and versioning of data items within the system.

This system, including the DataFinder as a graphical client, has been designed independently, to be flexible for use in many different scenarios. However, while implementing problem solving environments like TENT have been kept in mind for coherent interoperability. TENT itself is able to integrate seamlessly with the DataFinder data management scheme.

PROBLEM SPECIFIC IMPLEMENTATION

Within the scope of the AUTO-OPT [AUTO-OPT] project some extensions to the TENT environment have been implemented. These include a complete rework of the Python scripting API for easy wrapper implementation [Fischer et al. 2000], abstraction of process distribution for scheduling systems and Globus Grids [Globus], and data management system integration.

The system's fitness for automotive tasks has been demonstrated with a complex distributed sample optimization process. The process includes

- a multi-disciplinary optimizer,
- a parametric shape modeller,
- a model merger,
- a structural mechanics solver (FE),
- a crash test analysis tools,
- and several utility components (converters, value extraction, workflow logic).

In this scenario (figure 2) some applications were bound to specific hosts or platforms for execution, some were executed on scheduling systems LSF, and some were executed using a Globus Grid [Schreiber et al. 2005].

In the project the demonstration process to be conducted was the optimization of a main chassis beam with variable material thicknesses and several geometry parameters. SFE-CONCEPT [Heiserer 2004] took these to generate a fully meshed model, while avoiding defined engineering constraints. The meshed model then was converted and merged with non-changing parts of the automotive frame using Medina [Westhäußer 2003]. This compound model was then subject to analysis by the rest of the workflow.

To the optimizer DesParO [Bäuerle et al. 2004] in the workflow, the solver is a sub-workflow in a loop (in figure 2 enclosed by the dashed line). Both are attached to the coupling control component. The loop is forked for two independent analysis paths after the compound model creation. This created the first possibility for parallelization. In one branch a number of differently parameterized crash tests (PAM-CRASH and

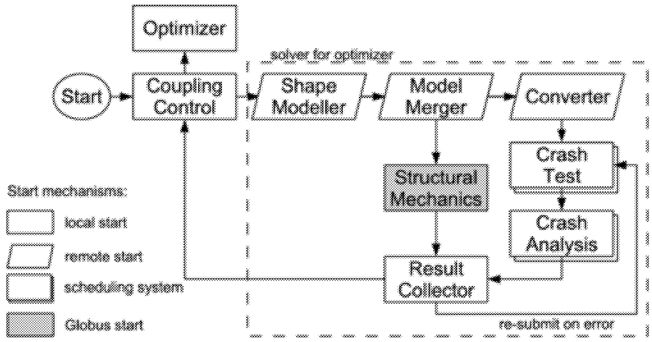


Figure 2: Implemented compound workflow application.

PAM-VIEW) was conducted. These tests were submitted to the scheduling system, and the results were collected again on individual completion (task farming). On these, the number of available floating licenses for PAM-CRASH had to be considered as a restriction for task farming, to prevent a decrease in efficiency. In the other branch PERMAS [Fischer et al. 2004] conducted a structural mechanics and eigen frequency analysis. After performing all numerical steps, the resulting files were analyzed to extract scalar values. These are collected and forwarded back to the optimizer as feedback to the quality of the generated parameter set.

The challenge for the optimizer was to determin an optimal set of parameters for competing goals in the objective function. Structural strength and high eigen frequencies are opposing weight of the design and energy absorbing deformation in the crash.

When complex and very long lasting optimizing simulations like these are set up to run automatically without the engineer's direct interaction, an unwanted behavior of the process may occur. Some unpredicted circumstances – as unavailibility of application's licenses, diverging simulation results, job rejections of the scheduling system, etc. – may require the engineer to intervene with the fully automated workflow processing. To accomplish this, the engineer needs feedback from the integration system on the current progress within the workflow, the state of applications, simulation code internal numeric states, etc. For this purpose the state of each application within a component is monitored, output is analyzed to determin erroneous abortions, output from files and on the standard streams (standard out, standard error) is parsed, etc.

This information is made available for the user's feedback in various ways:

- logging messages (GUI panel with system state information, warnings, errors)
- monitoring plots (numerical values in internal or external graph panels)
- results (current multi dimensional result sets passed to external visualization tool)

Through the user interface the engineer may then cause the workflow to pause execution at certain points to save unnecessary processing time and therefore costs.

COMPARISON

The market for problem solving environments features a wide variety of competitors with particular strengths in different areas. These commercially available systems usually provide a high level of maturity regarding their implementations for the designated target market.

EngineFrame for example covers the utilization of Grid Computing more extensively. CAE-Bench's domain is a seamless data management tracking all manipulations, data forking, etc. Other systems are focusing on other aspects while disregarding or neglecting some other facets. Commonly found in this field are optimizers featuring alternative algorithms. Systems like these are modeFRONTIER, iSIGHT, Pacelab, Pointer Pro, ModelCenter, and others.

CONCLUSION

With the combination of the integration environment TENT with the data management system DataFinder one gets a high flexibility to integrate, distribute, and manage complex simulation and optimization scenarios. Other environments provide very mature strenghts in other fields and often ease of use – at the cost of flexibility and extensibility.

The bandwitdh of strengths in detail are

- completely open interfaces,
- flexible extensibility of the user interface,
- feasibility to integrate arbitrary applications,
- conduction of interactive simulations,
- utilization of distributed environments and Grids,
- and an integrated data management based on open standards.

Further enhancements of the implementation are easily thinkable. It would be fairly easy to provide an interactive process control to the distributed optimization loop. A user could pause a workflow for alterations, check-point, migrate, and re-parametrize it, etc.

The data management features are essential for enabling computations in highly distributed systems. Data transfers directly between the components and through shared data servers provide the key to this goal.

The employed data management system introduces organizational means by annotation information through meta data stored with the content. This gives larger engineering teams the tools to structure, manage, and retrieve all relevant information from simulations. Additionally user management, access control, versioning, and server side location of data completes the toolbox for collaborative team work.

Further development on TENT is taken in the EU funded Grid Provenance project. It will enable simulation tracking beyond pure data storage and logging information. With provenance recording component interactions, causalities, and data linking information can be tracked and preserved for even complex simulation and optimization workflows. This gives means for sealess processe's provenance tracking for liability reasons and larger scale analyses.

ACKNOWLEDGEMENTS

The presented work was part of the AUTO-OPT [AUTO-OPT] project. Developments were in part sponsored by the German Ministry for Education and Research under the grant 01 IR B01 A.

References

[AUTO-OPT] The AUTP-OPT project web site: http://www.auto-opt.de/

[Bäuerle et al. 2004] Bäuerle, C.; Thole, C. A.; Trottenberg, U. 2004. "DesParO – A Design Parameter Optimisation Toolbox using an Iterative Kriging Algorithm." *ERCIM News,* No. 56, 32 – 33.

[Fischer et al. 2000] Fischer, J.; Schreiber, A.; Strietzel, M. 2000. "Script Wrapper for Software Integration Systems." In *Proceedings of the 8th International Conference on High-Performance Computing and Networking* (Amsterdam, NL, May 8 – 10). Springer-Verlag, London, UK, 560 – 563.

[Fischer et al. 2004] Fischer, R.; Kirchgäßner, B.; Müller, J.; et al. 2004. "Effiziente Akustiksimulation mit PERMAS." In *Proceedings of DaimlerChrysler CAE-Konzern-Forum* (Stuttgart, Germany, Nov. 9 – 10).

[Forkert et al. 2004] Forkert, T.; Kloss, G.; Krause, C.; Schreiber, A. 2004. "Techniques for Wrapping Scientific Applications to CORBA Components." In *Proceedings: 9th International Workshop on High-Level Parallel Programming Models and Supportive Environments* (Santa Fe, NM, April 26). IEEE, Piscataway, NJ, 100 – 108.

[Globus] The Globus Alliance web site: http://globus.org/

[Heiserer 2004] Heiserer, D.; Zimmer, H.; Schäfer, M.; et al. 2004. "Multidisziplinäre Optimierung parametrischer Fahrzeugkomponenten: Entwicklungen im Karosseriebau." *VDI Report 1846,* (Würzburg, Germany, Sept. 29 – 30). VDI, Düsseldorf, Germany.

[Kloss 2004] Kloss, G. 2004. "Technical-Scientific Data Management with the DataFinder." Presentation at the *European Software AG Users' Conference* (Berlin, Germany, November 10).

[Schreiber et al. 2005] Schreiber, A.; Metsch, T.; Kersken, H.-P. 2005. "A Problem Solving Environment for Multidisciplinary Coupled Simulations in Computational Grids." *Future Generation Computer Systems,* Volume 21, Issue 6, 942 – 952.

[WebDAV 1999] RFC 2518. 1999. "HTTP Extensions for Distributed Authoring – WEBDAV." Network Working Group. WebDAV web site: http://www.webdav.org/

[Westhäußer 2003] Westhäußer, M. 2003. "Wie kann der Berechnungs-Prozess für Gesamtfahrzeuge verbessert werden?" *Simulation – Fachmagazine für FEM, CFD und MKS,* Issue 2/2003.

A MULTI-USER MODELING ENVIRONMENT TO CREATE AND PARAMETRIZE MATERIAL FLOW SIMULATION MODELS IN D3 FACT INSIGHT

Wilhelm Dangelmaier
Mark Aufenanger
Daniel Huber
Christoph Laroque
Kiran Mahajan

Heinz Nixdorf Institute, University of Paderborn
Fürstenallee 11
33102 Paderborn, Germany
E-Mail: (whd|marka|huber|laro|kiran)@hni.upb.de

KEYWORDS
Simulation, multi-user modeling, d3 FACT insight, client-server architecture, material-flow simulation

ABSTRACT

This paper presents the development results of a two-dimensional multi-user modeling environment for the simulation tool d3 FACT insight. By using this modeling component, simulation experts were able to develop material flow models more effectively, and they can work together on different computers in different places at the same time. This makes model development more efficient as compared to when a single modeling expert has to model the entire problem alone. Hence, an aggregation of sub-models becomes possible. To realize a multi-user modeling tool, many difficulties like conflict or access-right management, have to be handled. In addition, the paper describes the development of the client server architecture in d3 FACT insight to enable distributed development.

MOTIVATION

Future orientated factory and product planning needs a collaboration of all competencies along the product lifecycle at a very high level. On the one hand, the product development times are getting shorter and shorter and on the other hand, there has to be more and more product variants. This is a big challenge for corporations. That is why the planning vision of the digital factory in research is followed with insistence (Marczinski 2005).

By using the digital factory concept, product planner and designer are getting a modeling and simulation tool, which allows them to optimize production places and products or to test planned changes before they take effect in real systems. Furthermore, the developer, planner and carrier dispose the same data pool as the other vigilantes of the company (e.g. engineers or advisors). Thereby changing processes can significant efficiencies or savings be realized faster, because many experts can work on the same assignment simultaneously. There is not only an upcoming transparency for the specialists. In addition, the costs become reduced too, because patency and correctness of data is given and can directly be optimized by simulation. So the simulation is an encouragement in designing systems. Furthermore already existing or conceptual systems could be modeled with the use of simulation.

However, current process-simulation and -modeling tools (e.g. eM-Plant, Taylor ED or Quest) support the team based development and simulation but only in an inadequate way. For this reason, the chair of business computing, esp. CIM at the University of Paderborn has developed a multi-user capable modeling module for the material flow simulation program d3 FACT insight.

This modeling module allows developers to design, parameterize and change simulation models together in teams. Collaborative work includes two essential aspects: space and time. Spatial viewed, collaborative work can be arranged face-to-face or in a distributed group. Temporally focused it can be arranged synchronously or asynchronously. This leads to a contemplation of collaborative work by building a space-time matrix (Johansen 1991). The multi-user modeling tool should support these different coopeation scenarios.

RELATED WORK

Today integrated packages like eM-Plant or Taylor ED are used for material flow simulations (Klingstam and Gullander 1999 or Mueck and Dittmann 2003). With these tools, modeling experts are able to model and execute models of production processes in one simulation environment. The whole process of modeling, execution, analysis and modification takes place on one person's computer. Because these tools do not support teamwork, only one single user is able to work with the model at the same time, in the same session. If more than one user has to edit the same simulation model, it has to be done sequentially (Dangelmaier and Mueck 2003).

An established approach to solve the multi-user support problem, is to operate with several sub-models, where the whole model consists of sub-divisions. Each disjoined sub-model can be designed by another team member. To calculate the whole model, a person has to integrate all sub-models into a large aggregated model. In this approach, an integrating user must recognize all dependencies between the sub-models. This is nearly impossible when a large simulation scene is going to be designed. If changes are required, more than one user cannot modify the integrated model at the same time.

Due to the lack of versioning, it is difficult to reconstruct the changes of other team members during the modeling process. After the integration of an entire model, it is difficult to separate the sub-models again in their buildingblocks, for example, if one sub-model has to be replaced by a newer version. To solve this problem of interaction, the modeling process of a material flow simulation has to be regarded closely.

THE MODELING ENVIRONMENT

The module-based modeling and simulation environment for virtual process planning and control was developed at the chair of business computing, esp. CIM at the University of Paderborn. It is based on a central data management of standardized objects, a controlled access to the necessary data, a modeling component to create and modify simulation models and a computation component to calculate and evaluate models. The computation component will not be addressed in details in this paper.

Models

Models used to simulate a material flow typically consist of blocks representing the modeled process entities and marks that stand for factors of the production (e.g. products, machines or parts). If two or more blocks have a production relationship between them and interact among each other, they link together by linking connections. While the model is simulated, the marks follow these connections at a calculated time.

Many of the models are too big to be managed effectively. For this reason the simulation models are organized in an hierarchically way. Figure 1 illustrates possible different layers and their level of aggregation.

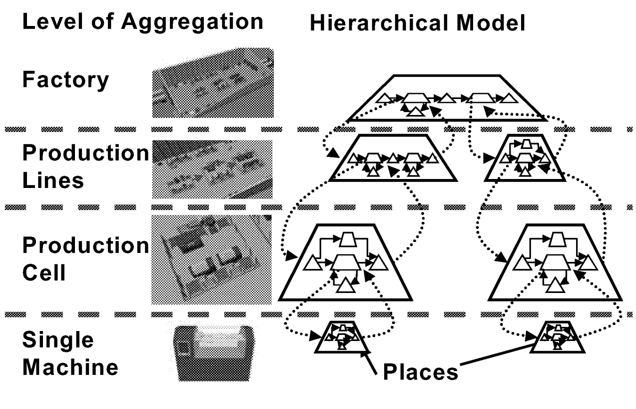


Figure 1: Hierarchical models for material flow simulation (cp. Dangelmaier et. el. 2004)

Several machines are aggregate able to one bigger production cell, represented as one place or buildingblock. Connected buildingblocks become aggregated to lines and lines to production stages. Links only exist within a sub-model. To interact on a more aggregated level, the blocks have special connectors, input- and output-channels. The user can work with the completely aggregated building block at once. For this reason models become manageable and can be edited in an effective way.

System architecture

To handle huge simulation models and the teamwork for many users, a capable environment is needed. To offer an effective and high-capacity system, 3-tier architecture (Bengel 2002) is used. Within this architecture there is a presentation layer on the client side, a persistent data layer, the database, and a transaction- and application-control-server. This three-step model is very comfortable in caching mechanisms, because the many database accesses during the modeling process are very cost intensive in resources (Tanenbaum and Steen 2003). Figure 2 shows the system architecture and the components of the entire modeling environment used for the 3d FACT insight modeling tool (Dangelmaier et. el. 2005).

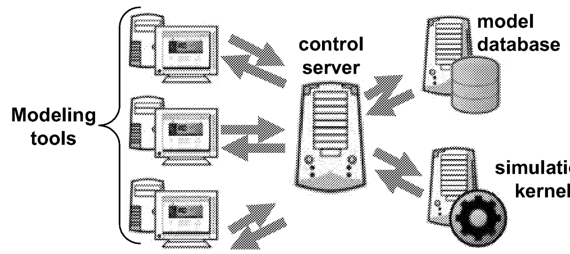


Figure 2: Modeling environment architecture

The communication between the modeling tools and the control server happens on save way, by using the secure socket layer. Therefore, it is guaranteed that no one without the adequate key can disturb or listen to the communication between the tools and the server.

The modeling tool allows designers to load more than one model into the tool and to work on all their loaded models. Thereby different users can load the same simulation model and work autarkic with this model or together in a team. A clearly differentiation of all currently opened models by the control server is essential. Therewith it is warranted that every modeler gets the active data for the models currently opened by himself. In addition, the changes made by a user to an opened model, only have to take effect on this model. It is not allowed to influence the same model, which was autarkic opened by other users. During the common work of a modeling experts team on a model, the changes, made by one team member, have to be transferred to all the other team members.

According to changes, every parameterization has to take effect only on the changed model. Therefore, the control server generates a container and a unique key -ShareID- for every opened model. By the use of this key, every currently loaded model can be identified clearly. A SharID is created by using the current control server milliseconds time value. This method is easy and effective because for example, the generation of a random number takes more calculation time.

In Addition, the uniqueness of the number has to be tested. Testing the uniqueness needs an average of $O(n*\log(n))$ and in worst case $O(n^2)$, if the fast Quicksort algorithm is used (Cormen 2004). Instead of the random number generation, the milliseconds method did not need saving and administrating a key list. All generated keys are automatically unique because of the time factor.

Access control for simulation models

By the use of authentification and authorization, the access to the models can be limited. For this reason, every modeling expert has to authorize himself when starting the modeling tool; it is not possible to use the tool without authorization. Simulation models and their components have a restricted access to their data areas. Every buildingblock contains an access list with four access entities:

- modify-right
- use-right
- delete-right
- administrate-right

When setting the mentioned rights, it is possible to set different access levels to a buildingblock. Figure 3 shows the different possibilities to set and mix the different access-rights.

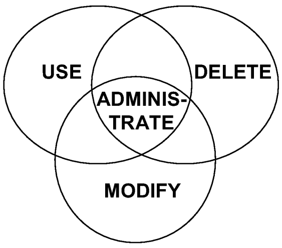


Figure 3: Different rights and the possible overlapping

To delete an object the modeler must have at least the delete-right. The use-right suffices if a modeler can apply the object or not. Furthermore, it allows users to create element instances. Changing the object in size, behavior, etc. is controlled by the modify-right. The administrate-right is the highest level of object-access; it contains the other three rights. By this, only modelers who own the administrate-right are able to change the access right for the object. They are able to change their own as well as other users' rights. Nevertheless, it is not possible to delete the elements creator whereby the exclusion of at least one modeler for element is prevented.

The access-rights were stored in a binary way. The delete-right is represented by the first bit, use by second, modify by third and administrate uses the fourth.

$$G = \sum_{i=1}^{n} R_i \geq n * 2^R \qquad \begin{array}{l} n = \forall \text{ selected objects} \\ R_i = \text{acccess - rights for element } i \end{array} \qquad (1)$$

The calculation of the different right for the selected objects uses the formula above (1). By the use of this formula, a fast calculation, of access-rights, during the modeling process is possible. To work with the selected objects two conditions have to be fulfilled: The corresponding rights as well as the lock-flag (cp.section *Conflicts during the modeling process*).

Furthermore, every object has more dates, e.g. unique number, version number and the actual data content. The locking entry on the data block level defines whether a modeler is qualified to access the element. The actual data content is represented by the data entry, which is responsible for the object's logical behavior. In addition, the object's connection links to other models or sub- models are stored in the data content. Figure 4 illustrates the formal design of a buildingblock. Restoring older stats or versions of an element is possible by using a unique number and an actual version number.

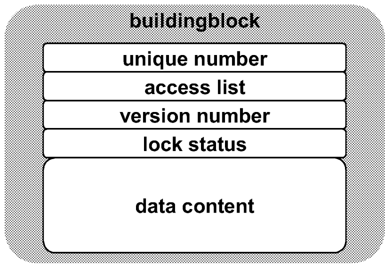


Figure 4: Buildingblock data-architecture

All simulation data is saved in a central storage. For this reason multiple access mechanisms to get this data areas are developed. The control server caches the loaded models, with which someone is currently working. On every users action the servers gets a control-message to evaluate. It updates the cached model and distributes the answer-message to all applied modeling tools that had currently opened the changed simulation model.

The Modeling Process

To afford teamwork in developing a simulation model coordination mechanisms are required. During the modeling process, the model and its sub-models have to be managed. Therefore, several interaction techniques are used to prevent errors. It is necessary to allocate behavior types to the used objects. This parameterization defines how objects can be placed, respectively integrated in the scenery. The major types are:

- create and use
- select
- delete
- edit and move
- connect

These types are considered more detailed during the next paragraphs. In General, they are executable by every user but there are some restrictions by the access rights, explained in the *access control for simulation models* section.

Create and use: During the development of a simulation model, creating objects is a fundamental step. The existing objects are stored and archived in component libraries in the model database. They can be loaded into the modeling module. After loading a model, all its elements can be used if the modeler has enough access-rights on them.

By using a drag and drop mechanism, a loaded model or some of its components can be placed in other modeling scenarios. Figure 5 shows the modeling tool with one open scene (out of three) on the right side and two loaded libraries on the left side.

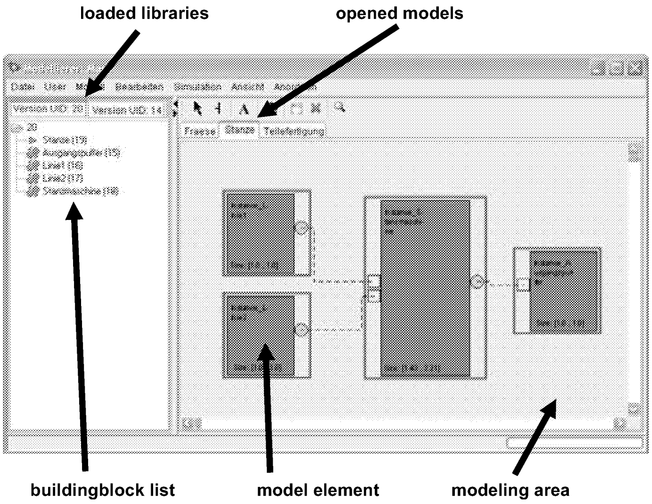


Figure 5: Modeling module

Select: Selecting objects is the precondition to apply other types of interaction. Selection is required before any movement or other actions on a focused object can be generated. When selecting an object, communication between the modeling tool and the control server starts. A control-message was transferred from the tool to the server. This control message contains e.g. the objects unique number (q.v. Figure 4) and the user-identifier. The control server looks in the cached model, and evaluates if the user has enough access-rights to accomplish the required action. Afterwards an answer-message is sent back to the modeling tool, and to all other applied tools. It causes them to execute the required selection action.

To display the selection action, the users get a visual respond on the objects. The objects are colored blue if the selection was successful, red if another user is blocking the object and gray if no one is currently trying to select or is working on the object (these colors can be changed in the tool properties). Thereby the users can see if their selection was successful.

Delete: It is essential that objects are deleting able to make the development work flexibly and comfortably. Furthermore, object deletion allows a recreation of the scene at any play and any time. Therefore, errors and mistakes can be repaired. Similar to the selecting process, the deleting process of an object demands a message transfer too.

Edit and move: To give more flexibility it is necessary to move or edit placed objects. This option refers to the objects layout and allows an organization of the simulation scene. Modelers with enough access-rights can select an object and move it within the scene. Furthermore, the user can modify the objects size and its variables, to influence the elements behavior. Also it is possible to adjust some variables that influence the objects behavior during the simulation. By changing calibrations, the material flow within the object changes. By the use of moving features, it is possible to clarify the activity structure or the hierarchical order.

Connect: To generate the relation between the objects it is possible to set some connections. For example, connections describe how object-children response to its object-parent. In addition, the connections determine the flow of the marks during the simulation process.

As long as the working environment is limited to one user at the same time and the same model, all the described actions can be handled easily. Nevertheless, through the possibility of a multi-user interface, several conflicts can be generated by the application of several types at the same time on the same model. The next paragraph describes these conflicts and how they have been solved.

Conflicts during the modeling process

The system follows the ideal of a multi-user system to model and simulate parallel in any situation, still upcoming conflicts had to be handled. For example if two or more persons have significant ideas or defaults to be directly integrated in the simulation model at the same time. Figure 6 illustrated one possible error situation, when two or more modelers want to edit the same element at the same time in the same simulation model. This conflict has to be solved by the control server.

According to this, possible conflicts are, for example, the access on the same object, as shown in Figure 6 or the delete of an object, which is necessary for another. The next paragraph will point out the most important methods to solve the conflicts during the modeling process.

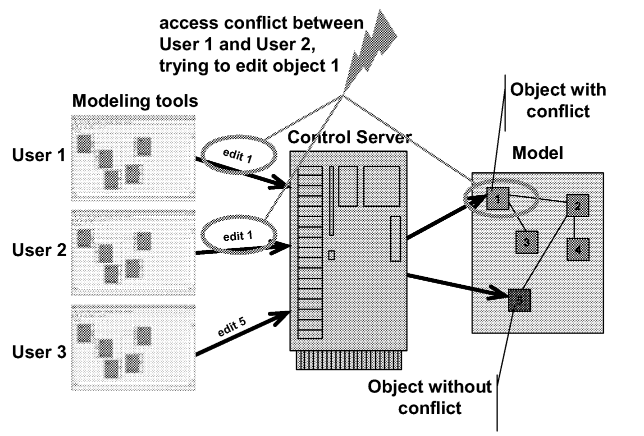


Figure 6: Possible conflict during multi-user design process of one simulation model

Conflict handling

In order to handle the possible conflicts, two methods have been developed. In these methods, the locking and versioning mechanism enable multi-user work on one simulation model at the same time and will be described in the following paragraph.

Locking: By the use of lock mechanisms, several users can work on the same model. As long as they work in different parts of the model, no conflicts are generated. If two users want to access the same object, it is locked by the first user and stays locked until he finished his adjustment or ends his modeling session. Only the user who owns the lock-flag can work with the objects. When a user finishes his modeling session all the buildingblocks in which he is still locking become unlocked automatically. Therefore, it is guaranteed, that objects formally locked by the user could be access able after he has logged out. The control server administrates a list with all the locked objects and the lock-flag owner in it. So it is possible to unlock all elements locked by a modeler also if he is locked out. Furthermore all currently blocked objects are delivered to users who enter the modeling team after the work session has already started.

Because the users did not know automatically, the locked objects have to be visualized, to assist the users in recognizing which objects are currently locked. Therefore, locked objects become painted in different colors (the different signal colors are mentioned in *the modeling process* paragraph).

Remembering that the model bases on different levels of hierarchies, the implementation of the locking method is more difficult. The lock of a specified object is deeply connected to the lower (more detailed) levels of the simulation model. Furthermore, the upper elements connected to the locked object have to be marked, so that they also cannot be changed. Figure 7 illustrates an example of the locking mechanism. Despite this, the system allows users to work parallel on the model on each level whenever two objects are not connected in their hierarchical order. Conflicts can only occur when two different users work in the same branch. The users will directly or indirectly take influence on sub- or superior objects by editing sub-models of a locked node from a higher level.

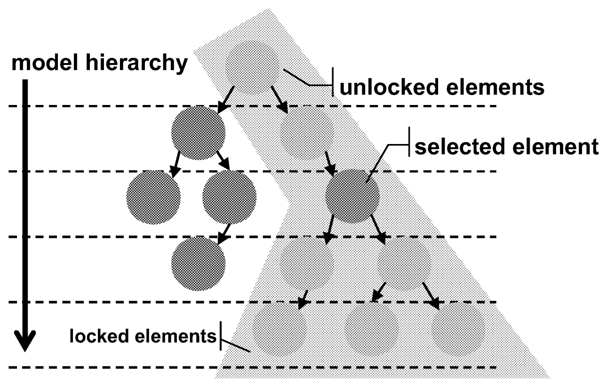


Figure 7: Extended locking in a model

Versioning: While saving the different changes of the model, a new version number is given and the control server (cp. Figure 2) deliverers all updates to the other team members' modeling tools. If a new version number is given, the changes always have to be recorded, so that former versions can be computed. Every object of the basic model has its own timeline. Therefore, it is possible to reproduce all its changes over the different versions.

Through the versioning, it is possible to generate older versions from the new ones, and therefore it is only necessary to save the actual version number, if an experiment is configured.

FUTURE PROSPECTS

Until now, the system supports multi-user work at the same time at and different place by the use of locking mechanism and an access-rights management system. Only authorized users with an adequate right are allowed to change the model. Furthermore, working at different times is also supported. The use of a versioning mechanism allows to restore different states of a simulation model.

In the next development steps of the d3 FACT insight tool a high quality three-dimensional modeling tool with multi-user support is planned. By the use of a three-dimensional modeling component, the design process of material flow models can be accelerated. A more intuitive design is is envisaged; the objects will no longer be represented by boxes but by realistic looking three-dimensional models. Thus, the designer of a model becomes integrated into the scene he is working on.

REFERENCES

Banks, J. 1998. *Handbook of simulation: Principles, Methodology, Advances, Applications and Practice*, page 3, John Wiley & Sons, INC.

Bengel, G. 2002. *Verteilte Systeme: Client-Server-Computing für Studenten und Praktiker*, page 18 et seqq., Vieweg.

Cormen, T. H. 2004. *Introduction to algorithms, an introduction*, pages 127-144, Oldenburg

Dangelmaier, W.; Laroque, C.; and Mueck, B. 2005. „d3 FACT insight – Mehrbenutzerfähiges Modellieren und Simulieren von Materialflüssen verschiedenartiger Organisationsformen". In: Gausemeier, J. and Grafe, M. (Hrsg.), *Augmented & Virtual Reality in der Produktentstehung*, Bd. 167, pages. 147-155, HNI-Verlagsschriftenreihe.

Dangelmaier, W.; Laroque, C.; Mueck, B.; and Renner, P. 2004. "Improved Process Planning by a Material Flow Simulation with Multi-User-Support". In: Geril, P.; Baake, U; Herbst, J.; and Landeghem, R. V., editors, *11th European Concurrent Engenieering Conference: Worldwide Partnerships and Mergers*, pages 59-63, EUROSIS.

Dangelmaier, W. and Mueck, B. 2003. "Simulation in Buisness Administration and Management – Simulation of production processes". In: Mohammad S. Obaidat and Gerogios I. Papadimitriou, editors, *Applied System Simulation – Methodologies and Applications*, pages 382-396. Kluwer Academic Publishers.

Johansen, R. 1991. "Teams for tomorrow". In: *Proc. 24th IEEE Hawaii Intl. Conf. On System Science*, pages 520-534, IEEE Comp. Soc. Press.

Klingstam, P. and Gullander, P. 1999, "Overview of simulaiton tools for computer aided production engineering". In: *Computers in Industry*, No. 38, pages 173-186, Elsevier.

Marczinski, G. 2005. "Digitale Fabrik – mit dem 4-Stufenmodell zum Erfolg". In: Gronau, N. and Scholz-Reiter, B., *PPS Management: ERP Systeme in Produktion und Logistik*, Nr. 2, pages 38-41, GITO.

Mueck, B. and Dittmann, N. 2003. *Marktanalyse Materialfluss-Simulatoren*, Nr. 11, ALB-HNI Verlagsschriftenreihe, Paderborn.

Tanenbaum, A. S. and Steen, M. 2003. *Verteilte Systeme: Grundlagen und Paradigmen*, page 66 et seqq., 2003, Pearson Studium

BIOGRAPHY

WILHELM DANGELMAIER was director and head of the Department for Cooperative Planning and Control at the Fraunhofer-Institute for Manufacturing. In 1990, he became Professor for Facility Planning and Production Scheduling at the University of Stuttgart. In 1991, Prof. Dangelmaier had become Professor for Business computing at the HEINZ NIXDORF INSTITUTE; University of Paderborn, Germany. In 1996, he founded the Fraunhofer-Anwendungszentrum für Logistikorientierte Betriebswirtschaft (ALB).

MARK AUFENANGER studied business computing at the University of Paderborn. Since 2005, he is a research assistant at the group of Prof. Dangelmaier, Business Computing, esp. CIM. He is mainly interested in material flow simulation in general, simulation of logistics systems and logics in simulation models.

DANIEL HUBER studied industrial engineering at the University of Paderborn, Germany. Since 2005, he is a research assistant at the group of Prof. Dangelmaier, Business Computing, esp. CIM. His main research interests are simulation of logistics systems in virtual environments and multi-resolution modeling.

CHRISTOPH LAROQUE studied business computing at the University of Paderborn. Since 2003, he is PHD student in the graduate school of dynamic intelligent systems and research assistant at the group of Prof. Dangelmaier, Business Computing, esp. CIM. He is mainly interested in material flow simulation and the “digital factory”.

KIRAN MAHAJAN studied mechanical engineering at the Delft University of Technology in The Netherlands with a specialization in production engineering. Since 2004, he is a research assistant at the Heinz Nixdorf Institute. He has a bachelor’s degree from the University of Pune (COEP), India. His research interests are interactive analysis of manufacturing simulations and 3D visualization.

Virtual Factory: an educational workbench for job shop scheduling

Jan Soons, Hendrik Van Landeghem

Industrial Management, Ghent University, Belgium
Jan.Soons@UGent.be, Hendrik.Vanlandeghem@UGent.be

***Abstract*—In this paper we give a description of the concept of the "Virtual Factory", a workbench for job shop scheduling. Furthermore, we describe the input and output that is given to the Virtual Factory and how this can be used to illustrate the effects of different scheduling rules in factory logistics.**

***Keywords*— Virtual Factory, factory logistics, simulation, education, Enterprise Dynamics**

I. Introduction

Simulation has been acknowledged as a powerful way to support training in complex concepts and systems. Its graphical interface and user interaction allows it to be the heart of many game-format educational softwares [1].
The Virtual Factory is created as a simulation model in Enterprise Dynamics. The simulation data (input and output) is stored in a database.
The goal of the virtual factory is to create an educational workbench on which the effects of different scheduling rules towards the global factory performance can be studied.
Currently, the layout of our sample factory is fixed, but in a future version this layout will be generated from the database. Our sample layout is based on the features of the Goldratt game, but the size is much smaller. It is based on 7 work centers, 4 resources (machines) and unlimited storage of raw materials [1],[3].

II. The model

The virtual factory is a job shop environment implemented in a simulation software. We created a model for this simulation software, which mainly consist of two parts: a **material flow** model and an (job) **order flow** model.[3]

The **material flow** model implements the factory layout. We have 3 product types which can be ordered. Each product type has a different BOM (Bill of Material), but needs partly the same resources. A place in this production flow where a given resource processes the same materials in the same production step is defined as a **work center (WC)**.

In our example every work center needs one resource (operator), and since there are more work centers than resources, this means that the resources are shared between the work centers

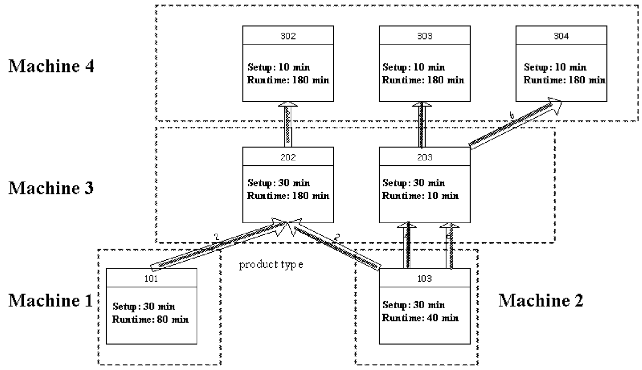


Figure 1: The material flow model

The work centers in the factory are **job driven**. To start a job, 3 initial conditions must be satisfied:

1) *A resource must be available in the resource pool*
2) *All (raw) materials must be available*
3) *A job instruction must be supplied to the work center.*

For the implementation of **(3)** we needed to create a **job processing system** in our model. It is important to make a distinction between an order and a job. An **order** is placed by a customer to obtain a product. It is the instruction to start the production of a product. For every product, the different steps in the production process are done at the **work centers**. The instructions for these processes are called "**jobs**". To schedule the virtual factory one must schedule all the jobs for the work centers. As a consequence it is possible to:

1) create intermediate storage of materials
2) to reallocate this intermediate storage for another order.

III. The job scheduling mechanism

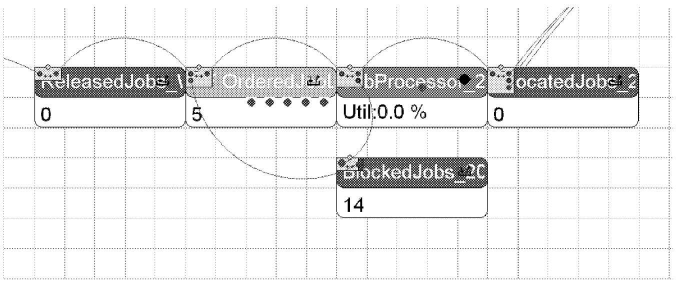


Figure 2 Job selection

The job selection is illustrated in Figure 2. In this part of the job processing system, the selection of the released jobs is done. Every released job is sorted on priority in a list OrderJobList. When this ordered list is complete, the jobprocessor takes the job with the highest priority first and checks if the necessary materials are available. If these materials are available, the job is sent to the model (allocatedjobs) and the materials are allocated. If the materials are not available, the job remains blocked. When new materials are produced, all jobs will be sorted and reevaluated again.

In the model the jobs are executed when all the materials are available.

IV. INPUT

In this section we will give a broader description about the virtual factory's input.

The assignment for the practicum is to compare 5 scheduling strategies: (1) **FIFO**, (2) **EDD**, (3) **EDD with delayed release dates**, (4) **SPT** and (5) **the optimal strategy**. [4]

The input of the practicum is an excel worksheet in which the scheduling of the jobs is done by the student. The output of the virtual factory is a set of Excel worksheets that reflect how the factory has executed the schedule of the student.

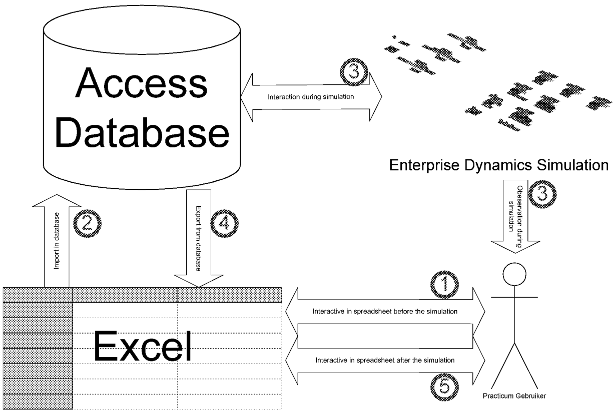


Figure 3: Software components of the Virtual Factory

As shown in Figure 3 the input worksheet is imported in a database before the simulation starts. After the simulation the results are exported from the database to the output worksheets.

The input sheet of the virtual factory consist of one worksheet (Figure 4). This worksheet (generated with VBA) contains a list of al the necessary jobs to produce the orders. To implement the scheduling rule, each job is assigned a priority and a release date (and time). The release date is the time when the job enters the simulation. Only after this time, the job can be selected for execution. After releasing the jobs in the simulation, the jobs will be executed. The job with the highest priority (highest number) will be executed first, if the materials are available.

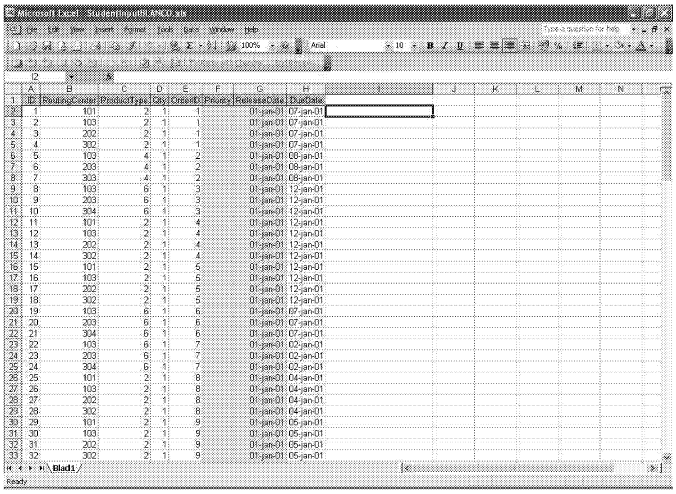


Figure 2 Input sheet

V. OUTPUT

During the simulation, one can watch the ongoing production in the virtual factory so a rough indication of the performance is already available during simulation.

After the simulation, a lot of data is exported to three excel worksheets. The first worksheet contains global factory information, the second contains information about the delivered orders to the customers and the last worksheet contains information about the executed jobs.

The global factory information contains the workload of the different work centers. For all the work centers where one machine is assigned, the sum of the loads of the assigned work centers equals the load of the machine (e.g. machine 4).

The worksheet with the order results (Figure 3) shows when an order is finished. If these results are compared with the due dates of the orders, the average tardiness, lateness, makespan and WIP can be calculated.

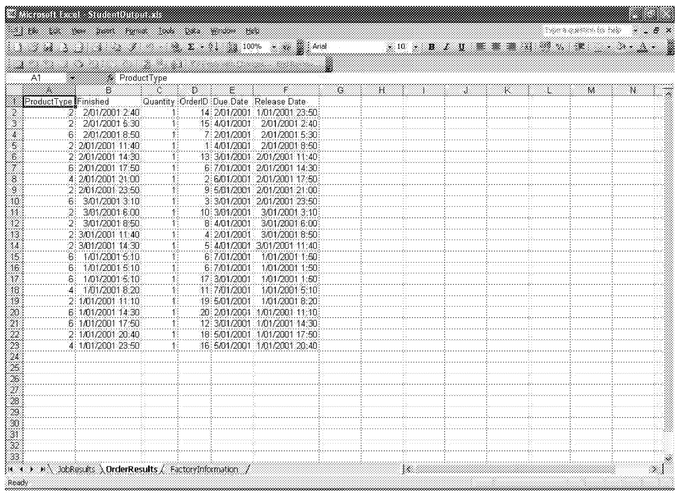


Figure 3 Worksheet with order results

The last worksheet contains the results of the executed jobs. It is on this level where we can find the most interesting information of the simulation. The sheet contains the actual start and stop time for all of the executed jobs.

With this output, one can compare the scheduled data and the results of the simulation and start analyzing the differences.

Two nice representations of the results from the job database are shown in Figure 4 and Figure 5.
The first one shows the number of jobs in process (WIP). The first graph gives the start times of the jobs (cumulative) and the second graph gives the end time of the jobs (cumulative). This gives a good idea of how much WIP there is (vertical distance between the two graphs) during the entire simulation. The average job lead time is obtained as the horizontal distance between the two graphs.

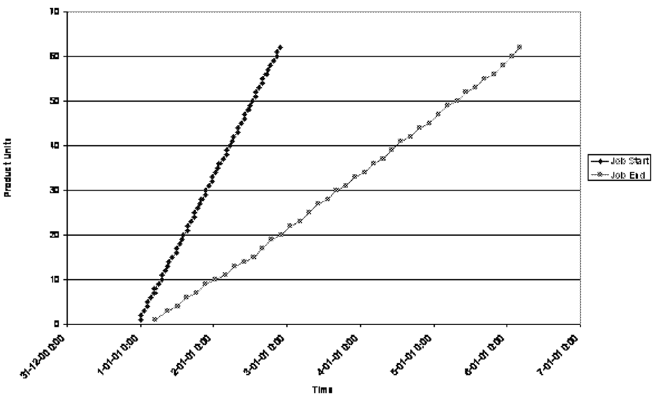


Figure 4 Jobs started and finished

The second representation is a Gantt chart, with an overview when the jobs were executed. On the vertical axis, the different jobs are represented and on the horizontal axis some fixed time frames are represented (e.g. each time frame represents a period of 3 hours). The result of this representation is shown in Figure 5.

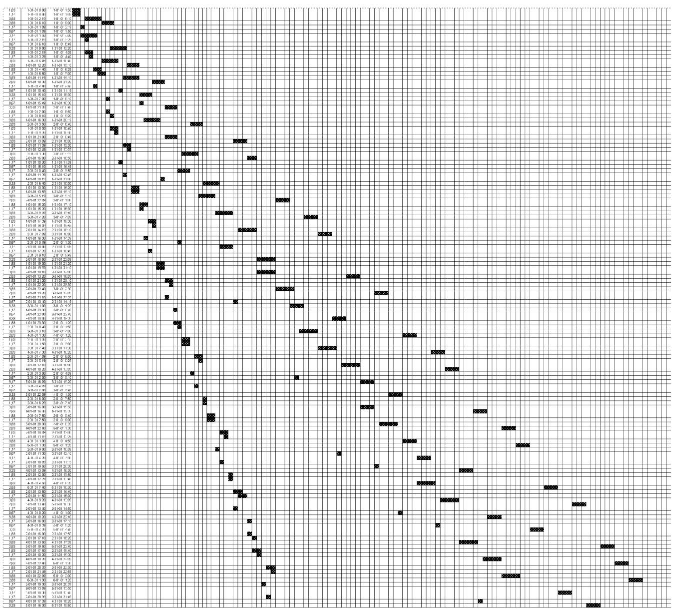


Figure 5 Gantt chart of jobs

VI. CONCLUSIONS

After organizing some first test labs with this version of the virtual factory (version 1.03), both in Belgium and Mexico, we can make some conclusions.
The idea of scheduling with a release date and a priority for the jobs is very flexible in the practicum. Another successful thing is the manipulation of the input and output in a worksheet. Students are used to work with a spreadsheet so that manipulating the data in a spreadsheet is much easier then doing this in simulation software like Enterprise Dynamics.

Another advantage of processing the input and output in excel is the possibility of simulating real systems, like small enterprises. In this aspect the layout of the virtual factory can be adapted.

The current version of the virtual factory is not in a final state. We plan to generate dynamic factories, which will also be read from the database. Also we plan to generate the input for the simulation from MRP software, so that the schedule can be constructed automatic and that students can measure the effects of MRP software.

Acknowledgements
Part of this research was sponsored by the European Commission, as part of an ALFA Project "Virtual Engineering Workbench"[5].

REFERENCES

[1] Riis J., Smeds R., Van Landeghem R., (Eds) Games in Operations Management, Kluwer Academic Publishers, 2000.
[2] Van Landeghem H., Acevedo M., Virtual Factory Version 1.03, January 2005
[3] Soons J, Ontwerp van een virtual factory, Master Thesis, Ghent University, August 2005
[4] Hopp, Wallace J. Spearman, Mark L. , Factory physics : foundations of manufacturing management McGraw-Hill, New York (N.Y.) 2000
[5] Website alfaproject: http://www.virtualsimlab.dep.ufscar.br/alfa

CE-ENHANCED LEAN MANUFACTURING

Value Stream Maps and Critical Path Method for production and procurement planning: a Make To Order company

Flavio Tonelli
Laura Piccinino
University of Study of Genoa
Street all'Opera Pia, 15
16145 Genoa, Italy
E-mail: {flavio.tonelli|laura.piccinino}@diptem.unige.it

KEYWORDS
Lean Manufacturing, Value Stream Map (VSM), Critical Path Method (CPM), Rough Cut Capacity Planning (RCCP).

ABSTRACT

The concept of Value Stream Map (VSM) is inserted as key point in Lean Manufacturing (LM). VSMs, originally named "material and information flow maps", allow a one page diagram representation of the whole process required to realize one or more final products. Because of this they are a common tool used to describe production processes and to underline the business areas in which wastes are identified. Nevertheless the aim of this paper is not to introduce a practical example of VSMs implementation, how to define a new approach that uses VSMs as tool to support production and procurement planning according to scarce resources capacity over a critical production path. This approach is articulated on 5 sequential steps and integrates the classical technique of VSMs with tools characteristic of planning theory like Critical Path Method (CPM) and capacity planning techniques like Rough Cut Capacity Planning (RCCP).

INTRODUCTION

Senior managers often state, "We are not globally competitive," and usually attribute this to high labor costs (McDermott, 2002). VSMs outline that reasons for poor competitiveness are usually due to a waste abundance inside organizations (Emiliani and Stec 2004). Lean Enterprise Management (LEM) is a comprehensive management model, evolution of the well known Toyota Production System, including Lean Manufacturing, Assembly, Accounting, etc… According to the manufacturing aspects the LEM proposition is to produce in slender and flexible way, pursuing a continuous struggle to every form of waste ("Muda"), both that this is an unnecessary inventory, defects, waiting, overproduction, transport or inappropriate processing. In other words it is a struggle to every activity that doesn't give added value to the product in terms of final customer point of view. Today, this issue is particularly important because European industrial world is going toward a dangerous delocalization of the productive sites in more economical advantageous areas. Nevertheless in order to compete price (workforce, materials, …) is not the only element on which companies should look for. Supply Chain and Product excellence are relevant areas of development. Processes can be optimized by eliminating wastes and focusing on value added operations. VSMs are the right tool allowing practitioners to identify waste in individual value streams and, hence, find appropriate route to removal, or at least, reduce this waste. VSMs, originally named "material and information flow maps" are concretized in a one page diagram that represents the whole process followed to realize the final product (Emiliani and Stec 2004). Nevertheless, from the analyses conducted in these years on the business companies that have introduced some lean methodology, it is clear the greatest problem of the implementation phases depends on the not correct definition of the value flow through the productive process (Rother and Shook 2003).
In fact, following the five phases of lean methodology (Womack and Jones 2002) it is possible to:

1. find an agent of change;
2. find an advisor to submit the process of transformation;
3. underline a problem that motivates the action of business change;
4. define the flow of value for all products families;
5. start the waste elimination.

Typically the fourth phase is skipped in order to introduce an immediate implementation of the operational phase. The consequence of this choice is a limited and, often, least improvement in terms of product quality and stakeholders satisfaction. From these premises, it appears clear the important role that can be carried out by VSMs. These allow to underline, in visual and intuitive way, every phase of the productive process focused on the products family (whenever product families are a suitable and useful representation for a low volume high mix production).
Authors, looking at the current situation inside the company, have deepened and applied the first two phases. This paper is a technical and practical view of well known methodologies (VSMs, CPM, RCCP), applied to support the planning and control processes focusing only on value added operations.

CASE STUDY DEFINITION

This approach has been developed functionally to a Make To Order (MTO) company in which the Authors are leading a LM change management process. The company specializes in heavy duty equipment so it is characterized by limited mix of products, high lead times and job shop organization. This particular operational world faces every day with the following problems:

- limited mix of products;
- elevated lead times (in many cases several months);
- job shop production, activated on order.

The company has a customers portfolio and knows in advance due dates for each product of a production lot. Starting from the company issues the most important problem was identified with the impossibility to schedule and dimension production phases and work centers capacity because the effective production routing is, usually, known only with a low detail level. Even if a Material Requirement Planning (MRP) solution seems to be a solution to plan production and procurement the products are so different each to other (even inside the same lot) to require, often, a complete review of the bill of material. From this point of view for some products the production strategy is more like Engineering To Order (ETO) than Make To Order (MTO). In this contexts fixed Bills Of Materials (BOM) and reliable routings are a dream and scarce resource capacity planning if often a nightmare at least accomplished with an high level of detail. In fact, MRP can be performed only for BOM commonalities and it isn't representative for critical resource capacity planning. An RCCP solution could be acceptable in order to verify scarce resource overloads for resources belonging to the critical path. Nevertheless an RCCP solution requires planning bills, planning routings, logical resources (groups of physical critical resources) and, finally, a job shop model on which check related capacity diagrams and identify overflows. Many of this information are provided by VSMs. Starting from this consideration Authors decided to use the VSM representation to collect and represent data required from a simple but effective RCCP performed on critical resources adding value to production process.

PROPOSED APPROACH

The proposed approach is based on sequential steps that integrates the classical technique of VSMs with tools characteristic of planning theory like CPM and capacity planning techniques like RCCP.

VSM represents the flow of all operations (both on added value and not) demand for the product production, from the row materials reception to the ended products delivery. Each activity is represented by standard icon ("Data Box", that reports operational value required), and the arrows represented the precedence relationship.

In particular, the implementation of a VSM, is articulated in four distinct steps:

- selection of the product family;
- creation of a current-state map (AS-IS): diagram of actual activities and relationship;
- creation of a future-state map (TO-BE): definition of a new diagram of activity and relationship, improved through lean techniques and best practices
- implementation of the future-state map through structured continuous improvement activities.

Besides this technique is particularly useful in all productions characterized by a linear flow, not there where production is complex with the overlap, on the same job centers, of different products characterized by different times of production (job shop).

VSM is translated into a network. In the network representation, each activity is represented by an arc pointing in the direction of progress and the node of the network establish the precedence relationship.

CPM methodology is based on the network to calculate the time schedule for the activities flow. In particular CPM produces the following information:

- diversification of the activities in critical and noncritical;
- determination of the Total Productive Lead Time.

An activity is critical if there is no leeway in determining its start and finish time, besides is noncritical if the start time can be advanced or delayed.

To determine the critical path, CPM involves the two following passes:

- Forward pass: determines the earliest occurrence time of the events;
- Backward pass: determines the latest occurrence time of the events

Defining:

$\square_j$ = Earliest occurrence time of event j

Δ_j = Latest occurrence time of event j

D_{ij} = Duration of activity (i; j)

Forward Pass: the computation starts at node 1 and advances to the end node.

Initial step $\square_1 = 0$

General step j for each node (i) linked directly to node j by incoming activities and the earliest occurrence times of events have been computed

$\square_j = \max_i \{\square_j + D_{ij}\}$

The forward pass is completed when $\square_n$ is computed

Backward Pass: the computation starts at node n and advances to node 1.

Initial step $\Delta_n = \square_n$

General step j for each node (i) linked directly to node j by outgoing activities and the latest occurrence times of events have been computed

$\Delta_j = \min_i \{ \Delta_i - D_{ji}\}$

The backward pass is complete when Δ_1 is computed.

An activity (i, j) is critical is:

$\Delta_i = \square_i$

$\Delta_j = \square_j$

$\Delta_j - \Delta_i = \square_j - \square_{i} = D_{ij}\}$

CPM provides the Finite Assembly Schedule (FSA). FSA is calculated scheduling critical activities one right after the other. The noncritical activities start as early as possible. So

slack period can be used to absorb unexpected delays in the activity execution .
RCCP defines the load for each critical work centres (bottlenecks, highly utilized resources, etc.). For these resources, RCCP attempts to balance workloads, through the introduction of work on Saturday, double turns, delaying non critical activities, etc.
So it is defined a feasible FAS.

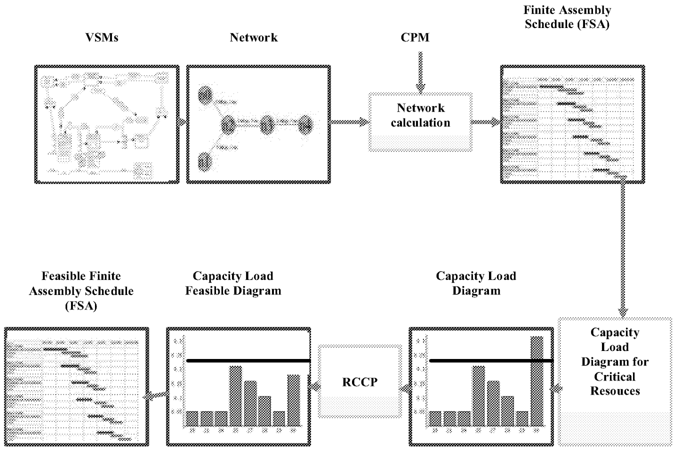


Figure 1: Sequencing Steps Approach

A SAMPLE OF PRACTICAL IMPLEMENTATION

The practical implementation realized by the Author is based on the following phases:

- **Phase 1:** definition of the product macro-families;
- **Phase 2:** representation of the main production flows for each macro-family by using VSMs;
- **Phase 3:** definition of the Critical Path (by using CPM technique and setting activities starting date with tardiness criteria), for each macro-family starting from VSMs;
- **Phase 4:** definition of Master Production Schedule (MPS);
- **Phase 5:** leveling of resource loads using RCCP technique.

Phase 1: Definition Of Products Macro-Families

The focus of the study is not the productive cycle of each single item, because it reverts the management of an immense and detailed data structure, in which the reference to the general targets would be lost. The focus is the productive cycles of a limited number of macro-families.
For the determination of the macro-families the correct methodology is the product-process matrix, but in the case proposed, it would be result not understanding for the production complexity. However Author gathered together products that, through operator investigations, result to cross the same production centers and to be characterized by comparable production lead time. Production lead time is the operative time during which an operator work on a product.
Throughput time is the sum of all activity production lead time and set-up time.

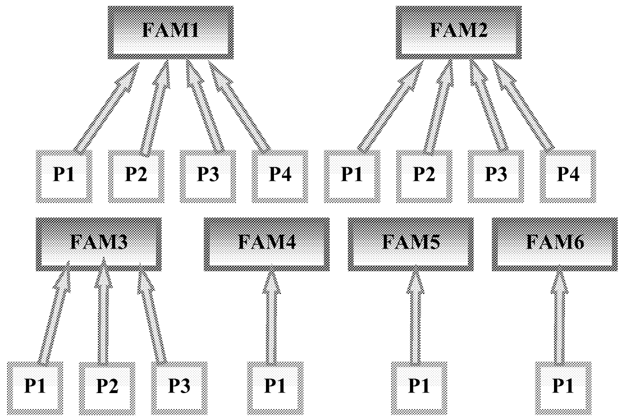


Figure 2: Example of macro-families defined

Figure 2 represents the six macro-families defined. As it appears, FAM 4, FAM 5 and FAM 6 are constituted by only one product because these last introduce unique features that are not comparable with any other.

Phase 2: Representation Of The Main Production Flows For Each Macro-Family By Using VSMs

The activities that allow the product production involve different organizational departments and the process complexity is featured by a general increment related to the realization phases. To reconstruct the correct productive flow, is essential both to give questions to operators that have cycle complete vision, both to notice physically and personally the data characterizing the material and information flow.
VSM layout is based on standard icons:

- Material Flow Icons;
- General Icons;
- Information Flow Icons.

Each icon can be used singularly or combined with others, according to the requirements of the individual value stream.

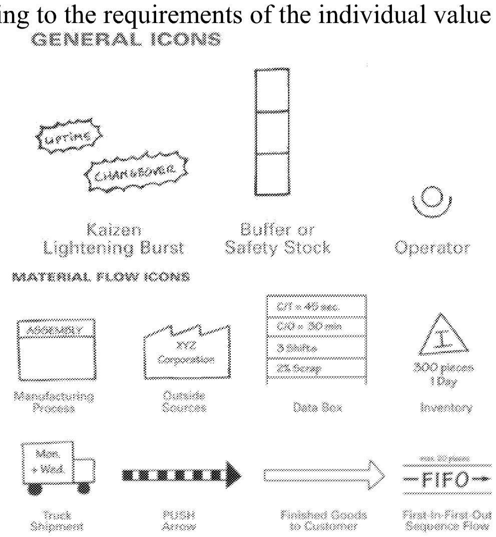


Figure 3: VSM Icons

VSMs report values entirely brought to the materials flow. The salient inserted data have an operational character and concern the definition of each macro-operation, in terms of:

- average working time;
- average set-up time;
- middle production time (sum of average working time and average set-up time);
- number of operators involved;
- department of affiliation (suitable graphically through a color identification of the department);
- macro-operation immediately upstream and downstream.

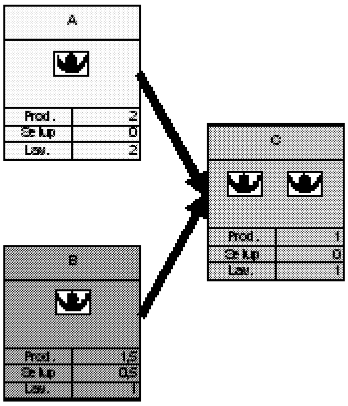


Figure 4: Example of VSM part

In this first phase of study, data are not related to the stock levels present through the productive process, neither to the information flow. In fact the construction of the future-state map is far from the following study focus. As over mentioned the actual target is the realization of a framework that uses VSMs as tool to support the production planning. VSMs allow to determine the correct sequence of the productive operations and to underline value added activities (critical activities).

Phase 3: Definition Of The Critical Path

Once defined macro-family VSMs the study focus on the Critical Path. VSMs data, in terms of task, production time and precedence relationship are been inserted in CPM based tool. It outlined the critical activities, that determine the productive LT for the macro-family, and the noncritical, that are characterized by a slack time and are allocated as soon as possible.

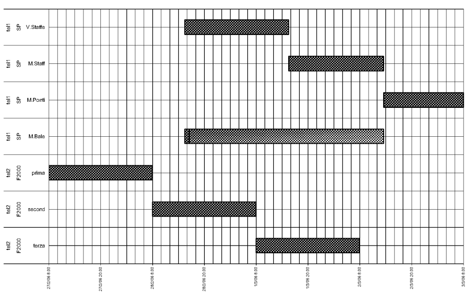


Figure 4: Critical Path

Automatically the Critical Path is transformed in a Gantt diagram correlated by capacities data.

Gantt capacities diagram is a graph with time (in days), on the abscissas, the department name on ordinates and the total capacities required by each macro operation as data (in hours).

Total Capacities Macro-Operation
= operators number required * macro-operation production time

Phase 4: Definition of Master Production Schedule (MPS)

Considering the Gantt capacities diagram and the customer orders, it's possible to define the MPS. Author released a sheet that calculates automatically, bucket for bucket, the capacity needed to each department. The procedure for the compilation of the planning sheet is to:

1. identify the macro-family for each product of a customer order;
2. to plot the Gantt diagram (obtained from the macro-families VSM) starting from the promised customer date;
3. repeat operation 1 and 2 until all customer orders exhaust.

The planning sheet provides, bucket for bucket, the capacity required to each department as sum of the capacity of each single macro-operations of each macro-family presents in the same bucket. The capacity required is red if superior to a the max department capacities or to a value defined by customer, otherwise is green.

Max Department Capacity = total department operators number * daily times job.

Phase 5: Levelling Of loads resources using RCCP technique.

Verified capacity overloads, for a certain department in a certain bucket, RCCP technique attempts to balance workloads:

- adding required capacity;
- delaying noncritical activities.

In the first case is given the possibility to introduce a quantity of additional capacity to reply to the production overflow (work on Saturday, double turns, etc).
In the second case the users choose to delay noncritical activities in a bucket not yet completely saturated.
Finally it is outlined the available capacity. It is red if it is inferior to a % of the max capacity, otherwise it is green.

Available Dep. Cap. = Max Dep. Cap. + Add. Dep. Cap. – Required Dep. Cap.

CONCLUSIONS AND FUTURE DEVELOPMENTS

This paper is mainly focused on the use of the VSMs as tools to create a clear and correct layout of the business productive process and to allow the realization of a planning. In particular the results obtained concern the possibility to:

- create a complete and correct vision of the business productive processes;
- know in advance the loads of all activities;

- underline the overflow of production and to introduce the necessary corrective actions (for exemple anticipate of a part of the activity not critic in temporal bucket not saturate or introduction of additional capacity through overtime shift, Saturdays working, double shift, etc).

This tool has just been introduced so quantitative results in terms of Key Performance Indicators improvement are not still available. Besides in the introduction we have underlined as VSMs is above all an useful tool for the implementation of lean technique. For this reason, the future objective of this study will be to complete the layout of the present-state map, with the introduction of all data related to the flows of information and progress to the search and elimination of the business wastes with the creation of the future-state map.

REFERENCES

Chase R.B.. Jacobs F.R.. Aquilano N.J.. Grando A.. Sianesi A.. 2004. “Operations Management for Competitive Advantage”. McGraw-Hill.

Middleton M.R.. 2004. “Data Analysis Using Microsoft Excel, third edition”, APOGEO.

Vollmann T. E., Berry W. L., C. and Jacobs F. R.. 2004. “Manufacturing planning and control systems for supply chain management”.McGraw-Hill

Emiliani, M.L., Stec, D.J.. 2004. “Using value-stream maps to improve leadership”. *The Leadership & Organization Development Journal.*

Rother, M. and Shook, J. 2003. “Learning to See: Value Stream Mapping to Create Value and Eliminate Muda”. *Lean Enterprise Institute.*

Hines, P., Rich, N.. “The seven value stream mapping tools”. *Lean Enterprise Research Center, Cardiff Business School.*

Galgano, A.. 2002. “Le tre rivoluzioni: Caccia agli sprechi e raddoppiare la produttività con la Lean Production”. Gruppo Galgano.

Womack, J. P. and Jones, T. J. 2002. “Lean Thinking”. Angelo Guerini and associati..

Womack, J. P. and Jones, T. J. 2002. “Lean Thinking”. Angelo Guerini and associati..

Taha H. A.,2004. “Operations Research: an introduction”.Prentice Hall

A GA-based Approach to the Resource Allocation Problem in the Global Manufacturing Supply Chain

Xiao You, Lingling Li
Nanyang Technological University, Singapore
National University of Singapore

KEYWORDS

Supply Chain, GA, Resource Allocation.

ABATRACT

Resource allocation decision has been identified as a crucial issue for the successful implementation of global manufacturing strategies. Classical models and optimization methods have been proved inadequate for such a large and complex problem. A new extensive model of the global resource allocation problem (GRAP) is developed in this paper which incorporates the product structure and supplier priority constraints. The constraint-based genetic algorithm is applied to obtain an optimal solution for this large-scale problem.

1. Introduction

Under the constant pressure of global competition, diversified market, and product customization with short life cycle and low cost, companies are forced to shift from a centralized, vertically integrated and single-site structure into one that is decentralized, horizontally integrated and global distributed (Ford 1990). The global manufacturing strategy has been widely applied to transcend national boundaries and leverage capabilities and resources worldwide (Pontrandolfo and Okogbaa 1999). Since global manufacturing activities might be dispersed and carried out in diverse locations, resource allocation decision has been identified as a crucial issue for the successful implementation of global manufacturing strategies. Such a global resource allocation problem (GRAP) is much more complex than that for the procurement, production and delivery of a simple commodity, not only for the volume and complexity of transactions, but also due to its dynamic and heterogeneous manufacturing environments (Gaonkar and Viswanadham 2001).

The GRAP problem is primarily concerned with an optimal assignment of product flows along the supply chain. It arises when constructing a supply chain network between multiple supply chain entities like suppliers, plants, distribution centers and retailers. As a comprehensive problem, GRAP requires a systematic and holistic approach to consider product-related characteristics, supply and demand networks, factory loading, inter-enterpriseand intra-enterprise transactions. The complexity arises from the stochastic variability in production demand and supply, as well as manufacturing and transportation time due to long chains of foreign plants (Shobrys and White 2000).

The traditional approach for GRAP is to derive an analytical model and find an optimal solution through programming methods like LP/IP/MIP. However, linear programming could only handle small size problem. The computation time and required memory would be exponentially increased with the complexity of the supply chain. What's more, GRAP is a highly constrained problem that numerous constraints lie within and between supply chain facilities, including capacity, throughput, transportation etc. Some non-linear constraints like product family architecture, supplier priority make it extremely difficult to get an optimal solution for such a NP-hard problem.

This paper primarily focuses on strategic planning phases to allocate resource along the supply chain. We propose an integrated multi-commodity, multi-echelon analytical model to formulate the typical GRAP problem. A constraint-based genetic algorithm approach is developed for this highly constrained complex model.

2. Analytical model

This model focuses on a multi-commodity, multi-stage and procurement-production-distribution GRAP in a dynamic network configuration environment. It aims to minimize the cost through the entire supply chain, subject to supplier, plant and DC capacity and throughput constraints and customer demand requirements. Total cost includes fixed cost, variable cost and transportation cost etc.

In this study, we use the following indices: $i \in I$, a set of products produced by plants; $j \in J$, a set of candidate suppliers; $k \in K$, a set of potential plants; $l \in L$, a set of DCs; $m \in M$, a set of customer zones; $r \in R$, a set materials provided by suppliers.

Input parameters:

f_j^S Set-up cost for each supplier j
f_k^P Set-up cost for each plant k
f_l^D Set-up cost for each DC l

δ_{ik} Standard cost at plant *k* of product *i*

c_{rjk}^{SP} Unit transportation cost from supplier *j* to plant k for material *r* ($/unit)

c_{ikl}^{PD} Unit transportation cost from plant *k* to DC *l* for product i ($/unit)

c_{ilm}^{DZ} Unit transportation cost from DC *l* to customer zone *m* for product *i* ($/unit)

p_{rj}^{S} Unit purchasing cost of material *r* from supplier *j*

p_{ik}^{P} Unit producing cost of product *i* at plant j ($/unit)

p_{il}^{D} Unit cost of throughput (handling and inventory) for product *i* at DC *l* ($/unit)

Φ_{rj}^{S} Max production capacity of supplier *j* for material *r*

Φ_{ik} Maximum production capacity of plant *k* for product *i*

Φ_{k}^{P} Maximum production throughput of plant *k*

Φ_{ln}^{D} Minimum throughput of DC *k* (units/period)

Φ_{lm}^{D} Maximum throughput of DC *k* (units/period)

λ_{ri} Standard unit material *r* used by product *i* (unit)

α_j Service level of supplier *j*

β_k Service level of plant *k*

γ_l Service level of DC *l*

D_{im} Average demand for product *i* at customer zone

Decision variables:

X_{ik} Volume of product *i* produced at plant *k*

Y_{il} Volume of product *i* received at DC *l*

Z_{rj} Volume of material *r* provided by supplier *j*

Q_{ikl} Volume of product *i* transferred from plant *k* to DC *l*

R_{rjk} Volume of material *r* transferred from supplier *j* to plant *k*

P_{ilm} Volume of product *i* transferred from DC *l* to zone *m*

C Total cost of the entire supply chain ($/period)

$U_j=\{0,1\}$ indication variable whether a supplier *j* is open

$V_k=\{0,1\}$ indication variable whether a plant k is open

$W_l=\{0,1\}$ indication variable whether a DC l is open

$T_{lm}=\{0,1\}$ indication variable whether DC l serve zone m

The objective function is given as:

$$MIN\ \ C=\left[\sum_{rj} p_{rj}^{S} Z_{rj}+\sum_{rjk} c_{rjk}^{SP} R_{rjk}+\sum_{j} U_j f_j^{S}\right]$$
$$+\left[\sum_{k} V_k f_k^{P}+\sum_{ij} p_{ik}^{P} X_{ij}+\sum_{ikl} c_{ikl}^{PD} Q_{ikl}\right]$$
$$+\left[\sum_{l} W_l f_l^{D}+\sum_{kl} p_{kl}^{D} Y_{il}+\sum_{ilm} c_{ilm}^{DZ} P_{ilm}\right]$$

Subject to:

$$\sum_{k} R_{rjk}\le \Phi_{rj}^{S},\ \forall r,j \quad (1)$$
$$\sum_{k} R_{rjk}=Z_{rj},\ \forall r,j \quad (2)$$
$$\sum_{j}(R_{rjk}.\alpha_j)\ge \lambda_{ri}X_{ik},\ \forall r,i,k \quad (3)$$
$$\sum_{l} Q_{ikl}\le \Phi_{ik}^{P},\ \forall i,k \quad (4)$$
$$\sum_{il} Q_{ikl}\le \Phi_{k}^{P},\ \forall k \quad (5)$$
$$\sum_{k}(Q_{ikl}.\beta_k)\ge Y_{il},\ \forall k \quad (6)$$
$$Y_{il}=\sum_{m} P_{ilm},\ \forall i,l \quad (7)$$
$$P_{ilm}\gamma_l T_{lm}\ge D_{im},\ \forall i,m \quad (8)$$
$$\Phi_{ln}^{D}\le \sum_{i} Y_{il}\le \Phi_{lm}^{D},\ \forall l \quad (9)$$
$$\sum_{l} T_{lm}=1,\ \forall m \quad (10)$$
$$X_{ik},Y_{il},Z_{rj},Q_{ikl},R_{rjk}\ge 0,\ \ \forall i,j,k,l \quad (11)$$
$$U_i,V_k,W_l,T_{lm}=0\ or\ 1,\ \ \forall i,k,l \quad (12)$$
$$v_i+v_j-q.\,u_l\ge 0 \quad (13)$$
$$(-u_m-u_n-\ldots-u_p)+kv_i\ge 0 \quad (14)$$
$$v_i\ge v_j \quad (15)$$

The mixed-integer linear objective function minimizes the total fixed and variable cost. It contains three parts that come from different stages in the supply chain: 1)the raw material purchasing cost, transportation cost and set-up cost of supplier; 2) the set-up and production cost of plants plus the transportation cost from plants to DCs; 3) the inventory holding cost and set-up cost of DCs plus the transportation cost from DCs to Customer Zones.

3.1 Product structure constraints

Product structure specifies the quantity and ingredient of the material to produce the final product. It is the fundamental relationship between the materials which are the output of supply chain. Thus, the product structure is the core element of the supply chain and should be incorporated into the supply chain modeling. Unfortunately, the common graphical or tabular expression format of product structure, make it difficult to translate into a mathematical expression.

Two logical rules are extracted from the product structure:

I. If product P_1 is produced, then all of the components C_1,C_2,C_3 should be produced.

II. If component $C_1$is produced, then all of the materials R_1,R_2,R_3should be produced.

Assume $\overrightarrow{P_i}$ represents product P_i is produced; $\overrightarrow{C_i}$ represent component C_i is produced; $\overrightarrow{R_i}$ represent material R_i is produced. I, II are converted to rules:

III. $\overrightarrow{P_1}\Rightarrow\overrightarrow{C_1}\cap\overrightarrow{C_2}\cap\overrightarrow{C_3}$ IV. $\overrightarrow{C_1}\Rightarrow\overrightarrow{R_1}\cap\overrightarrow{R_2}\cap\overrightarrow{R_3}$

Assume $u_i=\{0,1\},u_i=1$ represent $\overrightarrow{P_i}$. $v_i=\{0,1\},\ \ v_i=1$ represent $\overrightarrow{C_i}$. $w_i=\{0,1\},w_i=1$ represent $\overrightarrow{R_i}$. The cardinality rules would be represented as following inequalities:

V. $v_1 + v_2 + v_3 - 3u_1 \geq 0$ VI. $w_1 + w_2 + w_3 - 3v_1 \geq 0$

The inequalities V, VI are used in the GRAP model as product structure constraints.

3.2 Supplier priority constraint

Suppose the plant y need at least one of the suppliers x_1, x_2 to provide material, then we have inequalities $-x_1 + y_1 \geq 0, -x_2 + y_1 \geq 0$. This represents if plant y is open, then either x_1 or x_2 or both of them should be chosen. It is very natural that company has preference for specific supplier. For example, in the same situation, we want supplier x2 more than supplier x1. How to incorporate this constraint into the model? We can use the inequality $x_2 - x_1 \geq 0$. This inequality has two results: 1) $x_1 = 1, x_2 = 1$. 2) $x_1 = 0, x_2 = 1$. This means supplier 2 would be selected first. By introducing this inequality, the supplier priority is incorporated into the mathematical model.

Supplier priority constraint reduces the possible combination of selected suppliers in the supply chain that decreases the computation complexity and improves the algorithm efficiency. Suppose there are five suppliers 1, 2, 3, 4, 5 in the supply chain. if the SC need three suppliers open, it has 10 (C_5^3) possible combination such as (1,2,3) (1,4,5) (2,3,5) etc. But if the priority of supplier is specified as: supplier 1 is better than supplier 3, and supplier 4 is better than supplier 2, then there just has 6 possible combinations. Since each supplier has different competency and concrete running condition, the constraint of supplier priority makes the model more practical. Through the supplier priority constraint, our model is not only limited on the cost factor, other performance measures also could be incorporated. For example, we may define the priority based on quality or lead time, on-time delivery etc.

4. GA-based approach for GRAP

Global supply chain configuration problem has the characteristics of large, complex, poorly understood search spaces. GA shows great potential for its simplicity and its ability to search through the complex decision spaces. It deals with a coding of the problem instead of decision variables.

To apply a GA solution to the GRAP, we construct an internal representation of the solution space to be searched, which is called a gene-code, and define an external function to evaluate each individual of a population. According to the progress of generations, individual change based on their gene-code variations is generated by the operators of crossover and mutation.

4.1 Chromosome syntax representation

In the global manufacturing supply chain, the number of involved parameters for optimization is very large. Traditional binary vectors used to represent the chromosome are not effective in such a large-scale dimension. To overcome this difficulty, a real-number representation is therefore used.

The following is the syntax representation of the chromosome:

B(s_1)B(s_2)…B(d_n)Q(s_1p_1)Q(s_1p_2)…Q(s_1p_n)Q(s_2p_1) …Q(s_2p_n)…Q(s_np_n)

It consists of six sub-strings. The first, second and third sub-strings are binary digits representing opened/closed suppliers, plants and DCs. The last three sub-strings are real number to represent the shipping amount between different supply chain stages.

A gene-code of a supply chain configuration is represented as Fig.2:

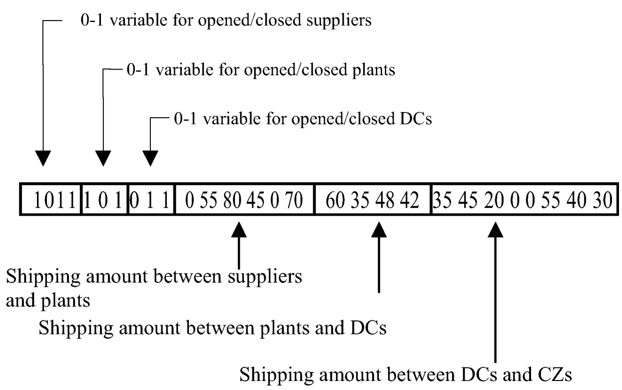


Figure 2: A chromosome example

4.2 Handling of configuration constraints

As indicated above, the problem of supply chain configuration is one that is heavily constrained. It spans from demand, capacity, product architecture and so on. These constraints limit the set of solutions from which an optimal solution is to be found. Under such a situation, a blind search and selection strategy tends to generate a high percentage of infeasible results and therefore inefficient. Therefore, constraint handling becomes a central issue during the application of GA method to supply chain configuration. We divide the constraints on the supply chain into two types of constrains on the GA: gene constraints and chromosome constraints. The gene constraints only impact on one gene itself such as the production capacity, DC throughput etc. The chromosome constraints should be considered by the whole chromosome including the production structure, supplier priority and entity pairwise relationship etc.

4.3 Initial population production

The population initialization technique used in the GA approach is a random real-number initialization which is performed by Procedure1. Procedure 1 creates a starting GA population filled with randomly generated real-number strings. The population size is set not less than twice the length of the vector of the chromosomes. To achieve higher efficiency, the first step of building population is to generate genes that satisfy all the gene constraints. The second step is to build chromosome using the available genes. The candidate chromosome could be added to the population if and only if it meets the chromosome constraints.

Procedure 1

Step 1: Initialize parameters: index j=1, a population size s and population P={Φ}
Step 2: Randomly generate genes and choose those satisfy gene constraints to genes pools G{}.
Step 3: Generate chromosome p_j by combining the available genes in G{}.
Step 4: Check the chromosome p_j whether satisfy chromosome constraints.
Step 5: If p_j is feasible, go to step 6, else go to step 3,
Step 6: If p_j is new and different from any previous individuals, then P =P+ p_j, j=j+1, else go to step 3,
Step 7: If j > s, then stop; else go to step 2.

4. 4 Evaluation and fitness

Evaluation is applied to evaluate the fitness of the chromosome for the consideration of a give objective function. The objective function value is considered as the evaluation rule of every individual. Sometimes the best and the worst chromosomes will produce almost the same numbers of offspring in the next population, which cause premature convergence. In this case, the effect of natural section is therefore not obvious. The objective function proposed in section 3 would be used as fitness function to evaluate the chromosomes.

4.5 Crossover

Crossover is an extremely important component of GA, which makes it different from all other conventional optimization algorithms. The crossover is done by exchanging the information of two parents to provide a powerful exploration capability. It is executed by exchanging the same parts of gene-codes between two individuals. In the proposed GA two-point crossover is used.

4.6 Mutation

By modifying one or more gene values of a chromosome, mutation creates new individual to increase the variability of the population. We use inversion and displacement mutation operations. The inversion mutation selects two positions within a chromosome at random and then inverts the substring between these two positions. The displacement mutation selects a sub-string at random and inserts it in a random position.

4. 7 Overall algorithm procedure

GA for the global supply chain configuration problem is outlined below:

Begin
$t \leftarrow 0$;
Initial the population of parents P(0);
Modify P(0);
Evaluate P(0);
While (not termination condition) do
Begin
Recombine P(t) to yield the population of offspring C(t);
Modify the population of C(t);
Evaluate C(t);
Select P(t+1) from P(t) and C(t);
$t \leftarrow t + 1$;
End
End

Where P(t) and C(t) are the population of parents and offspring in current generation t, respectively.

5. Case study

The proposed model and computation approach has been applied on a world-leading mobile phone company. It is a multi-national company located in Germany and serves three major markets: Europe, America, and East Asia. We would like to use this model to help the company to investigate the impact of new customer zone on its network structure and product family combination.

We implement our proposed GA solution by using visual C language. For comparison, we also use AIMMS linear programming software to get optimal solution for relatively small size problem. To see the efficiency and the effectiveness of the solution, we design several scenarios for the GRAP to test the algorithm. Each scenario is different in the structure of network and the size of product family. They are tested by three numerical experiments with different initial parameters on population size and generation period. The configuration of scenarios is listed in Table 1.

Table 1: The configuration of GRAP scenario

Scenario	Product Family	suppliers	Plants	Distribution Centers	Customer Zone
I	2	4	3	3	4
II	3	8	5	6	8
III	4	10	8	10	6
IV	4	15	10	12	10

The first scenario is a small size problem. It contains four suppliers, three plants, three DCs and four CZs. The basic constraints on demand, capacity, production and transportation cost are listed in Table 2.

Table 2: The capacity and set-up cost for supply chain

Entity \ Attribute	1	2	3	4
Supplier Capacity	700	1000	850	900
Plant Capacity	900	1150	1380	/
DC Throughput	1250	1080	1050	/
Plant Set-up Cost	2500	2200	2800	/
DC Set-up Cost	1200	1800	1000	/
Customer Demand	580	430	460	350

For the scenario I, we set the initial parameters as population size 30, crossover rate 0.9 and mutation rate 0.01. The result of scenario is shown in Fig 3. It can be seen that the best solution is reached by opening three suppliers, two plants and two DCs.

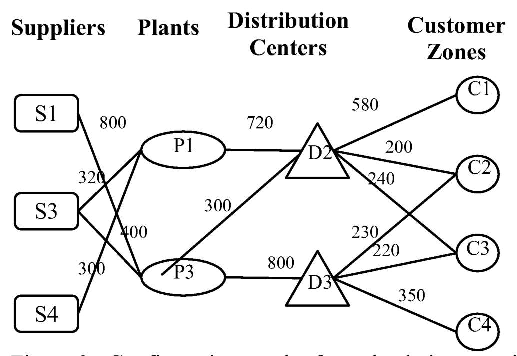


Figure 3: Configuration result of supply chain scenario I

Next, we also have the numerical experiments for relatively larger size problems as in the scenario II, III and IV. We developed all capacities and demands in considering all constraint in the model. The comparison of GA with the optimization package (AMISS) is also provided in Table 4.

Table 4: Optimal solution for test scenario

Scenario	AMISS			GA		
	Time	Objective	Memory	Time	Objective	Memory
I	42	45950	80	218	46275	55
II	367	47460	235	292	47935	92
III	673	49680	384	404	50148	120
IV	/	/	/	485	51842	144

The overall solution time of GA is within 500s and optimization package shows the exponentially increasing solution time with respect to the problem size. The solution quality measured by the percentage gap (GA solution value-optimal objective value/optimal objective value) is about 1%. It is shown in the above test that, for small size problem, our proposed method can give the optimal solution in all of the time and has better average fitness. For relatively large-scale problems, our proposed algorithm can give better heuristic solution. It is also shown that the approach has better computational time and memory required for computation. It is obvious that the GA approach can obtain near-optimal values and when the problem larger, it still can provide reasonable solution. However, LP method cannot handle much large-scale GRAP problem.

6. Conclusion

A novel GA approach to solve the GRAP incorporating multi-echelon process and production family balancing is proposed. Both simulation and comparison results show that this scheme is effective as well as efficient. The experimental results show that our algorithm not only can give better heuristic in most of the scenarios, but also has better performance in the sense of computation time and required memory. It can be used to solve highly complicated and non-linear functions of a realistic production problem. In this paper, a large scale GRAP problem with a production family is illustrated, and a near-optimal solution is reached in a relatively short time. The major benefit of the existing technique is to provide a quick response to changing market requirement, modify manufacturing capacities and production to meet the production need and fulfill the needs of customers.

7. References

Ford, O.D. 1990. *Understanding Business Markets: Interactions, Relationships and Networks*. Academic Press, San Diego, CA.

Gaonkar, R. and Viswanadham, N. 2001. "Collaboration and information sharing in global contract manufacturing networks." *IEEE-ASME Transactions on Mechatronics* 6, NO.4, 366-376.

Pontrandolfo, P., Gosavi, A., Okogbaa, O.G. and Das, T.K. 2002. "Global supply chain management: a reinforcement learning approach." *International Journal of Production Research*, 40, NO. 1,.1299-1317.

Shobrys, D.E. and White, C.D. 2000. "Planning, Schedulingand Control Systems: Why Can They Not Work Together." *Computer Chemistry Engineering*.24, No.2, 163-178.

PRACTICAL APPLICATIONS AND EXPERIENCES

THE "V" OPEN SYSTEM DEVELOPMENT MODEL: AN INTEGRATIVE CONCURRENT ENGINEERING DEVELOPMENT STRATEGY

Cyrus Azani, D.Sc.,
Open Systems Joint Task Force/Northrop Grumman Mission Systems
1851 South Bell Street Suite 102
Arlington, VA 22202 USA
Cyrus.azani.ctr@osd.mil

KEY WORDS: Open Systems, Open Architecture, Modular Open Systems Approach, The "V" Open Systems Development Model, Concurrent Engineering, Systems Development,

ABSTRACT

Countering the unknown, but increasingly sophisticated competitive strategies of the 21st century in an environment of rapidly changing technologies, complex interrelationships, and cut-throat competition requires non-traditional development approaches. When a new system is conceived, an organization must be able to (1) Plug & Integrate: bring together and configure the systems/products needed, and (2) Plug & Play: quickly assemble and interconnect subsystems/components to meet the immediate and emerging needs, and (3) leverage commercial technology and practices in real time. Having the flexibility to configure and integrate systems demands a totally new systems development strategy. The current system development strategies employed by organizations have to a large extent lead to development of closed systems that act as end to themselves and use proprietary technologies, products, processes, and standards that are unaffordable and obsolete. The reality is that system development is no longer a static endeavor based on time-frozen requirements. It is an ongoing process that, in most cases, must be based on an evolutionary development and deployment strategy to avoid building systems that become obsolete before they are even deployed. We also need to refrain from concentrating on system engineering concepts as end into themselves and be able to capitalize on latest change management principles in building and upgrading systems. What is needed is superior development strategy based on open systems and integrated product and process teaming concepts that will enable an organization to rapidly adapt to changes in requirements and technology and will lower the development as well as support and sustainment costs.

In this paper, the author will (1) explain and discuss the Modular Open Systems Approach (MOSA), (2) propose the integrated "V" Process as an effective concurrent engineering strategy for developing open systems, and (3) offer a number of principles to guide the application of the proposed methodology. By following the "V" Open Systems Developement Model the system developers and integrators should be able to successfully take on new challenges and effectively address the multitude of issues related to planning, development, deployment, and continuing evolvement of complex systems. The proposed methodology is simple and easy to follow and understand, and will allow for maximum use of creativity. The proposed system development methodology will also lay the groundwork for building an evolving and flexible open architecture based on modular design and widely-used commercial interface standards. Such architectures will be sustainable, cost effective, and technology/vendor neutral.

INTRODUCTION

A. The Open System Concept

The system architectures used by a great majority of organizations lack robustness and versatility required for 21st century operations and competition. They have been developed as stove-piped structures with minimum adherence to modular design tenets and employ interfaces defined by proprietary standards. Lack of adequate attention to modular design tenets and absence of widely-used and consensus-based standards have created complex development, integration, and sustainment challenges and multifaceted issues such as:

- Rising support and maintenance costs of systems.
- Inability to sustain the system capability at or

above requirements specified.

- Lack of accurate and timely estimates of current and projected component, subsystem, and system capability and life cycle costs in spite of evolving needs and technologies.
- Inability to predict or detect degradation with sufficient lead-time to mitigate the risks associated with a sole source of supply and technological obsolescence before impacts to system functions are realized.
- Difficulty to mitigate risks to the overall system functionality and performance by replacing degraded modules before system degradation.
- Excessive costs and difficulty to integrate legacy systems capabilities with those of newly developed systems to create interoperable system of systems capabilities greater than the sum of capabilities of constituent parts.
- Absence of viable business case analysis and effective strategies to develop open systems and migrate closed legacies into open systems to meet challenges related to evolving needs, technological change, and diminishing supplier availability.

These challenges can effectively be addressed by implementation of an open systems approach to development. It is apparent that the traditional system development strategies may no longer be appropriate tools for developing dynamic and evolving applications and systems. These strategies assume that all the requirements for a new system are known in the beginning of the development process and can be frozen in time or are less likely to change. The traditional approaches also assume that various technologies used for constructing systems are static and are subject to minor changes. Moreover, the traditional development strategies are slow, do not emphasize collaboration among various stakeholders, and result in excessive total system life cycle costs (Azani and Khorramshahgol 2005a).

There are numerous definitions of open systems. These definitions vary within disciplines, industries, and organizations. But there are some common themes that most of these definitions contain. (Roark and Kiczuk 1995). Open systems are systems with permeable boundaries or well-defined standardized interfaces that enable exchange of energy (e.g., electrical current via a wall plug), material (e.g., replacement of components/parts with equivalent components from competitive sources) and information (e.g., interoperability with other systems) with the environment (Azani 2001). Generally speaking, open systems are defined as systems that employ modular design, use widely supported and consensus based standards for their key interfaces, and have been subjected to successful validation and verification tests to ensure the openness of key interfaces (Azani and Flowers 2005).

An open system development strategy leverages other companies' investments and practices to make systems more adaptable to changes in requirements and emerging technologies, and extends the life span of a system. A *closed* system strategy is one that uses privately held, proprietary standards and specifications that are only accessible to a single vendor, or a small group of favored vendors (Azani 2001). By using an open system strategy, systems engineers and information technology managers can use widely available open systems-based products rather than deal with products of proprietary technologies, standards, and specifications. Consequently, system upgrades can be accomplished faster and cheaper because an open architecture facilitates incremental technology insertion rather than large-scale system redesign. Systems will also be more supportable due to continuing access to competitive sources of supply throughout the system life cycle.

An effective system/architecture must apply open interface standards for most if not all the key interfaces used in a system. Key interfaces are common boundaries or connections between modules that are subject to rapid technology turnover and/or increased failure, or have requirement for commonality and/or interoperability (Azani and Khorramshahgol 2005b). When such interfaces are defined by open standards, they permit: (1) quick and effective communication among systems and subsystems comprising a system; (2) rapid configuration and reconfiguration of systems; (3) affordable development and supportability through continuous access to multiple sources of supply; and, (4) increased adaptability to upgrade and change as new technologies become available or requirements evolve. Figure 1 depicts the distinction between key and non-key interfaces within a system of systems construct.

B. The Development Strategies

The best means for effective implementation of an open system is implementation of a Modular Open Systems Approach (MOSA) process. MOSA is an

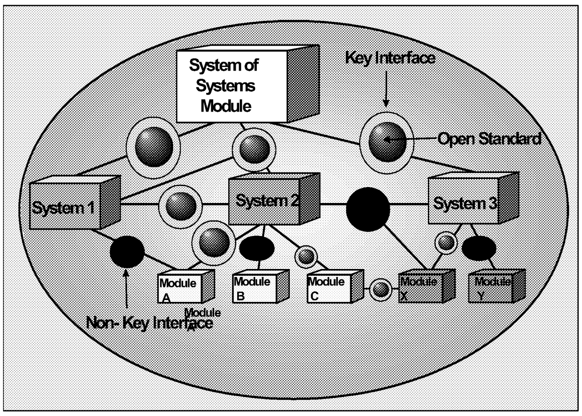


Figure 1: Key versus Non-key Interfaces

integrated business and technical strategy that employs a modular design and, where appropriate, defines key interfaces using widely supported, consensus-based standards that are published and maintained by a recognized industry standards organization (Azani and Flowers 2005). The MOSA process employs three major steps depicted in Figure 2. It first assesses the appropriateness of pursuing an open systems development strategy. Next, it utilizes five principles to develop an open systems architecture. Finally it verifies and validates the implementation progress.

The MOSA Process could be implemented via a traditional water fall model, a spiral development approach, or a “V” process model. The waterfall model introduced by Royce (1970), and initially used for software development. Waterfall strategy is a single-step-to-full-capability strategy. It is a simple sequential model in that there are clean phases which do not repeat processes carried out in previous phases. They are distinct in time. The outputs of one phase fall clearly into the input of the next. The waterfall model has been extended to other development efforts in various industries such as biotechnology. It is also being used in information systems, such as network design, and process reengineering efforts. The traditional Waterfall model does not allow concurrent (parallel) development and testing.

The spiral development model was originally developed by Barry Boehm (1986). This is commonly known as the spiral model because each development cycle progressively approaches some ideal solution. The spiral build-test-fix-test-deploy process is repeated several time till the product/system is fully developed. By cycling through the requirements, specifications, design, coding, and implementation to progressively refine the solution, the risks can be evaluated, quantified, and mitigated. The spiral development strategy reduces the development risk by incremental development, prototyping and testing, and then refining the design for improvement.

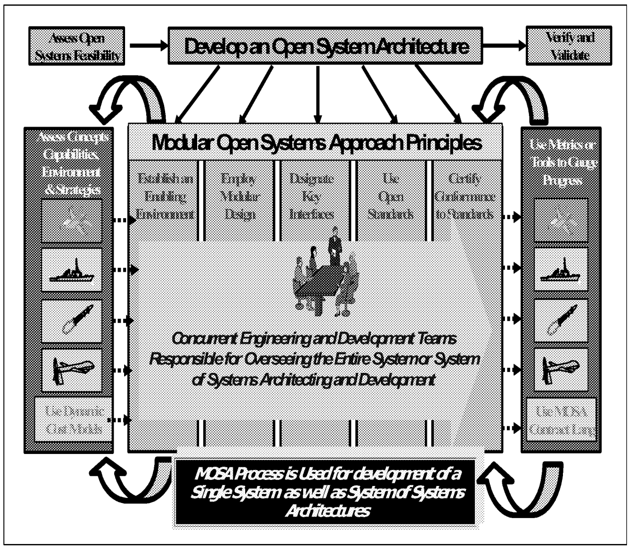


Figure 2: The Modular Open Systems Process

The “V” Process model was originally introduced by Paul Rook (1986) and is mainly used in developments which must be heavily decomposed for design and which is complex. In such cases, it is often necessary to check that the outputs of a stage not only satisfy the specifications of its inputs, but that those outputs meet the requirements in some way of the real world application. The “V” model allows parallel development and testing. The left-hand side of the “V” represents refinement of specification and the right-hand side represents concurrent testing. The advantage of this model lies in continuing feedback that exists between processes in both arms of the “V” because the correctness of each step is verified before proceeding to the next step.

THE “V” Open Systems Development Model

Figure 3 shows the principal steps for implementation of the MOSA process within a “V” development context. These phases/steps have been configured and constructed based on sound systems engineering, consulting and business experiences of the author, and examination of reasons for development project failures. The proposed Open System (OS) “V” Process is applicable to new as well as legacy systems and should be implemented by an integrated product/process development team hereby referred to as Concurrent Engineering and Development (CED) team. The CED team must include all of the stakeholders involved in the acquisition and employment of the system. It should at a minimum include those who generate requirements, design, build, test, operate and maintain the system for its lifetime. The actual make up of a particular team is the responsibility of the organization executives.

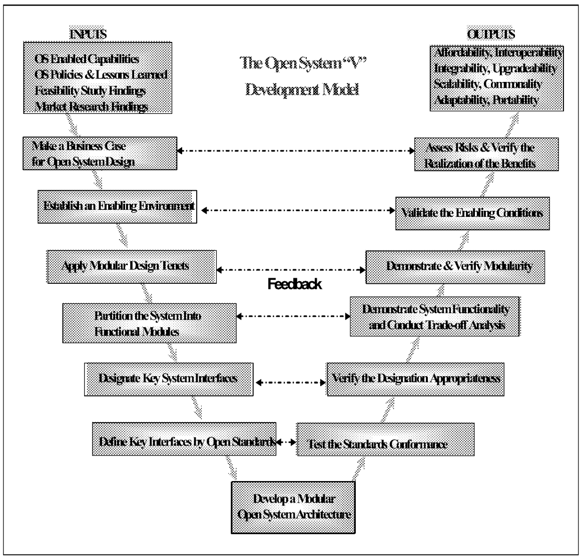


Figure 3: The Open Systems "V" Process Model

A. Model Inputs

The main required inputs that the model needs for initiation are a list of specific capabilities or requirements that could effectively enabled by following an open systems strategy, lessons learned from previous open systems development projects, findings of previous feasibility studies and market research/analysis, and organizational policies that govern consideration and implementation of open systems. The market research findings should be aimed at identifying consensus or de facto interface standards and commercial or non-developmental products built to these standards. The lessons learned from previous development projects and previous business case analysis conducted would assist in performing a better feasibility analysis and reducing the overall risk of developing an open system. Following requirements or capabilities may create suitable conditions for use of an adaptable open architecture :

a. When the organization experiences a high rate of growth and market expansion.
b. When the technology is not proven and is rapidly becoming obsolete.
c. When we need capability to quickly reconfigure systems in response to emerging needs.
d. When seamless and high-speed digital information exchange among diverse entities/elements is essential.
e. When business strategies call for diversification and quick modernization of legacy systems.
f. When the requirements are defined in terms that enable and encourage the use of commercial and non-developmental item equipment. Such requirements may address relative maturity, integration and modification requirements for non-developmental items.
g. When there is need for commonality of hardware, software, and support systems; continuing access to commercial technology/products from multiple suppliers; and simplified cost effective maintenance.
h. When requirements call for application of modular, reusable, portable, extensible, and non-proprietary hardware or software.

B. Make a Business Case for an Open System Design

Generally speaking, the user needs are expressed in requirement documents as capabilities rather than technical solutions and specifications. Such needs are then explored by a series of competitive, parallel, short-term concept studies focused on defining and evaluating the feasibility of alternative concepts and to provide a basis for assessing the relative merits (i.e., advantages and disadvantages, degree of risk, etc.) of these concepts. The overarching goal of using MOSA is to develop the best overall value solution over the system's life cycle that meets the user's and customers operational requirements. The fulfillment of this goal is dependent upon an effective assessment of the feasibility of using widely-supported commercial interface standards for the proposed system. Making a business case for use of open systems via a dynamic cost-estimating model will facilitate such feasibility (Azani and Khrramshahgol 2005a). Cost models used must consider differences between alternative open and closed systems.

Development of an open system is not risk/cost free. Consequently, the systems engineers and development project managers should not blindly follow the "OS "V" model and develop an open system. That is why the corresponding verification and validation aspect of this phase (the right-hand side of the "V") should be focused on risk assessment of a closed system alternative, application of dynamic cost models to estimate the cost savings resulting from an open system implementation, and final verification to ensure that the open systems benefits have been realized.

C. Establish an Enabling Environment

Effective implementation of the "V" OS Development Model is dependent upon establishment of a supportive environment. Such environment is characterized by supportive strategies such as technology assessment, technology transfer and development, technology insertion and integration, and technology support and sustainment strategies. For example, support and sustainment strategy should address non-developmental and/or commercial items whose design is not controlled by the organization, and the performance and interface specifications of which may be unilaterally changed by the suppliers. Employment of a standardized systems engineering process (e.g., IEEE 1220), adequate MOSA concept and "V" Process implementation training and experience, allocating adequate resources to the implementation of the open systems development process, and appointing an executive as a MOSA advocate are among other enabling factors needed for effective implementation of the "V" OS Development Model.

The corresponding verification and validation aspect of this step (the right-hand side of the "V") should be focused on preparation of a checklist to verify/validate that all the enabling factors and strategies have been taken into consideration and are in place before the initiation of the development and implementation.

D. Apply Modular Design Tenets

Modular design is a design where functionality is partitioned into discrete, cohesive, and self-contained units with well-defined interfaces that permit substitution of such units with similar components or products from alternate sources with minimum impact on existing units (Azani and Khorramshahgol 2005c). To design a system for change, the programs must adhere to four major modular design tenets identified in Figure 4. These tenets determine the degree to which modules are cohesive (contain well-focused and well-defined functionality); encapsulated (hide the internal workings of a module's behavior and its data); self-contained (do not constrain other modules); and highly binded (use broad modular definitions to enable commonality and reuse). By following these tenets, each module will be designed for change and the interface to each module is defined in such a way as to reveal as little as possible about its inner workings which facilitate the standardization of modular interfaces.

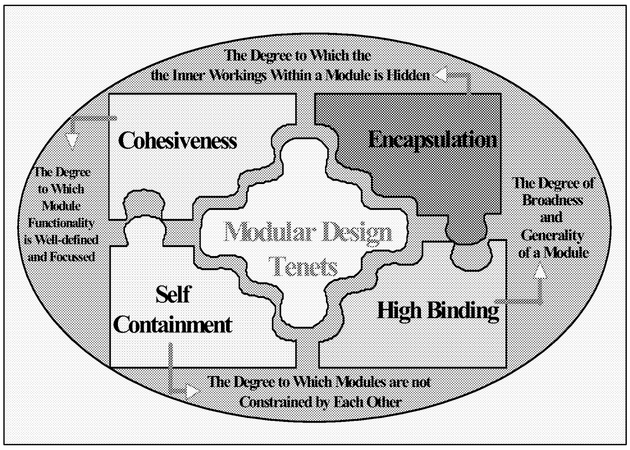


Figure 4: The Modular Design Tenets

The corresponding verification and validation aspect of this step (the right-hand side of the "V") should be focused on verifying that the designed system and its components are indeed encapsulated, decoupled, and self-contained. Modularity could be demonstrated by prototyping and by replacing the existing modules/ component with equivalent modules/component from competitive sources to ensure that the system complies with modular design tenets and could still functions without major redesign.

E. Partition the System Into Functional Modules

To undertake this step, we need to use the systems engineering process with greater reliance on interface control to support functional partitioning and modular design. We need to decompose the system/mission requirements into a small number of system building blocks using reference models and interface management to provide direction for the process (Azani 2001b). We also need to prototype the system, subsystems, and components to demonstrate the integration of the system using proposed interface standards/definitions. We also use prototypes to demonstrate standards and standards-compliant products. We need to demonstrate that potential interface standards and specifications will achieve the required system performance.

The corresponding verification and validation aspect of this step (the right-hand side of the "V") should be focused on demonstrating system functionality through modeling and simulation to display system capabilities, and through conducting trade-off analysis, prototyping, and reconfiguration of functions and modules.

F. Designate Key Interfaces

Early in the system definition effort, all key interfaces should be identified and well-defined. As mentioned earlier, key system interfaces include interfaces where the technology turnover is rapid on one or both sides of the interface, design risk is high on either side of the interface, and the system elements on one or both sides of the interface exhibit a high failure rate or are very expensive. We need to review and revise, if necessary, key interface identification and definitions. We also need to determine if there is an existing reference model we can use, or if we need to develop an appropriate one for the concept(s) under consideration. The reference model will be used to:

- Identify rapidly changing technologies applicable to the proposed concepts,
- Identify proposed subsystems which are likely to grow or evolve over each proposed concept's life,
- Identify high lifecycle cost drivers,
- Identify interfaces likely to be affected,
- Establish a list of key interfaces, and
- Identify consensus based or de-facto standards that could potentially be used to define/control the key interfaces.

The corresponding verification and validation aspect of this step (the right-hand side of the "V") should be focused on verifying that all the candidate key interfaces have appropriately been designated as such, are well defined, and their performance is properly demonstrated.

G. Define key Interfaces by Open Standards

The preference is always on use of open standards those that are certified by recognized standards organizations and are widely used by the industry first, then de-facto standards, and finally proprietary standards. We need to evaluate the market acceptance and maturity of potential standards to determine if they support needed capabilities and mission requirements.. Selected standards should provide access to non-developmental items and commercial items that are available from multiple sources. We also need to assess the impacts of new technologies to determine likely evolutions of standards. If emerging standards are used, then the organization should consider participating in appropriate standards bodies to ensure standards definitions meet the organizational requirements. We may ask the following questions when selecting standards:

- To what degrees are the candidate standards based on mature technology and are currently used in production?
- Are there current products available, built to the candidate standards?
- Are several vendors complying fully with the candidate standards?
- Do the candidate standards have a defined test suite or conformance criteria to eliminate the time and cost of testing?
- Are the candidate standards produced and maintained by a reputable accredited standards organization?
- Are the extensions to the standards by suppliers likely to make the system dependent on a single source of supply throughout the lifecycle?
- Are the standards selected compatible with one another? Do they overlap?

After the standards are selected we need to prepare system profiles. A system profile contains all of the individual standards, and interface requirements selected for a system. The profile also lists for every interface standard, the tailoring instructions and the selected optional extensions if applicable. Although the system profile references the interface standards at the desired levels of openness, the interfaces below the specified level may also be designated by open standards if the provider so desires. The level chosen to define the open interfaces should be at a level supported by the industry and consistent with supportability planning. For example in digital electronics, that level may be the line replaceable (LRU) and shop replaceable (SRU) level. We need to establish initial level of openness objectives consistent with interoperability, integrability, upgradeability, affordability objectives, and industry practice (based on market research).

The corresponding verification and validation aspect of this step (the right-hand side of the "V") should be focused on demonstrating system openness using modeling and simulation and validating the openness of subsystems, and/or components at the levels specified. Following are examples of means for fulfilling this objective:

- If contractors or consultants used, require contractors/consultants to certify that they have used widely accepted consensus standards for key interfaces at and above the specified levels of openness.

- Performing conformance and compatibility testing at and above the specified levels of openness.

- Asking independent test facilities to verify the use of open standards at and above the specified levels of openness.

- Developing innovative means such as indices e.g., a color-coded scheme, percentage, or numerical value to determine the extent of openness of system/subsystems.
- Ensuring that extensions to the standards by suppliers will not make the system dependent on a single source of supply throughout the lifecycle.

- Ensuring standards selected are compatible with one another.

- Demonstrating that the selected NDI or COTS products meet operational environment requirements, and the selected NDI or COTS products conform to the selected interface standards.

- Prototype systems, subsystems, and components to ensure their openness.

H. Develop an Open Systems Architecture

At this stage of the process, all the modules will be connected together via open interfaces to create an adaptable and flexible modular structure or architecture for the system. Architecture means different things to different people. Some definitions are complex, confusing, and depicted in very long paragraphs. For example those in Software Architecture for Product Families: Principles and Practice (Jazayeri et.al. 2000), or the one proposed in the UML Modeling Language User Guide (Booch et al. 1999). Other definitions such as the ones proposed by Zachman (1993) and Rechtin and Maier (1997) are simpler and more concise definitions of architecture. The best definition is perhaps the definition by the Institute of Electrical and Electronics Engineers that defines architecture as the structure of components, their relationships, and the principles and guidelines governing their design and evolution.

Generally speaking, there seems to be a consensus that the architecture is the structure not only of the system, but of its functions, the environment in which it will live, and the process by which it will be built and operated (Rechtin 1991). It is the highest-level conception of a system in its environment and includes the structure and behavior of the whole system in that environment, to show how the system will meet its requirements. (Emery et al. 1996). Additionally, an architecture is believed to be a base for reasoning about a system and goes through transformations, not decomposition, and uses series of representations that have different motivations, uses, representations, semantics, etc (Zachman 1993).

An open architecture depicts the structure of functional and physical modules, their interrelationships through open key interfaces, and the principles governing the design and evolution of such structure in a flexible manner. Different organizations operate at different levels of open systems maturity which entails different level of maturity for their architecture (Azani 2002). Open architectures rely on physical modularity and functional partitioning of both hardware and software to create the flexibility needed for replacing specific subsystems and component without affecting others. The open architecture supports the functional baseline and system specifications and is an effective blue print for developing and maintaining affordable and adaptable applications and systems. By developing an open architecture, the system integrators/architects will build flexibility into systems and will achieve enduring interoperability, integrability, affordability, adaptability, and supportability.

THE BENEFITS OF THE PROPOSED MODEL

Following are some examples of the benefits expected from implementation of an "V" OS Development Model.

Concurrent Development and Testing. The proposed model is a concurrent engineering and development model as it integrates development with testing and takes into consideration the overall life cycle issues of developing a system.This advantage is realized through the application of the "V" model itself.

Enduring Interoperability. Effective communication between subsystems in a platform and among platforms is one of the most important benefits realized by an "V" OS Development Model implementation.

Integrability. Systems and subsystems that can be tightly plugged in from disparate elements must, by definition, have the attribute of integrability. By following the proposed model, subsystems and systems will more easily be integrated together. Integrability is the ability to quickly interconnect and assemble systems, subsystems, and components in order to construct new ones in a synergistic and affordable way as needed. Common components, standardized interfaces that comport to those used in the commercial world, and wide availability of parts used in a system are the necessary elements of integrability.

Affordability. Affordability has always been close to the top of executives' priorities lists, but with current process and projected budget constraints its importance has increased. The "V" OS Development Model should be stressed for its ability to reduce the cost of systems not just in initial acquisition, but over the entire life cycle. Elements of open systems that impact affordability include economies of scale from commonality and reuse, flexibility to accept components from a wide variety of COTS sources, and competitive sourcing.

Adaptability. By following the "V" OS Development Model, a system will be designed in such a manner that can adapt to new requirements, systems, components and technology with minimal effort and cost. The following attributes are among the integral characteristics of adaptable systems:

1. Obsolescence mitigation. Obsolescence of both systems and components has long been a problem. With technology turnover cycles currently observed, this problem reaches unprecedented proportions. By managing to the natural cycle rates of encompassing technologies, an "V" OS Development Model can keep obsolescence from being a major problem.
2. Evolvability. Because both competitive threats and technology changes at an unprecedented rate a system must have the ability to evolve to new designs, including insertion of new technologies, replacement of obsolete technologies, and the accretion of new capabilities. All of these capabilities will become possible by following the OS "V" Development Model.
3. Structural Flexibility. Designs should be selected for the lack of dependence on fixed partitioning schemes. The fixed backbone of a system should have an adaptable structure to allow variation in partitioning and the specific collection of modules used in a system or system of systems. Use of modular design tenets by the model will enable such flexibility.
4. Design Robustness. Designs should be able to withstand unforeseen variation in the specific modules used, different manufacturers of the same module, and conditions under which the system is used. The end product of the model is an open architecture which by definition is a robust structure.
5. *Intra-system Integrability*. The system should be designed such that it can be easily matched to other systems used in a system of systems context. The proposed development model through parallel development and testing is an effective system of systems development model.
6. *Re-use*. One of the factors that affects the cost and other factors of system design is the degree to which it re-uses architecture, technology and products proven on other systems. Not only should systems be evaluated according to the degree to which they re-use previous technologies, but also the degree to which new or unique technology will be usable on other systems. The emphasis on applying the lessons learned from previous development efforts ensures ample opportunity for commonality and reuse by the proposed model.

Supportability. Through the application of a concurrent engineering process, the "V" OS Development Model can enable performance-based supportability. Issues to consider are the number of sources of supply, the maturity level of standards selected, and the projected support life of the system. Supportability should be given greater weight when programs are evaluated. Supportability requires consideration of the long-term impact, and uncovering possible unintended consequences, of decisions made early in the program.

THE GUIDING PRINCIPLES FOR EFFECTIVE IMPLEMENTATION OF THE MODEL

Achieving the benefits of concurrent engineering and development and effective application of the "V" OS Development Model will be contingent upon adherence to the following guiding principles:

1. Provide incentives and appoint open systems champion

The incentives offered must foster the cultural change necessary to realize the key open systems capabilities.

2. Involve the stakeholders in planning and implementation of the "V" OS Development Model

Involvement breeds acceptance. We must ensure that those responsible for the "V" OS Development Model planning and implementation, as well as those being impacted by such capability or lack of it participate in the decision making process and take ownership for the ensuing changes.

3. Plan for organizational and cultural changes needed for implementing the Process

The realization of the benefits of open systems requires changes in acquisition strategy, process, and program management. The acquisition process must concentrate on interrelationships among systems being acquired rather than on acquiring systems in isolation from each other. The acquisition process emphasis should be shifted from:

- consideration of short term cost of system development to total costs of ownership of systems capable to plug & play and plug & integrate;
- requirements frozen early in the acquisition cycle of a system to flexible and evolving plug & integrate and plug & play capabilities for system of systems;
- unaffordable supportability to Just in Time (JIT) supportability enabled through modularity, economies of scale and continuous access to multiple suppliers; and to
- technologies available during the early phases of the acquisition cycle to the ability to insert new technology, as it becomes available.

4. ***Create a fully synergistic partnership among departments, suppliers, original equipment manufacturers, distributors, contractors, and system developers***

Realization of the open systems benefits also demands a full partnership among various stakeholders. The implementation of an "V" OS Development Model and realization of open systems capabilities require constant change, and unless change becomes indoctrinated within the organization culture such capabilities will not be realized. Indoctrination will become a reality through involvement and commitment of the people responsible for realization of the OS Process benefits.

5. Ensure a Widespread use of Open Standards

The architectural flexibility required for *plug & integration and plug & play* is maintained through the application of open standards (i.e., widely used consensus-based international standards) for selected interfaces within a system. However, open systems benefits evaporate when the standards selected are not in widespread use by the commercial industry. The benefits of openness are realized when there are a large number of vendors offering a wide range of products compliant with open standards. Thus, key concepts of openness must be explained before the benefits of the OS Process based on "V" model be fully realized.

6. Develop explicit risk management strategies

We must also develop specific risk mitigation strategies to minimize the likelihood of occurrence and effectively deal with the risks associated with the implementation of the OS process. Some of the risks associated are:

- Resistance to change
- Creating more bureaucracy through additional organizational levels or reporting requirements.
- Fear of failure due to betting on wrong standards and forecasting errors related to estimated cost, benefits, or performance of a plug & integration and plug & play capability.
- Loss of control because the organization is no longer the driving force behind technological change
- Lack of adequate resources to fulfill the goals.

7. Enable partitioning and modularity with minimum effort and cost

The open systems benefits and capabilities can best be realized by suitable modular partitioning of the systems. If the interfaces are controlled at an appropriate level, then modularity consistent with the best economic and engineering requirements can be achieved.

8. *Use practical level of openness.*

Mindlessly adhering to an open systems concept simply because it is what is required should be discouraged. A closed system, selected after due and diligent deliberation may be preferred to a perfectly open system that will not do the job.

9. Be attentive of the maturity of standards.

Standards require relatively long periods to mature, and then after reaching maturity have a relatively short half-life. Some attention must be given to the nature of the standards selected. One that is too mature, despite being widely supported, may well disappear under the onslaught of superior new technologies. On the other hand, juvenile standards may not achieve sufficient support to merit consideration. However, it must also be communicated that conscientiously selecting a youthful standard, and being wrong, is still usually cheaper and quicker than creating a military unique standard.

10. Verify progress on a continuing basis.

The program managers and other acquisition authorities must ensure that the system design and support conform to the premises of the "V" OS Development Model. The ability to accomplish this goal will require training and education of the personnel working on the program. Objective metrics must be provided for evaluating the program's adherence to MOSA principles and the guiding principles discussed. Some examples of metrics are:

- *Obsolescence index***.** The percentage of obsolete products, components, subsystems, etc. (measured in replacement pain, e.g., schedule and cost) in a system at a specific period of time (e.g. when the system is fielded or being designed). If the obsolescence index is greater than a specified threshold value, then the system may not be sufficiently open to meet evolving program objectives and may be more susceptible to obsolescence.

- *Affordability ratio.* The projected total ownership cost (TOC) of the proposed open system over the projected total ownership cost when the system is designed with closed interfaces. If the ratio is equal or less than one then make the system more affordable by making it open.

$$\rho_{\mathrm{Aff}} = \frac{\mathrm{TOC\ (open)}}{\mathrm{TOC\ (closed)}}$$

- *The degree of openness.* The percent, ratio, or number of key interfaces in a system/subsystem that are supported by consensus based and/or *de facto* standards, adjusted for the system's/subsystem's level of openness.

- *Technology insertion.* Number of latest technologies successfully migrated, at a specific period of time (e.g. when the system is fielded or being designed), to a program or an organization as a result of open architectures.

CONCLUSION

Through harnessing the benefits of the "V" OS Development Model implementation, organizations will be empowered to effectively:

- Reduce the total ownership costs of systems through economies of scale.
- Plug and play based on standards-based and supplier neutral architectures.
- Upgrade systems components with minimum disruption in readiness as soon as technology changes and commercial products become available.
- Repair and maintain systems and components more easily and efficiently through application of Commercial Off The Shelf products that are readily available in the market place.
- Balance performance and affordability. As a result of using plug & play-based systems the need to sacrifice performance for affordability and/or supportability will be reduced.
- Build a more synergistic partnership among stakeholders through concurrent development, testing, and production. Such partnership will require noticeable leveraging of modern management practices exercised by other organizations and cooperation on setting standards.

Finally, the "V" OS Development Model is supportive of an organization move into the mainstream of change from the unique proprietary and rigid environment of the past. Such move is necessary in order to leverage the commercial investments in other countries and industries and quickly upgrade systems as needs and requirements change. Furthermore, The "V" OS Development Model is also compatible with the standardization and modular interchangeable parts movements that begun during the industrial revolution. These movements were to a large extent responsible for making products more affordable and reliable for the public at large. By implementing the "V" OS Development Model, an organization would also be able to quickly capitalize on the opportunities brought about by these movements, reduce its total cost of doing business, and develop systems more quickly and effectively

REFERENCES

Azani C.H., and Flowers, K. 2005. "Integrating Business and Engineering Strategy Through Modular Open Systems Approach," Defense AT&L Journal; pp. 37-40.

Azani C.H., and Khorramshahgol, R. 2005a. "A Methodology for Developing Viable And Cost Effective Customer Relationship Management Systems," Journal of Knowledge Management Practice, (August)

Azani C.H., and Khorramshahgol, R. 2005b. "Management of Technology Using a Modular Open System Strategy," Proceedings of the 12th European Concurrent Engineering Conference, Toulouse, France; pp. 5-12

Azani C.H. and Khorramshahgol R. 2005c. "The Open System Strategy: An Integrative Business and Engineering Approach for Building Advanced Complex Systems", Proceedings, the 9th World Multiconference on Systemics, Cybernetics and Informatics (WMSCI 2005), Orlando, Florida (July 10-13).

Azani C.H. 2002. "Enabling System-of-Systems Capabilities Via Modular Open Systems Maturity Models." Proceedings of the 2nd Annual CMMI Technology Conference, Denver, Colorado (11-14 November 2002)

Azani, C.H. 2001a. "The Test And Evaluation Challenges of Following an Open System Strategy," The ITEA Journal of Test and Evaluation, September/October 2001,Volume 22, No. 3

Azani, C. H. 2001b. "Joint Space Operations via Secured Integrated Network of Modular Open Architectures." Proceedings of the Joint Aerospace Weapons Systems Support, Sensors, and Simulation Symposium and Exhibition, 23-27 July 2001, San Diego, California.

Azani, Cyrus H. 2000. "The Open Systems Strategy: A Viable Business and Engineering Approach for Building And Sustaining Advanced Complex Systems." Proceedings of the Defense Manufacturing Conference, November 26-28, 2000, Tampa Florida

Boehm B.W. 1986. "A Spiral Model for Software Development and Enhancement" IEEE Computer, Vol. 21, No. 5, pp. 61-72 (May)

Booch, Rumbaugh, and Jacobson 1999. The UML Modeling Language User Guide, Addison-Wesley.

Emery D.E. and Hilliard R.F. and Rice T.B. 1996. "Experiences Applying a Practical Architectural Method" Proceedings of the 1996 Ada-Europe International Conference on Reliable Software Technologies pp. 471-484

Jazayeri M. and Ran A. and Linden F. 2000. Software Architecture for Product Families: Principles and Practice Addison Wesley Longman.

Rechtin E. and Maier M.W. 1997, The Art of System Architecting, CRC Press.

Roark C., and Kiczuk B. 1995. "Open System Avionics Architectures," IEEE Aerospace and Electronic Systems Magazine, Vol 10, No 9, (Sep)

Rook P. 1986 "Controlling Software Projects", Software Engineering Journal, Vol. 1, No. 1, pp 7-16 (January).

Royce W.W. 1970. "Managing the Development of Large Software Systems" IEEE WESCON Conference" pp. 1-9 (August).

Zachman J.A. 1993. "Managing Architectures" lecture delivered at Internal Revenue Service Headquarters, Washington D.C. (November)

ESA Open Concurrent Design Server

Arne Matthyssen, Massimo Bandecchi
Concurrent Engineering section,
ESA/ESTEC
TEC-SYE, P.O. Box 299, 2200 AG
Noordwijk NL
Tel. +31(0)71.565.3986
Fax. +31(0)71.565.6024
Email: Arne.Matthyssen@esa.int
Email: Massimo.Bandecchi@esa.int
www.esa.int/cdf

Petter Myrvang, Narve Mjos
Det Norske Veritas
NO-1322 Hovik
Norway
Tel: +47 67 57 99 00
Email: Petter.Myrvang@dnv.com
Email: Narve.Mjos@dnv.com
www.dnv.com

Kjell A.Bengtsson
EPM Technology AS
P.O Box 6629 Etterstad
N-0607 Oslo
Norway
Tel: + 47 23 17 17 17
Fax: + 47 23 17 17 01
Email: Kjell.Bengtsson@epmtech.jotne.com
www.epmtech.jotne.com

KEYWORDS

ESA, Concurrent Design Facility, Integrated Design Model, Open Standards, ISO, ECSS, Data Exchange, Distributed, Space System Engineering.

ABSTRACT

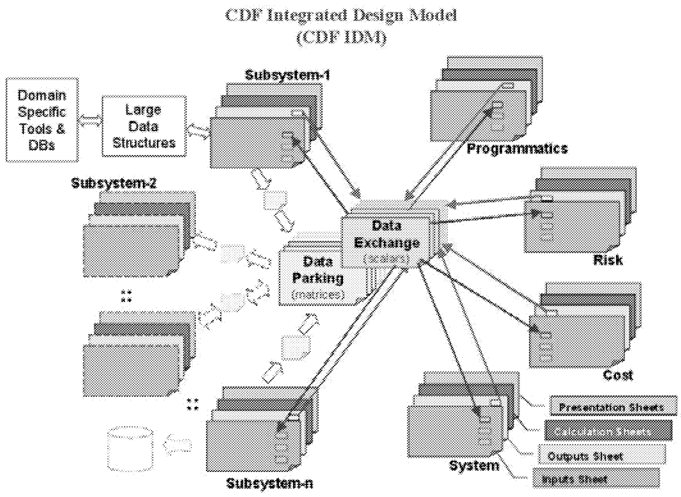
This paper describes a project from the European Space Agency for the creation of a concurrent design server, the Open Concurrent Design Server (OCDS), using open data exchange standards.
Based on the ESA's Concurrent Design Facility (CDF) methodology and Integrated Design Model (IDM) the OCDS has 2 main objectives a) the creation of a service to enable collaborative, concurrent and distributed engineering of space systems between different organizations, using standard object definitions b) an instantiation of a OCDS product to be distributed to the European Space Community as the first building block to implement their own concurrent design infrastructure.

BACKGROUND

The Concurrent Design Facility (CDF, www.esa.int/CDF) is a state of the art facility created at the European Space Research & Technology Centre (ESTEC) in Noordwijk (NL) with the main purpose of assessing and designing future space missions using modern concurrent engineering technologies and methodologies.

Since 1998 over 50 potential future missions, mostly scientific, but also related to Earth Observation and Human Spaceflight, have been assessed at conceptual design level, known as pre-Phase A/level 0, using the facility. In addition, the CDF infrastructure has been used to perform industrial work reviews, prepare specifications, and coordinate international project work.

The CDF infrastructure is based on the Integrated Design Model (IDM), an in-house developed tool, which allows integration of all the subsystem discipline tools and parameters in a consistent and effective design environment.



Figures 1: ESA CDF Integrated Design Model structure

The ESA CDF IDM has also represented the means to capture the technical knowledge and to document the "engineering views" that are required to run the design processes. The techniques and tools developed are now well tuned and established for the in-house application to the preliminary phases of the space project life-cycle. Furthermore the know-how and the models developed in CDF have gained a lot of interest among institutional partners, academia and industry in the last years.

DISSEMINATION OF KNOW-HOW AND TOOLS

The IDM is highly requested by several ESA Institutional Partners and the nucleus of the IDM (CDF Core-IDM) has already been delivered to CNES (French Space Agency), ASI (Italian Space Agency) and CSA (Canadian Space Agency) at their request. It has been/is being used by these organizations as an initial building block for the creation of their own design facilities. The idea is to develop these centres using a common data structure and representation in order to facilitate future interoperability and interchange, joint project work, and linking of these facilities for real-time cooperative (i.e. concurrent) engineering.

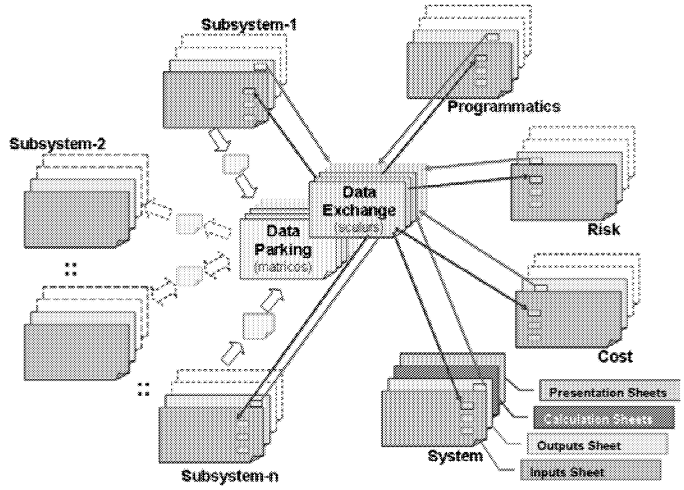


Figure 2: ESA CDF Core-IDM structure

The CDF Core-IDM is a subset, the nucleus, of the CDF IDM, as illustrated in Figure 2.
The CDF Core-IDM consists of:

- Discipline workbooks: a MS Excel® workbooks representing disciplines and/or space system subsystems. The workbooks include input and output sheets and empty calculation and presentation sheets, except for the System workbook. The System workbook does include the ESA CDF IDM calculation and presentation sheets.
- Data exchange: a MS Excel® workbook used as a database to store scalars.
- Data Parking: a MS Excel® workbook used as a database to store matrices.
- All communication features and interfaces to allow exchange of data between above mentioned parts.

A wider distribution of this nucleus and the generalised and standardised use of a common data model/representation could highly contribute to the creation of a global E-collaborative environment for the design and development of space missions. Academia would also benefit of the usage of this tool and related methodology and, eventually, contribute to its improvement.

It is ESA's intention to coordinate and integrate these efforts in Concurrent, Collaborative and Distributed Engineering actually on-going in the European space sector, sharing the experience of CDF with industry and other agencies. In order to achieve this goal ESA has taken the initiative to industrialise the CDF model into a product to be made available to the Space Community. ESA CDF has selected an Open Concurrent Design Server as the most appropriate way for the industrialization and distribution of a common data model, which will eventually fulfil the data sharing requirements in the Space Community.

OPEN CONCURRENT DESIGN SERVER

The OCDS activity will enhance the CDF Integrated Design Model and methodology by the introduction of standard information data models and Reference Data Libraries (RDL).

The activity will start with a review of the current CDF Core-IDM, taking into account the Space Industry's feedback. Followed by the preparation and demonstration of a data model according to open data exchange standards describing the parameters and their function in the space system design process. Initially the OCDS targets the Level 0 and Phase A design activities. Further design phases in the life-cycle will be analysed to ensure the evolution and continuity of the data model where required.

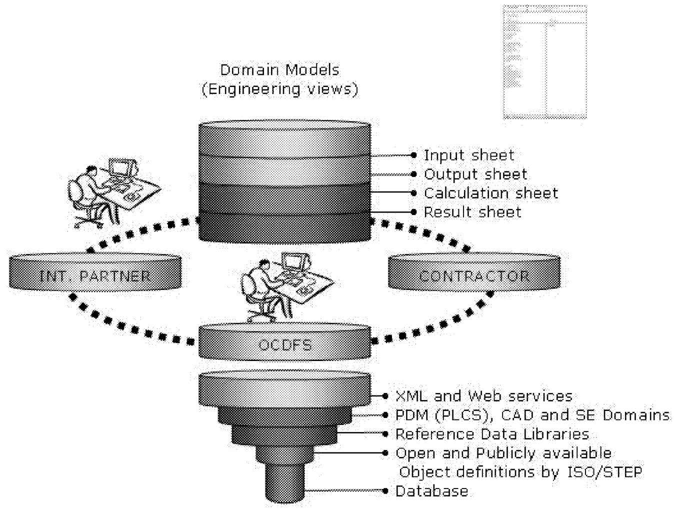


Figure 3: graphical representation of OCDS

Figure 3 shows the collaboration of different parties e.g. international partners and contractors using the OCDS.
The OCDS will include data models and RDLs readily available for the users to download and or to consult. XML, UML and EXPRESS are some of the data representation and exchange formats envisaged.

The OCDS activity also includes the definition, organisation and implementation of the infrastructure to support the delivery, maintenance and redistribution of the OCDS in the form of Community Software (i.e. "Open Source" restricted to the institutional partners of the Agency). One possible implementation of OCDS or a part of OCDS can be via a dedicated website.

STANDARDS

Standards envisaged to be used in OCDS activity are ECSS-E10 Part 7A (European Cooperation for Space Standardization, see [RD 1]), STEP-NRF (see [RD 2]) and ISO 10303 (see [RD 3]) which includes support for XML (Part 28) and an interface with UML (Part 25) and possibly ISO 15926 (see [RD 4]) which defines a complete information architecture that enables the interoperability of software applications (engineering disciplines) for the entire product life-cycle.

An important complementary activity has been undertaken by the ECSS Working Group (WG) in charge of the development of the Level 3 standard E-10-0A1 -

"Engineering design model data exchange (CDF)". The Terms of Reference of this WG include:

- the initiation of a global perspective of the parameter sets required to cover all project design phases,
- the creation of a list of system parameters, definitions and units having a common application for phase 0/A analysis and design reviews,
- the creation of a merged set of parameters, agreed by all Agencies and partner industry and institutes, defining a standard for model based data exchange.

BENEFITS

Other industrial sectors have achieved major quality and productivity improvements through the use of object model technology based on open standards for interoperability. Figure 4 shows a snapshot of a Reference Data Library explorer (browser-like interface incl. data model representation in EXPRESS) made available to the users via a website.

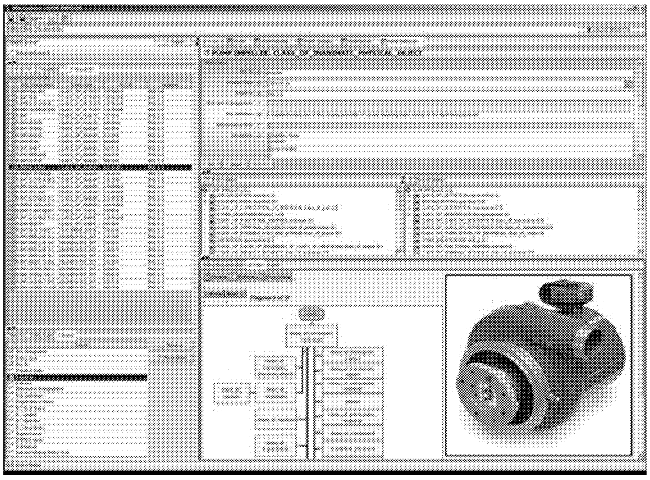


Figure 4: an RDL explorer view.

These methods and technology are now available to the Space Industry in the form of interoperable space information modelling objects. Interoperable object model technology also allows automated standards checking and cost estimating to better control project scope, schedule and cost.

This solution will help the European Space Community to:

- Increase Data Management Capabilities, including Life-cycle Data Management
- Support the information longevity objective
- Achieve a hardware and software independent solution
- Optimise the design process
- Consolidate Design Models in a repository based on open standards
- Improve communication to contractors and partners, using Open and publicly available standards (e.g. ISO, ECSS, ….)
- Streamline the communications to other ESA corporate applications using model based integration
- Connect data to product assurance (PA/QA) activities.

CONCLUSION

The OCDS combines experiences from the application in CDF of concurrent engineering to initial phases of space projects with the data modelling and standardisation used in product data handling and interchange processes.

The OCDS product represents the reference data model for the European Space Community and the first building block to create their concurrent design centres. It also can be seen as a platform for cooperation and an interoperability gateway for engineering data.

The OCDS project is the first milestone for the creation of a collaborative CE environment for the design and development of complex space systems at European level.

REFERENCES

[RD 1] ECSS E-10 Part 7A "Space engineering - System engineering - Part 7: Product data exchange"

[RD 2] STEP-NRF "Network-model and results format", developed by ESA (TEC-MCV).

[RD 3] ISO 10303 "Industrial automation systems and integration — Product data representation and exchange".

[RD 4] ISO 15926 "Industrial automation systems and integration -- Integration of life-cycle data for process plants including oil and gas production facilities"

LATE PAPER

Optimal Selection of the Joining Technology in the Automotive Product Development Process

V.I. Vitanov[1], J.W. Stahl[1], M. Ganser[2]

[1] School of Industrial and Manufacturing Science, Cranfield University, United Kingdom, V.Vitanov@cranfield.ac.uk, JanStahl@compuserve.com
[2] Research & Development Centre Munich, BMW Group, Munich, Germany, Markus.Ganser@bmw.de

Abstract: The correct selection of an optimal combination of joining technologies at the body in white process planning stage is critical for the overall quality and cost of a new car. This paper considers cost analysis of new design concepts in the preliminary phase of the automotive Product Development Process (PDP) followed by further optimal selection of appropriate joining technology. An optimisation approach to facilitate the design process will be discussed. The proposed approach can be useful for design and process planning engineers when selecting different joining technology combinations in the process of designing an automotive manufacturing system. It can lead to significant reduction of the manufacturing costs.

1. Introduction

Optimisation of the manufacturing cost management process plays an ever-increasing role in the global automotive industry. In particular, the development of new designs accounts for a great percentage of the final cost of a car [1]. A significant saving could be achieved if at the end of the design process of a particular subsystem, designers have in their disposal a method to help with the comparison of future manufacturing costs of different design options. Present practice in estimating the cost of new designs follows a lengthy procedure of adding up different costs that can take up to several months and requires input from several different departments. In the same time the real need at this stage is for a method that can contrast different newly developed designs in sense of cost and rank them appropriately according to specified criteria. The optimisation procedure discussed in this paper focuses on an approach that can become a vehicle for solving this important manufacturing problem. It focuses on the body in white process and more specifically on the selection of a correct combination of joining technologies taking into account investment, technology acquisition and variable manufacturing cost that has to be allocated to different joining stages from a constrained budget. This process can be followed later on by more accurate calculation based on process plans and information concerning the operation of the manufacturing facility. The aim of the proposed approach is to help with the cost analysis of different manufacturing options and to focus experts' effort on the most promising ones saving time, funds and increasing the overall quality of the product.

2. Overview of the Automotive Product Development Process

2.1 Product Development Process (PDP)

To successfully manage the entire vehicle development process, automotive companies use structured Product Development Process (PDP) approach. PDP allows for the overall synchronisation and control of the development processes in a specified timeframe, which varies among different automotive manufacturers. A typical example of the main activities is shown in figure 1. The PDP is subdivided into several phases starting with the Preliminary Phase, known as Early Phase of development, and finishing with the Start of Production (SOP) phase.

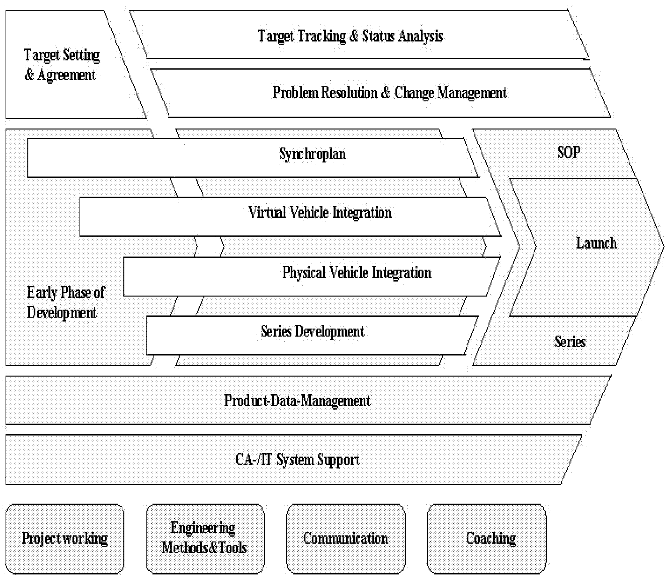
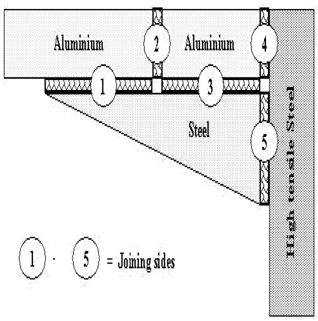


Figure 1: Example of the main activities involved Product Development Process.

An important part of the Early Phase of the PDP is the Concept Phase. Within the Concept Phase, that is, the first stage of the Early Phase, a Design trend definition on the basis of an initial design model can be provided. The problem in this part of the PDP is the lack of data and information concerning future costs of new materials, manufacturing processes, joining techniques etc. Under such circumstances of uncertainty a designer cannot make an informed decision concerning the future manufacturing costs of a new design concept without extra input from Process Planners and Plant Managers, which is a lengthy process [3]. In the same time almost 90 % of the future product costs are predefined at the Early Phase of the PDP by the Design and Planning Department at the Concept Phase [4].

2.2 Body-in white design process

At the early stage of development a coherent technical design concepts are introduced. This comprises the configuration of the vehicle, design characteristics and design trends, innovation concepts, component concepts and system concepts. Body-in-white (BIW) design is considered to be the most complicated part at the preliminary phase of the automotive PDP. This is one of the core competences of the automotive industry that is not normally outsourced. The integration of the design and planning activities at this stage are of critical importance and there is a need of techno economical models to serve as an interface between these product development activities. The idea behind the proposed approach is to use working contents available from the conceptual design concept, to calculate the investment costs for different joining technologies. Figure 2 shows a schematic of joining sides and materials that have to be transformed into working contents and joining techniques.



	Material	Length [mm]
1. Joining side	Aluminium - Steel	520
2. Joining side	Aluminium - Aluminium	275
3. Joining side	Aluminium - Steel	460
4. Joining side	Alu. - High tensile Steel	275
5. Joining side	Steel - High tensile Steel	350

Figure 2: Schematic of a wheelhouse and joining sides.

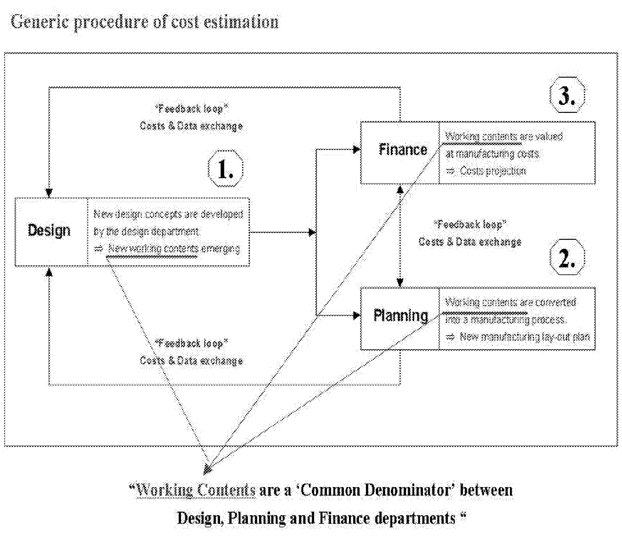


Figure 3 Schematic of the cost estimation procedure.

Fig. 3 shows the general idea of using working contents as a link between the design and production process.
They are a „common denominator"between the design, planning and finance departments because WC has technical and economical interpretation. They can be linked to joining techniques and therewith with cycle times.
1.) A designer knows that the individual BIW parts of his design concept have to be joined together by means of JT's and he can calculate the total number of WC's at the end of the BIW design process.
2.) A planner designing a production line aims optimising cycle time. Considering the number of WC's for the BIW joining process one can identify the cycle time diagram.
3.) Once the number of WC's has been ascertained for a given production line, the finance department is in a position to calculate the total investment and manufacturing cost by integrating cycle times with the various WC's.

The discrete nature of the joining process containing number of sequential stages suggests that dynamic programming can be used for the purposes of process optimisation. A designer has to make a decision regarding the optimal combination of joining techniques in respect of the manufacturing cost incurred. The optimisation procedure begins with the identification of the sides that have to be joined together, their joining length, and the material of the participating parts - as shown in Fig. 2 followed by the selection of appropriate joining techniques for each joining side, as it is indicated in figure 5.

3. Optimisation of the Joining Process Selection

Because of the space limitation the proposed optimisation method is explained below on the basis of a generic joining techniques case study.

3.1 Initial considerations of the joining technology selection

Considering the process of joining technology selection separately at every joining side, a designer has to choose among several alternative technically applicable joining technologies. Knowing the material required on each joining side, a list of joining techniques that are authorized for a particular joining side can be obtained. Following the cost estimation procedure indicated in Fig.3 a cost tables containing investment and manufacturing costs for every joining technology can be created as it is shown in Fig. 4 and the cycle time for a particular operation can be calculated which allows for the calculation of the cost of a working content specific for this technology. The procedure has to be repeated for all of the candidate technologies. The next step of the proposed methodology is to account for the specifics of the design and manufacturing complexity. For this purpose a set of coefficients that reflex the

Joining Technique:	Resistance spot welding	Steel - Steel		
Installation:	Flexibel manufacturing cell for the joining process of two 1000 mm **steel** sheet components by utilisation of spot welding			
Investment cost:	198.744 EURO		per spot:	0,00531 €/spot
Basis:	465 min/shift, 2 shifts/day, 244 days/year, for 7 years, 1412 spots/hour			
Manufacturing costs:				
	Type/Unit	Consumption (units/hour)	Costs (EURO/unit)	Costs (EURO/hour)
Manufacturing resources:				
- Extra wire				
- Bolts				
- Rivets				
- Screws				
- Rivet-bolts				
- Punch-rivet				
- Electrodes	1 pair per 4000 spots	0,1804	0,51 €	0,09 €
Electricity/water/air/gas:				
- Electricity for robot	6,6 KVA/%	6	0,047 €	0,28 €
- Elect. welding transformer	0,018 kwh/spot	25	0,047 €	1,20 €
- Electricity for turntable	4 KWh/%	2,8	0,047 €	0,13 €
- Cooling water	2 bar/ 4ltr. per min.	240	0,00005 €	0,01 €
- Inert gas				
- Compressed air	12 bar/ 2ltr. Per spot	2824	0,00005 €	0,14 €
Maintenance:				
a) Material				
- Cap-cutter	1 piece per cycle	0,0016	2.556,46 €	4,09 €
- Cutter	2 pieces per cycle	0,0288	20,45 €	0,59 €
b) Labour	Minutes	2	1,02 €	2,04 €
Sum manufacturing cost per hour without depreciation:				8,58 €
Depreciation from the total investment of:		465 min/shift, 2 shifts/day, 7 years	198.744 €	7,48 €
Sum manufacturing cost per hour with depreciation:				16,06 €
Example: 1000mm joining length (20 welding spots/ Spot distance 50mm)				
- Joining time:	44 sec			
- Robot time:	2 sec			
- Turntabel time:	5 sec		per Working Content:	
- Total cycle time:	51 sec	(1412 spots/hour)	per spot	0,01137 €/spot

Figure 3 Table containing investment and manufacturing costs

complexity of the design has been introduced. An example is shown in Fig. 4.

Spot welding			
Complexity level 1 JS Coefficient = 1,00 100% cycle time	1. Blank sheet thickness:	Variation of blank sheet thickness 0%, two-blank sheet-welding of equal blank sheet thickness, blank sheet thickness > 1,5mm	Plane part contour 100% cycle time
	2. Effective width:	Effective joining width according to DIN: 20mm < effective width of the joining blank sheet	
	3. Surface condition:	Blank sheet surface without discoloration, free of oil, free of grease, without coating, not galvanised	
	4. Accessibility:	Great degree of freedom for the welding gun, little steel in "secondary welding gun window", easily positioning of the welding gun because of an appropriate blank sheet contour ⇒ plane part contour	
Complexity level 2 JS Coefficient = 1,25 125% cycle time	1. Blank sheet thickness:	Two-blank sheet-welding of different blank sheet thickness 0,7mm < blank sheet thickness < 1,5mm, variation of blank sheet thickness < 20%	Curved part contour 125% cycle time
	2. Effective width:	Effective joining width outside of DIN reference, effective joining width narrow, 20mm > effective width of the joining blank sheet	
	3. Surface condition:	Blank sheet surface without discoloration, without coating	
	4. Accessibility:	Average volume of steel in the "secondary welding gun window", normal size of the welding gun (length of pole X < 300m), normal positioning of welding gun because of an appropriate blank sheet contour ⇒ curved part contour	
Complexity level 3 JS Coefficient = 1,50 150% cycle time	1. Blank sheet thickness:	Two or three-blank sheet-welding of equal blank sheet thickness 0,7mm < blank sheet thickness < 1,5mm, variation of blank sheet thickness > 20%	Stepped part contour 150% cycle time
	2. Effective width:	Joining width under limit, 15mm > effective width of the joining blank sheet	
	3. Surface condition:	Blank sheet surface with discoloration, heavily oiled, greased, galvanised	
	4. Accessibility:	Shortage of space, great amount of steel in "secondary welding gun window", big welding gun (length of pole X > 300mm), difficult positioning of welding gun because of a complicate blank sheet contour ⇒ stepped part contour	

Figure 4 Example of complexity levels for spot welding

Figure 5 provides an example of the process of allocation of joining technologies to the joining sides without considering budget constrains.

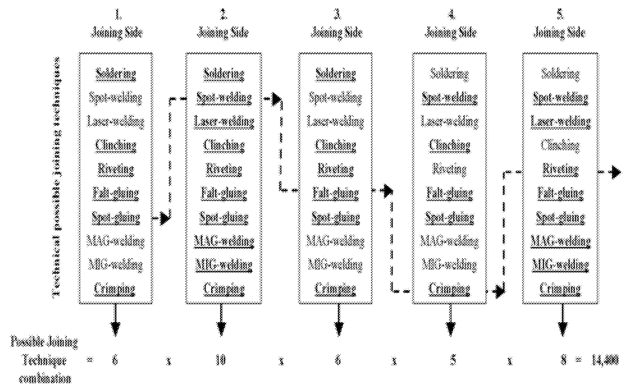


Figure 5: Unconstrained allocation of joining techniques to the joining sides.

The technologies that are indicated in Fig. 5 are based on the example shown in Figure 2 representing a wheelhouse with 5 joining sides and its material diversity. The allocation process is illustrated by the dotted line arrows. In this example only subsets of authorised joining techniques selected using expert opinions have been considered. These feasible technologies are underlined in the figure. Taking into account the number of authorized joining techniques for every joining side, the possible number of joining technique combinations has been calculated as 14,400. When constrains on the investment budget are considered and the optimisation goal is to minimise the manufacturing costs even from this oversimplified example it is apparent that an optimisation procedure is needed to support the decision making process. To emphasise further this conclusion it is worth mentioning that, in reality, in an automotive BIW wheelhouse design the expectation is for approximately 200 joining sides, each with an average of 4 appropriate joining techniques - which increases the number of possible combination up to 4^{200} to find a global optimum by exhaustive enumeration appears to be impractical. As it has been mentioned above this paper proposes an optimisation procedure based on dynamic programming method.

3.2 Optimisation procedure for allocation of joining techniques to joining sides

Dynamic programming (DP) has been selected as a mathematical procedure to improve the computational efficiency of the allocation of joining techniques to joining sides problem. The idea behind DP is to decompose the global optimisation problem into smaller, and hence computationally simpler, sub-problems called stages. They are defined as the portions of the problem that possesses a set of mutually exclusive alternatives from which the best alternative is to be selected. In the example discussed above, stages represent the joining sides. The computations at the different stages are linked through recursive computations in a manner that yields a feasible optimal solution to the entire problem.

Each sub-problem is then considered separately with the objective of reducing the volume and complexity of computations.

		Stage 1		Stage 2		Stage 3		Stage 4		Stage 5	
Joining side No.:		1.		2.		3.		4.		5.	
Joining side Length [mm]:		520		275		460		275		350	
Joining side Material:		Aluminium - Steel		Aluminium - Aluminium		Aluminium - Steel		Aluminium – High tensile Steel		Steel - High tensile Steel	
No.	Joining Technique \ Costs	c_1	M_1	c_2	M_2	c_3	M_3	c_4	M_4	c_5	M_5
1.	Soldering	25	15	15	10	30	20	--	--	--	--
2.	Spot-welding	--	--	10	15	--	--	15	20	20	30
3.	Laser-welding	--	--	30	20	--	--	--	--	15	10
4.	Clinching	15	40	10	20	15	30	10	15	--	--
5.	Riveting	20	30	10	25	20	10	--	--	10	25
6.	Falt-gluing	35	25	30	15	25	15	30	20	25	20
7.	Spot-gluing	30	20	40	10	40	15	25	10	30	15
8.	MAG-welding	--	--	15	15	--	--	--	--	20	25
9.	MIG-welding	--	--	20	10	--	--	--	--	15	20
10.	Crimping	25	20	30	20	25	25	30	20	15	15
c_j = Investment costs for the joining technology ; M_j = Manufacturing costs of the joining technology of joining side j											

Figure 6 Joining techniques - joining sides starting matrix.

In terms of the example in figure 2, each joining side defines a stage, with the first up to the fifth stages having up to ten alternatives respectively, as shown in figure 4. These stages are interdependent because all five joining sides must compete for a *limited* investment costs budget. The aim of the above DP problem is the optimisation of the sum of manufacturing costs: $M_{opt.} = M_1 + \ldots + M_5$ under the compliance of the constraint: $c_1 + \ldots + c_5 \leq 90$. In the present example, utilised data are deduced from real world sources.
The accomplishment of the DP algorithm results in the joining technique combinations assigned in Fig. 7, which achieve the optimal solution for manufacturing costs of 60 and investment costs of 85.
Five iterations were needed to achieve an optimal solution, which is a significant improvement in comparison with the enumeration approach that requires valuation of the entire set of 14,400 joining technique combinations.

Permissible Solutions:		1. Joining side	2. Joining side	3. Joining side	4. Joining side	5. Joining side	Investment costs	Manufacturing costs
1. Solution	[1,1,5,4,3]	Soldering	Soldering	Riveting	Clinching	Laser-welding	**85**	**60**
2. Solution	[1,9,5,4,3]	Soldering	MIG-welding	Riveting	Clinching	Laser-welding	90	60

Figure 7 Optimal solutions for the technology selection example

The importance of using an appropriate optimisation model increases with the increase in the number of the joining sides.

4. Conclusions

A central element of the proposed approach is the possibility of manufacturing cost estimation of different BIW design concepts in the preliminary work phase of the automotive PDP. This early cost estimation gives the BIW designer an opportunity to evaluate new design concepts in advance and to decide which will be best in terms of cost. Hence the design department will acquire a disclosure of the cost savings potential early in development process.
The development of a model linking working contents and joining techniques provides a common platform for decision making between the Design, Planning and Finance experts in the BIW design process. The use of Working Contents in the Early Phase of the design process also accelerates the process of data exchange between Design, Planning and Finance activities. Hence, the cooperation is intensified, and the number of time consuming information feedback loops is reduced.
The proposed generic joining techniques selection methodology considers the specific needs of the BIW design engineers. This methodology brings together a series of techniques, and provides a holistic procedure of optimal joining techniques selection. The computational procedure of the proposed selection method provides optimal combination of joining techniques for the design concept under scrutiny in terms of minimised manufacturing costs and compliance with technical and economic constrains.
The introduced DP optimisation procedure is applicable to a variety of optimisation and constraint criteria, such as the optimisation of the weight, energy, running costs etc. of the joining techniques process.
Optimisation at an early stage in the design and development process assists planning, finance and manufacturing experts in the selection of the most appropriate design concepts, subject to selected manufacturing criteria.

Acknowledgements

This research has been supported by the Research & Development Centre of the BMW Group, Germany, and the School of Industrial and Manufacturing Science, Cranfield University, UK. The authors would like to express their sincere thanks to Mr. Appold and Mr. Schrödinger for their attention and useful comments.

References

[1] Nitschke F., 1998, Markt- und prozessorientiertes Kostenmanagement von Entwicklungsvorhaben im Automobilbau. Dr. Kovač Verlag Hamburg.
[2] Ehrlenspiel K., 1998, Kostengünstig Entwickeln und Konstruieren, Springer Verlag Berlin Heidelberg.
[3] Beitz W., Birkhofer H. and Pahl G., 1992: Konstruktionsmethodik in der Praxis. In: Konstruktion, 44 (1992) 12, pp. 391-397.
[4] Botta V., 1996, Mitlaufende Kalkulation für eine frühzeitiges Kostenmanagement. KRP Kostenrechnungspraxis, Sonderheft1.

AUTHOR LISTING

AUTHOR LISTING