

Luigi Telesca
Katarina Stanoevska-Slabeva
Veselin Rakocevic (Eds.)



21

Digital Business

First International ICST Conference, DigiBiz 2009
London, UK, June 2009
Revised Selected Papers



 Springer

Lecture Notes of the Institute
for Computer Sciences, Social-Informatics
and Telecommunications Engineering

21

Editorial Board

Ozgur Akan

Middle East Technical University, Ankara, Turkey

Paolo Bellavista

University of Bologna, Italy

Jiannong Cao

Hong Kong Polytechnic University, Hong Kong

Falko Dressler

University of Erlangen, Germany

Domenico Ferrari

Università Cattolica Piacenza, Italy

Mario Gerla

UCLA, USA

Hisashi Kobayashi

Princeton University, USA

Sergio Palazzo

University of Catania, Italy

Sartaj Sahni

University of Florida, USA

Xuemin (Sherman) Shen

University of Waterloo, Canada

Mircea Stan

University of Virginia, USA

Jia Xiaohua

City University of Hong Kong, Hong Kong

Albert Zomaya

University of Sydney, Australia

Geoffrey Coulson

Lancaster University, UK

Luigi Telesca
Katarina Stanoevska-Slabeva
Veselin Rakocevic (Eds.)

Digital Business

First International ICST Conference, DigiBiz 2009
London, UK, June 17-19, 2009
Revised Selected Papers

Volume Editors

Luigi Telesca
CREATE-NET
Via alla Cascata 56D
38100 Povo, Trento, Italy
E-mail: luigi.telesca@create-net.org

Katarina Stanoevska-Slabeva
University of St. Gallen
Institute for Media and Communications Management
Blumenbergplatz 9
9000 St. Gallen, Switzerland
E-mail: katarina.stanoevska@unisg.ch

Veselin Rakocevic
City University London
School of Engineering and Mathematical Sciences
London EC1V 0HB, United Kingdom
E-mail: v.rakocevic@city.ac.uk

Library of Congress Control Number: 2009942428

CR Subject Classification (1998): K.4.4, J.1, H.4.1, I.6, C.2

ISSN 1867-8211
ISBN-10 3-642-11531-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-11531-8 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© ICST Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12830176 06/3180 5 4 3 2 1 0

Preface

This volume contains the proceedings of the First International ICST Conference on Digital Business (DigiBiz 2009), hosted by City University London in London, UK. This annual event had the main objective to stimulate and disseminate research results and experimentation on future Digital Business to a wider multidisciplinary forum that would allow the participants to cross the boundaries between research and business.

The scientific offering in e-business, e-commerce, and ICT in general is quite broad and spans many different research themes, involving several communities and methodologies. The growth and dynamic nature of these research themes pose both challenges and opportunities. The challenges are in having scientists and practitioners talk to each other: despite the fact that they work on similar problems they often use very different languages in terms of research tools and approaches. The opportunities on the other hand arise when scientists and practitioners engage in multidisciplinary discussions leading to new ideas, projects and products.

The first DigiBiz in London engaged the participants in discussions on the evolution of business models, the increased liquidity of competition through increased availability of information and the impact of redesigning the Internet for the current and future business landscape. Topics such as cloud computing, digital media, enterprise interoperability, digital business ecosystems and future Internet paradigms were debated and analyzed combining the technological and research aspects with a challenging business perspective. We tried to cover all aspects related to the success of digital business and to present the opportunities for the specifications of novel models, products and services that are, use and/or rely on ICT. In its first year, DigiBiz attracted some excellent scientific papers, application ideas, demonstrations and contributions from industrial research teams all over the world.

The 19 contributions published in this volume represent perspectives from different fields, including organizational theory, applied economics, information studies, communication, computer science, software engineering and telecommunication. A range of different methodologies have been used, spanning from qualitative to quantitative.

The papers deal with subjects such as virtual business alliances, organizational and coordination theory, model-driven businesses, efficient energy management, e-negotiations, future Internet markets and business strategies, agent technologies, recommender for active preference estimate, trust & reputation models for e-commerce, RFID architectures, decentralized system for ubiquitous VoIP services, reservation and incentive schemes for grid markets, end-user-driven service creation, digital media distribution, IP-TV, social media trails, P2P business and legal models. Collectively they demonstrate the dynamic and interdisciplinary nature of the business. Each of the chapters was competitively selected over 60 submitted works, using a rigorous peer-reviewed process.

The conference emphasized European leadership by featuring keynote speakers from the European Commission and contributions from various Commission-funded research projects, exhibiting the strength of European industry in the field. It also

reviewed innovative visions and ideas from key industry players such as SAP, IBM, BT and Sole 24 Ore, without disregarding regulatory aspects related to the evolution of the mobile business arena.

Many people have contributed to the success of DigiBiz. It was a privilege to work with these dedicated people, and I would like to thank them all for their efforts. The Organizing Committee as a whole created a collaborative working atmosphere, which made our task an easy one. Ian Cloete and Dirk Trossen supported us in the definition of the DigiBiz concept. Elmar Husmann actively contributed creating a challenging panel on cloud computing. Javier Vasquez made an enormous contribution to the development of innovative workshops for the conference, while Katarina Stanoevska did a great job assembling the Technical Program Committee and running the paper submission and reviewing process. Veselin Rakocevic meticulously oversaw the preparation of the conference proceedings. Finally, we would like to thank all the distinguished members of the Program Committee - the soul of our event. I would also like to thank our keynotes: Orestis Terzidis, Cristina Martinez, Steve Wright, Leonardo Chiariglione and Andrea Gianotti for their contributing an inspiring keynotes.

Organizing a conference which is the first of its kind is always an adventure. I was delighted to have been part of it.

Luigi Telesca

Organization

Steering Committee Chair

Chlamtac Imrich University of Trento, CREATE-NET, Italy

General Chairs

Kersten Gregory J. Molson School of Business, Concordia, Canada
Telesca Luigi University (Canada), CREATE-NET (Italy)

Research & Position Papers Committee

Kersten Gregory J. Molson School of Business, Concordia University,
Canada
Dirk Trossen BT Research (UK)

Business Leader Forum Committee

Cloete Ian International University (Germany)
Elmar Husmann IBM (Germany)

B2B Exhibit and Demo Track Chair

Biswas Shantanu Motorola (India)

Program Committee Chair

Stanoevska Slabeva Katarina MCM Institute University St. Gallen (Switzerland)

Workshops Chair

Javier Vazquez Salceda UPC (Spain)

Local Chair

Rajarajan Muttukrishnan City University London (UK)

Publication and Content Co-chairs

Veselin Rakocevic City University London (UK)
Dasun Weerasinghe City University London (UK)

Web Chair

Andrea De Luca CREATE-NET (Italy)

Publicity & Sponsorship Chair

Saritha Arun IBM (UK)

Conference Coordinator

Gergely Nagy ICST
Stela Stancheva CREATE-NET

Technical Program Committee

Mu Xia	University of Illinois at Urbana-Champaign, USA
Izak Benbasat	University of British Columbia, Canada
Dirk Neumann	Universität Freiburg, Germany
Ting-Peng Liang	National Sun Yat-sen University, Taiwan
Michael J. Shaw	University of Illinois at Urbana-Champaign, USA
Martin Bichler	Technische Universität München, Germany
Olivia Sheng	University of Utah, USA
Daniel Zeng	University of Arizona, USA
Michael Chau	The University of Hong Kong, China
Hsiangchu Lai	National Sun Yat-sen University, Taiwan
John Debenham	University of Technology, Sydney, Australia
Ryszard Kowalczyk	Swinburne University of Technology, Australia
Nikos Karacapilidis	University of Patras, Greece
Pierpaolo Pontrandolfo	Pierpaolo Pontrandolfo Bari Politechnico, Italy
Grazia Concilio	Politechnico di Milano, Italy
Tomek Szapiro	Warsaw School of Economics, Poland
Catholijn M. Jonker	Technical University Delft, The Netherlands
Jan Kok	NSN
David Kennedy	Eurescom
Fiona Williams	Ericsson
Felix Hampe	University of Koblenz, Germany
Petra Schubert	University of Koblenz, Germany
Hannes Werthner	Technical University of Vienna, Austria
Niels Bjorn- Andersen	Copenhagen Business School, Denmark
Stefan Klein	University of Münster, Germany
Reima Suomi	University of Turku, Finland
George Thanos	University of Athens, Greece
Joze Gricar	Slovenia
Bernhard Katzy	Universität der Bundeswehr München, Germany
Yao-Hua Tan	Free University of Amsterdam, The Netherlands
Christine Legner	University of St. Gallen, Switzerland

Walter Brenner	University of St. Gallen, Switzerland
Yves Pigneur	University of Lausanne, Switzerland
Ulrike Lechner	Universität der Bundeswehr, German
Hans-Dieter Zimmermann	University of Applied Sciences Chur, Switzerland
Giulio Destri	University of Parma, Italy
Sergio Gusmeroli	TXT Solutions
Fabio Casati	University of Trento, Italy
Alberto Marcati	LUISS, Italy
Josep Lluís de la Rosa	University of Girona, Spain
Giovanni Cortese	Interplay Software, Italy
Renata Guarneri	CREATE-NET, Italy
Pierfranco Ferronato	Soluta.net, Italy
Mike Boniface	IT Innovation Centre
Jason Finnegan	WIT/TSSG, Ireland
Flavio Bonfatti	University of Modena, Italy
Nagaraj Konda	UCE, UK
Javier Val	ITA, Spain
Miguel Vidal	Techideas, Spain
Oliver Christ	SAP
Christian Merz	SAP
Santi Atos	Origin, Spain
Juan Carlos F. Galindo	Telefonica, Spain
Veli-Pekka Niitamo	Nokia, Finland
Hristo Koshutanski	University of Malaga, Spain
Francesco Ricci	Free University of Bozen-Bolzano, Italy

Table of Contents

Reservation Schemes for Grid Markets	1
<i>Marcel Risch</i>	
A Computational Trust Model for E-Commerce Systems	9
<i>Mehrnoosh Divsalar and Mohammad Abdollahi Azgomi</i>	
A Recommender for Active Preference Estimate	19
<i>Andrea Malossini and Paolo Avesani</i>	
An MDA Based Environment for Modelling and Executing Business Negotiations	27
<i>Pierfranco Ferronato and Jason Finnegan</i>	
Analysing Requirements for Virtual Business Alliances – The Case of SPIKE	35
<i>Christian Broser, Christoph Fritsch, Oliver Gmelch, Günther Pernul, Rolf Schillinger, and Sandra Wiesbeck</i>	
Business Impact of Online Institutional Recommendation	45
<i>Mihaela Ion and Regis Saint-Paul</i>	
uCard: A Novel Privacy Friendly Radio Frequency Tag Architecture for Universal Loyalty Card	53
<i>Krishan Sabaragamu Koralalage and Noriaki Yoshiura</i>	
VERTIGO: Find, Enjoy and Share Media Trails across Physical and Social Contexts	64
<i>Luca Galli, Renata Guarneri, and Jukka Huhtamaki</i>	
VOICE: A Decentralized System for Ubiquitous VoIP Services	74
<i>Costas Kalogiros, Costas Courcoubetis, and Panayotis Antoniadis</i>	
Monitoring Ecosystem Characteristics on Negotiation Environments	82
<i>Nicolás Hormazábal, Josep Lluís de la Rosa, and Gabriel Lopardo</i>	
Incentives, Business Model and Required Technical Infrastructure for the Economic Aware Data Grid	90
<i>Dang Minh Quan, Orestis Terzidis, and Sonja Klingert</i>	
The Incoming Trends of End-User Driven Service Creation	98
<i>Zhenzhen Zhao, Nassim Laga, and Noel Crespi</i>	
The New Way to Think Television on the Web	109
<i>Riccardo Mangiaracina, Giovanni Toletti, and Luca Turba</i>	

The Wisdom of Sustainable Communities in the Digital Era: The Case of Efficient Energy Management	119
<i>Alessandro Tavoni and Luigi Telesca</i>	
P2P Business and Legal Models for Increasing Accessibility to Popular Culture	130
<i>Eitan Altman, Sulan Wong, and Julio Rojas-Mora</i>	
Disruption of the Digital Media Distribution	139
<i>Margaret Tan and Miguel A. Morales-Arroyo</i>	
Evolutionary and Pervasive Services	149
<i>Elia Conchione</i>	
Future Internet Markets and Business Strategies: An Analysis and a Proposal	156
<i>Man-Sze Li</i>	
Impact of Microfinance Operations on the Livelihood of the Clients: A Review of the Existing Literature	166
<i>Aliar Hossain, Mohammad Kamal Hossain, and Karim Mohammed Rezaul</i>	
Combining Organisational and Coordination Theory with Model Driven Approaches to Develop Dynamic, Flexible, Distributed Business Systems	175
<i>Javier Vázquez-Salceda, Luigi Ceccaroni, Frank Dignum, Wamberto Vasconcelos, Julian Padget, Siobhan Clarke, Paul Sergeant, and Kees Nieuwenhuis</i>	
Author Index	185

Reservation Schemes for Grid Markets

Marcel Risch

International University in Germany, Campus 3,
76646 Bruchsal, Germany
Marcel.Risch@i-u.de

Abstract. Grid markets today have very limited capabilities with regards to reservations. However, it is vital that customers are given the ability to purchase resources far in advance of the actual usage which necessitates the existence of a sold reservation model. While this is not an issue of great importance in today's Grid markets, reservation schemes must be developed to enable a sustainable operation of Grid markets. This paper explores a number of different reservation models that are in use today and evaluates their usefulness both from the point of view of resource sellers as well as from the point of view of resource buyers. It then goes on to propose a new reservation scheme for grid markets.

Keywords: Grid Markets, Reservations.

1 Introduction

One of the major issues that must be addressed if open Grid markets are to become a reality is the issue of reservations. Since grid customers will not only wish to buy resources shortly before they are needed but also far in advance of actual usage, reservations are vital to keep such an economy operating. While many markets have been developed and some have reached the production-level stage, reservations still remain largely ignored.

Such an option is a requirement if customers are to be allowed to order resources far in advance of the actual usage. This requirement, in turn, leads to the important question of which type of reservation should be used. There are a number of reservation types which are used in various settings. These will be analyzed with regards to a grid market so that the appropriate reservation scheme can be chosen.

The paper is structured as follows: in the next section, two currently existing grid markets will be explained with a special emphasis on their reservation schemes. In the following section, a number of reservation schemes which are currently in use in other fields are analyzed with regards to their suitability for grid markets. We then develop an improved reservation scheme which is based on a prepaid model. This section will be followed by an analysis of the future work.

2 Existing Reservation Schemes

2.1 Existing Grid Markets

The Distributed European Infrastructure for Supercomputing Applications (DEISA) [1] is an organization in which eleven high performance computing (HPC) centers are

interconnected to provide computing power mainly to researchers from participating countries. In addition to researchers, private enterprises can also use some of the computing power for which they have to pay. The procedure for gaining access to these resources is very complex, as can be seen from [2].

While the DEISA consortium reserves the resources, it does so at its own convenience. Customers do not have any influence on the time at which the resources are available but must wait until their assigned time. Therefore, this model cannot be considered to be customer friendly and is not suitable for a grid market in which customers can obtain resources for usage at their convenience.

The Amazon Elastic Compute Cloud (EC2) [3] is the first grid market in which anybody can buy resources. The resources can be bought in bundles for a fixed price and are charged for according to the actual usage. This system allows users to create images which are then uploaded to virtual computers and run on the Amazon.com servers. While this approach has many benefits, it is nevertheless not perfect from an economic point of view.

Users who want to reserve resources find themselves unable to do so. While they are able to add multiple resources to an existing resource pool, they cannot reserve the resources in advance. Thus, it can happen that when resources are scarce the customer cannot add urgently needed resources to his pool. Furthermore, once the resource is obtained, it can be used without any time constraints, which implies that Amazon.com is unable to determine when resources will become available for further use.

Tsunami Technologies [4] provides clustering services to paying customers. This company provides a number of services:

- On-Demand Cluster: A simple clustering scheme without any requirements. Users can use as much or as little compute power as they require.
- Subscription Clusters: In this scheme, users agree to use a minimum number of CPU-hours per month, in exchange for a lower CPU-hour rate.

The first scheme is identical to the Amazon EC2 scheme. The second scheme can be seen as a middle ground between a reservation and a reservation-less system. By requiring the user to use a certain amount of CPU-hours, the company can perform some basic planning tasks. On the other hand, it is not clear whether the resources are guaranteed to be available for the required amount of hours per month or whether they can be reserved for certain time periods.

2.2 Reservations in Other Markets

Since current markets do not handle reservations well, we have to look at other areas in which reservations have been widely used. There are a number of reservation models that currently exist which could all be applied to grid markets.

The restaurant-style is modeled after the reservation service offered in restaurants. In the case of grid markets, the customer orders a resource far in advance of the actual usage, telling the provider a starting time and date for the usage, as well as the duration for which the resources are to be used. The provider then reserves the resources ordered. Should the customer use the resources, they will remain reserved for the duration specified by the customer when the reservation was made. However, if the

customer is not ready by the starting time, the reservation is canceled and the provider can attempt to sell the idle resources on the grid market.

For the client, this model offers a number of distinct advantages over buying resources for in-house installation. Firstly, this model allows the customer to reserve resources before he knows if he actually needs them. Secondly, if the client wishes to cancel the reservation, there will be no consequences for the customer.

For the provider, this model has few advantages. The main advantage lies in the fact that the provider can see how high the demand for a certain resource is. On the other hand, the provider is not able to accurately forecast the income generated, due to the risk of cancellations which requires the provider to apply some strategies to offset the consequences of these cancellations, such as overbooking.

As opposed to the restaurant reservation scenario, where the customer had nothing to fear in the case of late cancellations or expiration of reservations, the airline-style reservation makes it the customer's responsibility to have all data ready on time. The reservation scheme is modeled after the airline reservation system in which the customer reserves a seat, pays in advance and then gets a ticket and reservation. In the case of cancellations, there are three different models to consider: in the first case, the customer is refunded the entire payment; in the second case, the customer is returned the payment minus a fixed fee; and in the last case, the customer cannot cancel the ticket at all. Thus, the risk incurred by a cancellation is not entirely shouldered by the provider in the last two cases. The first is identical to restaurant-style reservations.

The two remaining cases are much more advantageous for the provider as compared to the previous reservation scheme. Firstly, since the customer pays in advance, the provider can be certain of the income generated by the resources. If the customer has the job delivered on time, he can have access to all resources. Should the client not be ready by the starting time of reservation, the provider has the option of selling the resources on the grid market. In addition, the provider can be certain of a small amount of income being generated. Furthermore, if the provider is able to sell the idle resources, the income generated is increased by the cancellation fee, thus giving the provider an added bonus for allowing reservations.

For the resource user, this model is more risky, since a cancelled reservation can be costly. In the case that the resources are still needed and the reservation has been cancelled, the cost will be even higher, since the customer is forced to purchase the required resources on the grid market. As opposed to the restaurant-style reservation scheme, this reservation model puts a fixed amount of the costs for missing a reservation on the shoulders of the customer.

Since the first two reservation schemes have been focused on either party carrying the financial burdens of cancelled reservations, the third alternative is a mixture of these two. Its first point of change is the fact that reservations do not expire if the customer is late. Instead, the length of time that the resources will be kept available depends on the non-refundable down payment the client is willing to pay. An improved solution is a proportional model which can be described mathematically as follows:

$$\frac{P_{Down}}{C_{Total}} = \frac{T_{Available}}{T_{Reserved}} \quad (1)$$

In the equation, P_{Down} denotes the down payment made by the customer, C_{Total} denotes the total cost for the resources for the entire reservation period, $T_{\text{Available}}$ denotes the time for which the resources will be kept available without customer action, T_{Reserved} denotes the total reservation period. Thus, the duration for which the resources will be kept available even if the customer does not take any action is proportional to the size of the down payment. In other words, if the client pays 50% of the total resource cost, the resources will be reserved for half the time that the client requested them for.

The improvement of the model lies in its fairness and its firm guarantees. The client can be certain that the resources will be available for a certain amount of time which is influenced by the size of the down payment. The provider, on the other hand, can determine the likelihood that the resources will be actually used by the customer. A large down payment would indicate a high probability that the resources will be needed; while a low down payment would indicate that the customer is not certain if the reserved resources will be needed. This is proportional model is somewhat similar to the truth telling reservation scheme proposed in [5].

Additional advantages for the provider are that the customers are encouraged to take the reservation seriously, since the resources have been partially paid for. Furthermore, this model gives the provider the ability to plan the cash flow because he knows a lower bound on the income thanks to the prepaid reservations.

This model has very few disadvantages for both resource provider and client. The customer needs to ensure that the deadline can be met within the time frame specified by the down payment. The main disadvantage for the provider is that the hardware can remain idle because the customer doesn't need it but cannot be sold again on the grid market until the time reservation time has passed, depriving the provider of a second source of income.

3 An Improved Reservation System

Of the reservation schemes explained above, only the last shows some promise, since it strikes a balance between the risks shouldered by the customer and the risks shouldered by the provider. However, the main disadvantage is that the provider is very inactive when it comes to setting the size of the down payment. This new reservation model will rectify this shortcoming.

3.1 The Reservation Model

Since the airline-style reservation scheme is very advantageous for the provider, it should not be removed from the pool of possible reservation schemes in a grid resource economy. However, in order to level the playing field with regards to cancellations, the refund scheme must be changed to be more flexible than the currently existing schemes. In general, providers have three possible ways of calculating the size of the refund.

Ideally, the size of the refund should be proportional to the remaining time to the beginning of the intended resource usage. If customers cancel far ahead of the actual resource usage, they would get a large part of their money refunded. On the other hand, if the cancellation of the reservation is done shortly before the start of the reservation, only a small part of the down payment would be refunded.

This approach naturally leads to the question of how to determine the size of the refund. There are a number of possible mathematical solutions which will be detailed here. It is assumed that the customer makes the reservation at time T_0 and that the reservation will begin at time T_R while the cancellation is made at time T_C . The cost of the resources is assumed to be C_R while the refund is denoted by R .

The first and most basic of these refund models is the linear model in which the size of the refund is directly proportional to the time left to the start of the reservation.

$$R = C_R - \frac{C_R}{(T_R - T_0)} * T_C . \quad (2)$$

Thus the client is given an incentive to cancel the reservation as early as possible to maximize the refund. This equation is easy to understand, making the refund process very transparent for the customer. This can be seen in figure 1, where the refund percentage is linearly proportional to the time remaining till the reservation starts.

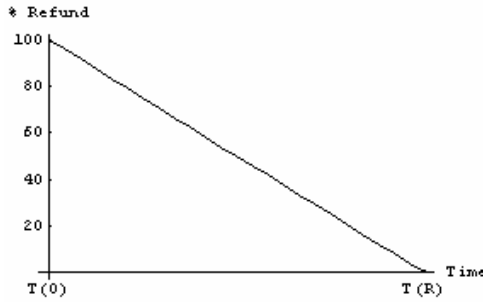


Fig. 1. Linear Model

An alternative strategy is to use a concave-up exponential function. In this case, the size of the refund would diminish very quickly but then fall slowly before the start of the reservation. This behavior can be modeled by the following equations:

$$\begin{aligned} R_1 &= \left(\frac{1}{1 + 100 * T_C} \right) * C_R \\ R_2 &= C_R \\ R &= \text{Min}[R_1, R_2] \end{aligned} \quad (3)$$

Therefore, unless customers cancel reservations early, they will not get a large refund, since R_1 will decrease very quickly. R_2 avoids the problem of having an infinite return if the reservation is cancelled immediately after being placed. However, if the reservation is cancelled quickly after being placed, the customer will get a large refund. This is illustrated in figure 2.

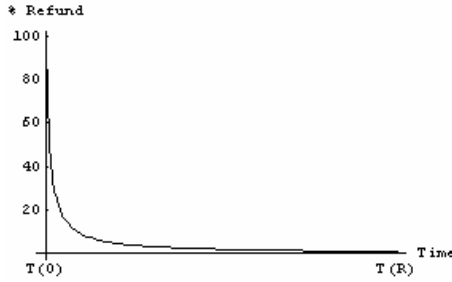


Fig. 2. The Concave-up Model

It can be clearly seen that the refund is very large initially but then drops sharply, thus encouraging the client to either cancel a reservation quickly or not to cancel it at all.

The final strategy is very customer friendly, since it will not punish the customer for cancelling late. Instead, a large part of the down payment will be refunded if the reservation is cancelled early enough. However, he will be penalized for canceling reservations late. Mathematically, this concept can be described as follows:

$$R_1 = \frac{\left(\frac{T_R}{T_C - T_R} + 100 \right)}{100} * C_R \tag{4}$$

$$R_2 = 0$$

$$R = \text{Max}[R_1, R_2]$$

The customer has an incentive to use the resources he has bought while at the same time is given some leeway to cancel the reservation. This is shown in figure 3.

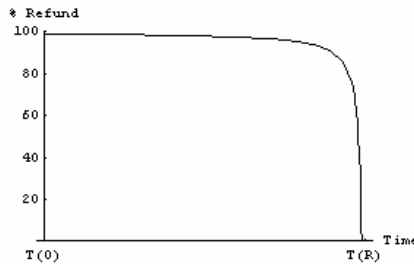


Fig. 3. The Concave-down Model

In this example, the user receives a large refund until shortly before the reservation begins. If such a scheme has been agreed upon with the provider, the customer can easily cancel the reservation without having to pay large fines.

3.2 Analysis of the Model

The prepaid model has a number of advantages for providers which will make this scheme very attractive to them. Since customers pay the moment the reservation is made, providers can be certain that they have a certain amount of income. The minimum income can be calculated by the time remaining to the start of the reservation. In addition, providers can choose from the three refund models the one which best fits the market situation and their own resource utilization. In essence, providers can choose the refund method as a way to attract customers or to reward returning customers.

For the customer, the mathematical model allows for a very transparent calculation of the size of the refund and thus allows the customer to confirm the provider's calculation. This ability to check the provider is another point in favor of such a scheme, since customers may not trust their providers.

As to which refund scheme is best under which circumstances is very difficult to say, due to the many parameters that must be considered: the demand, the willingness of customers to commit to resource purchases, the refund models used by other providers and the objectives of the provider. In the case of high demand and a high willingness to commit to resources, providers can choose the concave-up refund scheme, since customers will be concerned about being able to buy the resources they need. However, if the provider wishes to attract more customers, he may choose to use the linear model instead. This should attract customers who are not as certain of their resource requirements but still wish to participate in the grid market. It may also be the case that the provider chooses to offer the concave-down refund scheme to frequent customers as an incentive to return to this provider and not use other providers.

In the case of high demand and a low willingness to commit to resources, the provider can hope at best to use the linear model and may have to resort to the concave-down refund scheme. Using the concave-up model will only serve to discourage customers and thus be counterproductive when it comes to selling resources.

However, it is not only a matter of demand and the customer's willingness to commit to a purchase. If other providers offer mainly the concave-down or linear refund schemes, customers will prefer these, since they may not be entirely certain whether or not they need the resources. Furthermore, a provider needs to consider which refund policy other providers use. If most providers use the concave-up refund model, the provider may be able to attract more customers by using the less strict linear model. On the other hand, if most providers use the linear or concave-down models, using the concave-up model would be counterproductive, since customers would choose to buy resources from the more flexible providers. Thus, determining the best refund scheme is up to the provider who must take into account the current market situation, the actions of other providers while taking his own goals into account.

4 Conclusions and Future Work

In this paper we have analyzed a number reservation schemes used in grid markets today, as well as analyzing some reservation schemes that are used outside of grids. The former showed that in today's grid economy, reservations are very much neglected.

Since the additional reservation schemes examined were not satisfactory, a new model was developed which is based on the prepaid reservations where the size of the refund is based on the time remaining to the start of the reservation.

The suggested refund model allows providers to choose the function with which to calculate the refund. The choice depends on a number of parameters and must be carefully made so that customers will not be turned away to the competition. This determination of which reservation scheme to use at which point in time is the topic of further research.

References

1. DEISA - Distributed European Infrastructure for Supercomputing Applications (2008), <http://www.deisa.org>
2. DEISA Extreme Computing Initiative, DECI (2008), <http://www.deisa.eu/science/deci>
3. Amazon Elastic Compute Cloud, Amazon EC2 (2008), <http://aws.amazon.com/ec2>
4. Tsunami Technologies Inc. (2008), <http://www.tsunamictechnologies.com>
5. Wu, F., Zhang, L., Huberman, B.A.: Truth-telling Reservations (July 17, 2005), SSRN: <http://ssrn.com/abstract=771804>

A Computational Trust Model for E-Commerce Systems

Mehrnoosh Divsalar¹ and Mohammad Abdollahi Azgomi²

¹ ICT Group, E-Learning Center, Iran University of Science and Technology, Tehran, Iran

² Department of Computer Engineering, Iran University of Science and Technology,
Tehran, Iran

divsalarmehr@yahoo.com, azgomi@iust.ac.ir

Abstract. Today, one of the most important factors of successful e-commerce is creating trust and security. Undoubtedly, in absence of trust as a key element of online business communications, e-commerce will face a lot of challenges. Assessing trust, especially in the beginning of a commercial relation and generally, formulating the trust is a complex and difficult task. Trust may not be defined in a special framework even for customers and e-vendors. The reason is that for giving a definition for trust, different people, based on their cultural and familial background, believe in different features and dimensions in their mind. In the model proposed for computing trust in e-commerce, first of all based on past research three general fields are envisaged. Then, a computational model has been presented for evaluating trust. Finally a case study is presented to illustrate the proposed trust evaluation model. Future works are also discussed in addition.

Keywords: Trust model, trust evaluation, e-commerce, trust manager agent (TMA).

1 Introduction

Trust is very important in individual and social life. In absence of trust, many of social aims will be unachievable and inaccessible. E-commerce has not yet attained its full capacity mostly because of the absence of trust in relations of the agents of the transaction. Based on the research presented in [11], trust has several properties: - Trust is context dependent. -Trust describes the degree of belief in the trustworthiness of a partner. - Trust is based on previous experience. - Trust is not transitive. - Trust is subjective. Generally, if people get certain that electronic buying and selling will save the time and costs, their trust will promote. In this line, the models offered for assessing trust in e-commerce are very useful.

This paper will establish a trust evaluation model in e-commerce based on analyzing conception and component of trust in e-commerce. This trust evaluation model has two features: First, trust is evaluated dynamically and with a change in any of trust variables, the new value is calculated for trust. Second, it is optional. It means that it is capable to make partial study of trust on contexts presented in general dimensions of trust. The advantage of partial study of trust is that the trust parameters are mostly personal and each agent has his own beliefs which can serve as basis of trust

evaluation. The remainder of this paper is organized as follows. The envisaged dimension for trust is presented in Section 2. Main factors that influence the trust in e-commerce also are analyzed in this section. Computational Trust Model is introduced in Section 3. Definitions of the model and the components that are used in this model and also the trust management agent are presented in this section. In Section 4 a case study is given to illustrate the proposed trust evaluation model. Section 5 discusses conclusions, limitations and future works.

2 The Envisaged Dimensions

Based on past investigations the dimensions proposed in this study which will be a base for presenting the computational trust model have been presented as follows. As seen, three general backgrounds have been viewed for trust in e-commerce. They are:

- Institutional Trust
- Technology Trust
- Trading Party Trust

Institutional trust includes laws and regulations which facilitate commercial transactions. It consists of two groups, namely structural assurance and certificates of third parties. This category of trust mostly is focused on B2B transactions and structurally decreases distrust and uncertainty. It is like a state in which we do not trust the other party and use structures and control mechanisms as substitute of trust and apply a series of supervision mechanism. The testimonies of third parties which assesses an online trade in respect of security or organizational and procedural and then releases the results on web site for being visited by probable customers is an example of those mechanisms.

Nowadays, there are organizations which assess the commercial transactions in accordance with their special standards. The commercial companies coordinate their commercial processes with those standards and then receive approval certificate of those organizations. In [6], the infrastructural dimensions of an institutional trust have been examined. The structural guarantees include agreements and instructions dealing with organizational processes and the standards and procedures which lead to success of different services such as transportation, delivery, and processes.

Technology trust may be studied from two angles of view:

- Web site
- Security Infrastructure

Web site is the first thing the customer faces (especially for a person who visits an e-commerce site for first time) and can have desirable effect on purchase decision of the visitor. The particulars which may attract the attention of the consumer to the web site include appearance of the web site (such as beauty, convenience of using, accessibility, and speed) and the site's technical specifications (such as being up to date in respect of software and hardware) as well as transparency and standardization of the information. For instance, the complete specification of the product and company, the company's logo and trade name shall be given in the web site and the laws and contracts shall be observable. Off line attendance or contact (Tel No., Email address, fax No.) too is effective in winning the trust of the customer.

The security infrastructures which provide safe connective mechanisms deal with issues such as confidentiality and security of information related to persons and organizations. Generally, the mechanisms such as integrity, confidentiality, non-denial, access control, etc., are among the security mechanisms [2]. Because technical ability and security services of websites play important role in customer trust absorption and website is the first demonstrator of vendors' abilities, increase trading party trust.

As past investigations in [1], [2] indicate and especially in B2B transactions, three categories of trust, that is, competency, benevolence, and predictability of the trading party are considered.

In trading a product or a service one very critical matter is that satisfaction of the product may become the root of many later decisions. Since in doing a transaction is acquiring a special product or service so trust to the product is very critical and satisfaction of the product is one of the prerequisites of trading party trust. For evaluating the product trust a signature of one reliable reference may be used or the product trust index based on satisfied or unsatisfied customers may be given along with the specifications of the product.

3 A Computational Trust Model

Samples of models for evaluating trust in e-commerce have been presented in [3], [5], [10], [18], [19]. In this section, a solution for evaluating trust is proposed. The solution has two important features: 1) It studies trust in a dynamic way and by changing each of the variables of trust the new value of trust is calculated. 2) It is optional, that is, the consumer may measure the properties or features that he/she wishes to know their value. The importance of this aspect rests in the fact that since trust is a multi-dimensional issue, different individuals may concentrate on different dimensions of that and thus it is better for them to measure the features they attach importance to. The important point is that distrust is qualitatively different from trust. Lack of trust does not mean distrust because lack of trust may be originated from non-existence of trust-related information. In [15, 17], trust and distrust and their preliminaries and consequences in e-commerce have been studied. In the proposed model, an agent may be considered trusted in one context and may be distrusted in another context because of lack of needed information or bad history in past transactions.

Generally, there are two approaches to trust management. One is the centralized method in which for storing and searching the data some perfectly concentrated mechanisms are used. These mechanisms have one main central database for assessment and there is a central control. The second approach is the distribution methods. In this method, there is no central database and thus none of the agents have a general understanding of entire system. Assessment and management of trust are done locally by number of agents without using any central control.

In proposed model, for evaluating trust we assume a reliable reference which is trusted by both vendor and consumer. This reference is in fact the principal core of the evaluating system and may control the commercial processes too.

We call this trusted medium of the model "trust manager agent" or TMA and suppose that for specific subsets of network nodes there is a node with duty of trust management and the other nodes have assigned the duty of other subsets trust management. So, the trust management is distributed between trade agents.

It is possible that TMA enjoys the help of other references for completing its knowledge. In this case, the techniques of trust propagation shall be used. Samples of trust propagation models are in [13], [14], [16], [12]. In addition to that, since the intrinsic nature of trust is being temporary the agent at each intention of purchase will have opportunity to enjoy this TMA to complete its knowledge.

TMA's not only receive the information of transactions from the nodes of the set but also share information with other TMA's if requested. Neighboring TMA's know each other and confirm the validity of each other through cryptography.

In TMA database there is a table for maintaining the information of agents. The calculated value of trust in each context is allocated for each agent. The table is updated dynamically with implementation of transactions or propagation the information from other nodes. TMA saves the feedback of agents in its database too. The manner is such that the e-vendors send some information to TMA periodically about their transactions. Also the consumers reflect their assessment of their trading party. The reflections are kept by TMA.

We have following definitions:

1. Agent is an active entity who plays a role in trust model and fulfills different activities toward meeting the needs. Vendor, mediums, and trusted parties are examples of agents: $Agent = \{a_1, \dots, a_n \mid Agent(a_i), i \in [1, m]\}$
2. Transaction is a unit of action which is entangled with two agents. The result may be failure or success. The result of a transaction can be a value in $[0, 1]$ interval which points at the degree of success of the transaction.
3. The context of trust as was stated and considering the request of the agent, is defined as: $Context = \{I, T, Tp, Be, Co, Pr, G\}$

I: *Institutional Trust*, T: *Technology Trust*, Tp: *Trading Party Trust*, Be: *Benevolence*, Co: *Competence*, Pr: *Predictability*, G: *Goods*

In relation to trading party trust we have defined four context, namely, Be, Co, Pr and G. for future calculation, trading party trust itself and two other dimensions too have been included in the context.

One of the advantages for existing TMA is that an agent may need consultation only in one context. For instance, for a consumer, the trustworthiness of trading party is certain but trustworthiness of his technology is not perfect then consumer consults with TMA. We have $T(x)$ function whose ($x \in Context$) makes clear that in which context TMA assesses the trustworthiness of a_j .

3.1 Trading Party Trust

3.1.1 The Supplier Trust

One of the definitions proposed for trust is following definition: Trust is the subjective probability by which person A expects person B to fulfill an activity assigned to him/her in direction of the welfare of person A. The important element existing in this definition is "activity". Activity is in contrast with information and we can trust a man through considering a specific activity by that person and also we can trust a man based on the information provided by him/her. In computation model both of activities and information are envisaged. We have relation (1) for trade party trust evaluation.

$$Ta(Tp) = \sum_{i=1}^{I(a)} \alpha * S(i) / I(a) + \sum_{i=1}^{I(a)} D(i) / I(a) + \sum_{i=1}^{I(a)} B(a, i) / I(a) + \sum_{i=1}^{I(a)} G(a, i) / I(a) \quad (1)$$

Where $I(a)$ is total number of transactions of agent a and i indicates the transaction.

The relation $\sum_{i=1}^{I(a)} \alpha S(i) / I(a)$ is based on the past transactions history of the agent. $S(i)$

are transactions completed successfully and with satisfaction of the customer. Any successful transaction is graded "1". Success in transactions may be determined by the feedback mechanisms of e-commerce sites. Since feedback mechanisms play a very important role in studying the history of commercial transactions we attach positive point to having feedback mechanism in commercial sites. α is a coefficient which is used in calculating the final value based on the importance degree of satisfaction

and success. In formula (1), we have purposed the $\sum_{i=1}^{I(a)} D(i) / I(a)$ for information sharing factor that indicate benevolence of trading party. The relation shows the total number of files downloaded by e-consumers for the total number of transactions of agent a and $D(i)$ shows downloading in i^{th} transaction and is determined with due regard to transactions and according to the standard drawn up by TMA, for example we can envisage that each downloading receives "0.5" as grade. Maximum value of information sharing factor in integrated trust value is 0.5.

The $\sum_{i=1}^{I(a)} B(a, i) / I(a)$ relation shows the positive behaviors of the seller agent. Positive behavior is same as preference behavior such as giving prizes or prominence to some consumers or receiving the amount in installments that the seller may consider for some customers. $B(a, i)$ means the positive behavior of the vendor toward consumer in i^{th} transaction. Since this factor does not apply to all customers so it does not apply to all transactions and its value is between zero and one. Legal commitment or $G(a, i)$ are the guarantees given by the e-vendor to its trading party in i^{th} transaction and the $\sum_{i=1}^{I(a)} G(a, i) / I(a)$ relation calculates the number of commitment of the supplier

in all completed transactions. The commitments are not usually fulfilled for low-value transactions and in limited time. In fact, they change with the conditions of transaction and type of the service. We envisage 0.5 as maximum value of this factor in integrated trust value.

The above mentioned points show the behavior of online merchant and study transactions history of e-vendor. The researchers have found out that trust has positive relation with disclosure of information and case studies made on interpersonal exchanges confirm that trust is a pre-requisite for self-disclosure because decreases the risk existing in disclosing the personal information [9]. A very important issue in the field is the truthfulness and correctness of the information given to TMA by the agents sincerely and without intention of deceit. This means risk in the transaction that shall be assessed based on the value of transaction and its expected revenue. For preventing fraud, the customers are asked to reflect their opinions related to above matters after each transaction. It needs to be pointed out that since in proposed model trust context is optional, the agent who sends request to TMA may send the request on

special field. He/she may request the assessment of benevolence alone. In this case, in relation 1 the formulas related to past history and legal commitment become inactive and the formula will change to:

$$Ta(Be) = \left(\sum_{i=1}^{I(a)} D(i) + \sum_{i=1}^{I(a)} B(a,i) \right) / I(a) \quad (2)$$

When assessment of trading party competence is requested the formula will be

$$Ta(Co) = \sum_{i=1}^{I(a)} \alpha * S(i) / I(a) \quad (3)$$

If the agent wants to assess the trading party predictability, TMA will calculate the Formula (1) in the following manner:

$$Ta(Pr) = \sum_{i=1}^{I(a)} G(a,i) / I(a) \quad (4)$$

3.1.2 Product Trust

In this study it has been supposed that trust to the product affects on trading party trust. When the consumer makes its first electronic purchase and is not familiar with the products offered by the e-vendor, for investigating product trust he/she sends to TMA a product trust request.

TMA is capable to evaluate product if it has been signed by a trusted party or if the product has satisfaction index. Organizations which prove their skill in producing high-quality goods and services achieve high rank in trustworthiness scale. If the total number of goods and services offered for sale by agent a is showed by $G(a)$, the grade of product trust will have following form. In this relation, $S_g(a, g)$ is signed goods and g is the products. Any signed good is graded "1". Therefore maximum value for product trust will be "1"

$$Ta(G) = \sum_{g=1}^{G(a)} S_g(a, g) / G(a) \quad (5)$$

The above relation can be added to general relation of $T_a(Tp)$. TMA is able to verify the genuineness of the product signature.

3.2 Technology Trust Evaluation

Web site is the entrance gate of e-commerce. The appearance of web site may have considerable effects on continuation of purchase because beauty and efficiency of website are features which non-professional users too can realize. For evaluation the website, TMA makes a questionnaire accessible to the consumer and when the questionnaire is filled in TMA gives points based on its own policy and makes the points of the product accessible to the agent. For different agents, the points of the web site may differ because based on personal factors, tastes, and social and contractual status of persons and or organizations the points may differ. Therefore, the points are determined based on the completed questionnaire.

The questionnaire has two sections: website and information. Let's suppose that the points considered by TMA are as follows: The points of website section are 0.4 and the points of information section are 0.6. If all points are won, TMA points 1 for technology trust. The security infrastructure of website cannot be checked in this way and for evaluating it, received certificates should be examined. This examination has been made in section 3.3 accompanied by examination of other certificates.

3.3 Institutional Trust Evaluation

Traditional commercial communication relies on legal commitments in primary communication making phases which are used for management on uncertainties. Because of non-transparent legal environment of internet, this does not work well in online environments [6]. To decrease uncertainties of online environments, a great number of third parties are active for helping the transactions of the organization and special structures, guarantees, and laws are used. Institutional trust is like a belief the trustor has on the security of a situation because of the guarantees, safe networks and other efficient structures. In studying institutional trust we assume two general components: 1) Institution-based third party trust 2) structural guarantees and processes.

In first form, the mediums define the organizational supplies but in second situation, the supplies are directed in two-way relation and are not arranged by the medium (such as legal agreements between commercial agents and the standards like commercial messages, product information, infrastructure of information technology, etc.). The structural guarantees may be provided as they are ordered for meeting common needs of the agent. Reference [6] names bilateral institutionalized trust substituting the structural guarantees and divides it into three groups. These three groups are structural guarantees, facilitating conditions and situational normality. The states predicted in the contract between trading agents shall be prepared in accordance to a procedure or standard which TMA is able to evaluate. The relevant contracts shall be signed by parties and shall bear signature of TMA and time stamp. TMA keeps these signed documents in its database. If the consumer wants to know about institutional certificate of e-vendor, TMA make evaluation in accordance to relation 6.

$$T_{\alpha}(I) = I(C) * \alpha \quad (6)$$

Where C stands for certificates, $I(C)$ is the number of received certificates. The security certificates too are evaluated by this relation. Generally speaking, the point of the set of the certificates received by online company is determined based on the importance of the certificate which has been pre-defined for TMA. The degree of importance is determined by α . The maximum point we have chosen for that is 3.

4 A Case Study

For making the trust calculations we take an Internet service providers company in Iran, which offers services for finding access to local and world networks. We refer to this company as ANET. Since, it is not possible to have access to the information of this site concerning the sales (for becoming aware of the number and items of transactions and services of ANET company) we resort to knowledge of persons who have

had interactions with this website. We call these persons A, B, and C. A is a person who had commercial relations with company as a customer for about three years. A uses the Internet services of this company in framework of an ADSL contract which is extended often monthly. B has used the services of the company for six months in form of hourly subscription to dial up Service, and C has used ADSL ANET services for one year. First we consider trust to trade partner whose calculation has been presented in relation 1. In recent two years, totally, A has had 24 interactions with this online store and has been pleased with all these transactions except one. So, the result

of the relation $Ta(Co) = \frac{\sum_{i=1}^{I(a)} S(i)}{I(a)}$ is 0.96. This value for B and C is 1 and 0.92,

respectively. Concerning information sharing factor, the information related to access contracts and other services are available in a complete way. These information has been uploaded in forms and are accessible to the customers. Considering the activities of this company and the needs of customers, full point may be given to information sharing factor. Points that may be given for positive behavior factor is for two services: "little more" service (permits the users to have access to ANET services during a 5-day period after expiry of the contract period without paying additional fee) and giving rewards to any customer who introduces new customer to the company. A, in his 24 interactions with this website, has used "little more" service of ADSL for 23 times and had one reward for introducing new customer. Therefore, the positive behavior point of the company with customer A is 1, with B is 0.83 and with customer C is 0.92.

The obligations of ANET Company have been enumerated in the contracts for ADSL and Dial up services. One obligation relates to installation and commissioning the connection to Internet at "entrance" of modem for the subscriber. The second obligation relates to granting access right in each extension of the contract and rendering support services to the subscribers through telephone. It may be said that considering the need of the customer in respect of having access to the services of the company, the Company will receive entire points of legal obligations or $G(a, i)$.

Concerning product trust, generally speaking, ANET Company offers seven services. Though, according to the solution proposed in computational model, many of the customers are fully satisfied with this service but the services of the company do not bear signature of any special reference and no product satisfaction index has been registered for them and therefore, no point can be scored for product trust.

What pointed out above is all related to appraisal of the trust to trade partner. For appraising trust to technology, a questionnaire was send to persons. Based on opinion of A, totally 0.95 point was scored to the technology of the company. B and C gave points 0.9 and 0.85 respectively to online company for this aspect of trust. Concerning organizational trust, there is no certificate of special reference for this company and so no point can be allocated to the Company for this type trust.

Since the competence of the company is an important element appraised with due regarded to satisfaction of the customer, high importance is attached to that when points are given to trust. Therefore, in calculating the general scores of trust we have applied coefficient 2. Considering above points and as presented in calculative model, we will have following table for trust to ANET Internet virtual store. Results indicate that evaluating of A, B, C is almost similar and is in near range.

Table 1. Contexts of trust and their values

Trust dimensions	Trust context	A	B	C	Max
Trade party trust	competence	1.92	2	1.84	2
	Benevolence	1.5	1.33	1.42	1.5
	Predictability	0.5	0.5	0.5	0.5
	Products	0	0	0	1
Technology trust		0.95	0.9	0.85	1
Institutional trust		0	0	0	3
Total		4.87	4.73	4.61	9

5 Conclusions

In this paper, first of all three general dimensions for trust in e-commerce have been presented which is the base of later calculations of trust. The manner is such that some trusted parties (TMA) have been envisaged distributed in the network. TMAs have duty to evaluate, maintain and propagate trust for agents. The presented computational model has the ability of evaluating all contexts of the trust.

We present a case study to display efficiency and rationality of the proposed computational formula of trust. We didn't have access to e-suppliers sales and transaction information because of security problems. This limitation and the experiment condition lead us to do our case study with three consumers who purchase from a common e-vendor. In future, we intend to do a case study with different supplier and different shopping with more experiment and sufficient data to demonstrate the advantages of the proposed model. We also intend to implement a prototype of the model in a sample e-commerce system. Analyzing the detail of institutional trust especially in structural guarantee and process and presenting a solution for evaluating it in our computational trust model is another work we will try to do in future. Also we intend to design and assess a model for dimensions we discussed in this study and specify relations between dimensions.

References

1. Ratnasingam, P., Pavlou, P.A., Tan, Y.-H.: The Importance of Technology Trust for B2B Electronic Commerce. In: 15th Bled Electronic Commerce Conf. eReality: Constructing the eEconomy Bled, Slovenia (2002)
2. Ratnasingam, P.: Trust in inter-organizational exchanges: a case study in business to business electronic commerce. *Decision Support Systems* 39(3), 525–544 (2005)
3. Schlager, C., Penul, G.: Trust Modelling in E-Commerce through Fuzzy Cognitive Maps. In: 3rd Int'l Conf. on Availability (2008)
4. Huiwen, T., Feng, X.: A Study of Consumer Trust Formation in Electronic-Commerce. In: Proc. of the Int'l Symposium on Electronic Commerce and Security (2008)

5. Zhang, Z., Wang, Z.: Assessing and Assuring Trust in E- Commerce Systems. In: Computational Intelligence for Modelling, Control and Automation, Sydney, Australia, p. 124 (2006)
6. Pavlou, P.A., Tan, Y.-H., Gefen, D.: The Transitional Role of Institutional Trust in Online Interorganizational Relationships. In: 36th Hawaii Int'l Conf. on System Sciences, HICSS 2003 (2002)
7. Tan, Y.-H., Thoen, W.: Formal Aspects of a Generic Model of Trust for Electronic Commerce. In: 33rd Hawaii Int'l Conf. on system science (2000)
8. Xiong, L., Liu, L.: A Reputation-Based Trust Model for Peer-to-Peer E-Commerce Communities. ACM, New York (2003)
9. Metzger, M.J.: Privacy, Trust, Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication* (2004)
10. Mahala, D.: Trust Metrics, Models and Protocols for Electronic Commerce Transactions. In: 18th Int'l Conference on Distributed Computing Systems (ICDCS 1998), p. 312 (1998)
11. Abdul-Rahman, A., Hailes, S.: Supporting trust in virtual communities. In: Sprague Jr., R.H. (ed.) 33rd Hawaii Int'l Conf. on System Sciences, pp. 1769–1777. IEEE Computer Society Press, Los Alamitos (2000)
12. Androutsellis-Theotokis, S., Spinellis, D., Vlachos, V.: The MoR-Trust distributed Trust Management system: Design and simulation results. *ENTCS* 179, 3–15 (2007)
13. Karl, A., Zoran, D.: Managing Trust in a Peer-to-Peer Information System. In: 10th Int'l Conf. on Information and Knowledge Management, pp. 310–317. ACM, New York (2001)
14. Chen, R., Yeager, Poblano, W.: A distributed trust model for peer-to-peer networks. Technical Report, Sun Microsystems (2001)
15. Jinsook, C.: The mechanism of trust and distrust formation and their relational outcomes. *Journal of Retailing* 82(1), 25–35 (2006)
16. Marti, S., Garcia-Molina, H.: Limited reputation sharing in p2p systems. In: 5th ACM Conf. on Electronic commerce, USA (2004)
17. McKnight, D.H., Vivek, C.: Distrust and Trust in B2C E-Commerce: Do They Differ? ACM, New York (2006)
18. Qin, Z., Tian, B.: A Trust Evaluation Model for B2C E-commerce. In: IEEE Int'l Conf. on Service Operations and Logistics, and Informatics (SOLI 2007), pp. 1–6 (2007)
19. Nafti, S., Meziane, F., Khairudin, K.: A Fuzzy Trust Model for E-Commerce. In: 7th IEEE Int'l Conf. on E-Commerce Technology, CEC 2005 (2005)

A Recommender for Active Preference Estimate

Andrea Malossini and Paolo Avesani

Fondazione Bruno Kessler

Abstract. Recommender systems usually assume that the feedback from users is independent from the way a suggestion is presented. Usually the learning effort to estimate the user model is aimed at recognizing a subset of items according to the user preferences. We argue that acceptance or rejection is not independent from the presentation of the recommended items. Shaping a recommended item means to decide what kind of presentation can be more effective in order to increase the acceptance rate of the item suggested. In this work we address the challenge of designing an active policy that recommends a presentation of an item, i.e. which features use to describe the item, that will produce a more informative feedback from the user. The main contribution of this paper is a first principled design of an algorithm to actively estimate the optimal presentation of an item. An experimental evaluation provides the empirical evidence of the potential benefits.

1 Introduction

A basic step in recommender systems is the estimate of user model. Learning the user preferences is the premise to deliver personalized suggestions and to design adaptive information filters. The challenge is how to support the user to make a choice when the space of items is so huge that an exhaustive search is not sustainable [1]. The usual strategy is to take advantage from the feedbacks of the users to learn their preferences while are browsing a sample of items [2].

The elicitation of preferences takes place as an iterative process where the users get explicit what kind of items like most or dislike. Conversational recommender systems [3,4] were introduced to manage this process as a dialogue to shorten the estimate of the user preferences and consequently the recommendation of a compliant item. Step by step the goal is to converge on a subsample of items that includes the user's choice.

Usually this kind of approaches share the simplifying assumption that the choice of the user is independent from the representation of the items. Let consider a scenario where the purpose is to choose a movie. The choice of the user may be affected whether the movies is described by features like title, years and genre, rather than title, actor and rating. What might happens is that the user would accept our recommendation in the first case while would reject the other even though the item, e.g. the movie, is the same [5].

The objective of this work is to address the problem of learning what features affect the user acceptance of a given item. The working hypothesis is that the

process of decision making is biased by the information encoded in a subset of features. If such information is available the user may perform the assessment of an item and to decide whether s/he likes or dislikes. In the other case the user can't evaluate an item and the feedback may be perceived equivalent to a dislike option as above. We prefer to refer to this feedback as a third option called unknown. Therefore we may distinguish among three kinds of feedback: like, dislike and unknown.

It is worthwhile to remark that the two answers, dislike and unknown, are not equivalent. If we received an unknown feedback, it doesn't necessarily mean that the recommended item isn't suitable to the user. May be the same item, recommended with a different set of informative features, might be accepted at the next stage. If we received a dislike feedback the recommended item is equally rejected by the user but we may learn what kind of features affect the decision.

The ultimate goal is to design a recommender system that helps not only to select compliant items (items that users might like) but also to shape effective items (items that users might perceive as appropriate). Our approach is oriented to active sampling. The idea is to recommend those features that will produce more informative feedbacks from the user [6]. This work differs from other approaches based on active policy [2]. Usually the task of estimate the user model aims to learn the threshold values that discriminate between acceptance and rejection. We focus the estimate task to learn only what are those features that are relevant to take a decision. It might be conceived as a shallow user model that requires less effort, e.g. short interactions.

The motivation of this work derives from a different perspective on e-commerce where the selection of an item is not driven only by the preferences of a single user but more often is matter of negotiation [7] between two parties: producer/consumer, buyer/seller, tenderor/tenderee. While the buyer is interested to find something s/he likes among huge range of items, the seller aims to recommend items that provide larger revenues. In this scenario, much more competitive rather than collaborative, the general habit is to reduce the amount of elicited information since it might represent a competitive advantage for the counter part. In the following we will refer to a framework of bilateral negotiation [8].

The main contribution of this paper is a first principled design of an algorithm to actively estimate user preferences. An experimental evaluation provides the empirical evidence of the potential benefits.

2 Formal Definition

In this Section, we formally define the behavioural model of users playing bilateral negotiations.

We model a user using a workflow which encodes the sequence of actions to be executed; in particular, we define three functions, the *pref* function, the *shape* function and the *like* function, which encode the user's preferences and decision making procedures. While *pref* and *like* functions are supposed to be stationary and defined at the beginning of the negotiation, *shape* is conceived

as an adaptive function whose definition evolves through the negotiation by a process of learning. The way to shape an offer change according to the feedback collected from the counter part. After each stage of interaction the hypothesis on issues relevance is revised with respect to the evidence collected so far by the rejected offers.

The *pref* function, given the offer number in the negotiation, returns an item from the ordered preference list of the user. The user is in charge of selecting an item for the next offer. We indeed suppose that each user has a fixed list of items ordered in accordance with his utility. Formally, let U be a set of users, T be the item set, and k be the maximum number of offers (for each user involved in such negotiation) before the negotiation expires, we define, for each user $u \in U$, the *pref* function as $\text{pref}_u : I_k \rightarrow T$ where $I_k = \{1, 2, 3, \dots, k\}$.

The *shape* function, given an item, returns an offer to send to the counterpart. Practically, it selects which features of an item to include in the offer. In the next Section we give a formal description of the item set and the offer set. Formally, let O be the offer set, for each user $u \in U$, we define the *shape* function as $\text{shape}_u : T \rightarrow O$

The *like* function, given an offer, return a label (answer) which is related to the evaluation. Let $A = \{0, \perp, 1\}$ be the set of all possible answers to an offer, for each user $u \in U$, we define the *like* function as $\text{like}_u : O \rightarrow A$. The like function is invoked every time we need to evaluate a new opponent's offer; we emphasize that both the values of the issues and the presence or absence of some issues play a role in the definition of this function.

After the definition of the *pref*, *shape*, and *like* functions, we can focus on the user's workflow. Firstly, A calls its *pref* function which returns the item to propose. Then, A calls its *shape* function which provides a subset of issues to include in the offer. Finally, A sends the offer to B. Once the offer is received, B calls its *like* function to decide if accept the offer. If B rejects it, B will call its *pref* and *shape* functions to generate a counter offer, that will then send to A. At this point A will evaluate it using its *like* function. The negotiation go on until one of the two users accepts an offer or the time expires.

3 Computational Model

In this Section we develop an active learning algorithm that will explore the space of agreements by selecting a set of issues that are "relevant" for the opponent or yield an high "benefit" when assessing the opponent's preferences. This search will be carried out inside the present negotiation and will react according to the counter-offers proposed by the opponent.

As a first step, let us define the item set T ; we suppose that each item $t_i \in T$ can be described as a set of pairs of the type

$$t_i = \{(\text{attr}_1, a_{i1}), (\text{attr}_2, a_{i2}), \dots, (\text{attr}_n, a_{in})\},$$

where n , the number of features is fixed, and we denote with a_{ij} the value assumed by the j -th feature of the i -th item. The shape function defined in

Eq. 2 takes an item t_i and shapes the offer by selecting a subset of informative features. Formally this can be accomplished by setting some of the values a_{ij} to the special value NA (not available, i.e. missing value). Hence, we obtain the offer $o_i \in O$ as

$$o_i = \{(\text{issue}_1, v_{i1}), (\text{issue}_2, v_{i2}), \dots, (\text{issue}_n, v_{in})\},$$

where

$$v_{ij} = \begin{cases} a_{ij} & \text{if the } j\text{-th feature is an issue,} \\ \text{NA} & \text{otherwise.} \end{cases}$$

Essentially, in order to define the shape function we need to attach a score to each item's feature. This score can be obtained by using a variant of the active learning algorithm for feature relevance [9]. In order to adapt such algorithm we need to introduce some definitions in the following.

As stated in the previous Section, we assume that the negotiation will end after a certain number of offer exchanging k . We can, then, define the *user's content matrix* at step m (i.e. after m offers exchanged) as

$$V_m = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \dots & v_{mn} \end{pmatrix}, \quad (1)$$

where $m \leq k$. Analogously, we can construct a similar matrix, the *opponent's content matrix* at step m , which contains the issues' values of the offers received by the opponent. We denote such matrix as W_m , where $m \leq k$. Note that these matrices contain also missing values. As long as we receive and make new offers, the matrices will increase row by row.

V_m will encode our strategy for making offers, whereas W_m will contain features' values which are of interest for the opponent [9]. During the negotiation process, the user is able to collect a lot of information related to the opponent's preferences analysing the content of the received offers. These data can be used to estimate probabilistic models by considering each issue as a discrete random variable with a multinomial distribution (for continuous issue we can think of transforming the original continuous distribution to a discrete distribution, essentially because we expect to have very few values for each issues, hence a continuous distribution will bring no information).

Let us define then the *issue matrix* F_m and the response vector A_m at step m as

¹ In fact, we argue that the opponent, when composing an offer, will use item of interest with potentially interesting features' values.

$$F_m = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1n} \\ f_{21} & f_{22} & \dots & f_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m1} & f_{m2} & \dots & f_{mn} \end{pmatrix} \quad (2)$$

$$A_m = (a_1, a_2, \dots, a_m)^T \quad (3)$$

where

$$f_{ij} = \begin{cases} \Pr(\text{issue}_j = v_{ij} | W_{i-1}) & \text{if } v_{ij} \neq \text{NA} \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

represents the probability that the issue issue_j be equal to the correspondent value in the i -th offer we made given the content matrix W_{i-1} and a_i the i -th response to our i -th offer (which can assume values in $\{0, \perp\}$). Let g_j be a *relevance* function² of the issue issue_j computed using the matrix F_m , then, we can compute the *benefit* of issue issue_j in a new offer as⁹

$$B(\text{issue}_j) = \sum_{a \in \{0, \perp\}} \left[g_j(F_m, \text{issue}_j = v_{(m+1)j} | a_{m+1} = a) - g_j(F_m | a_{m+1} = a) \right]^2 \cdot \Pr(a_{m+1} = a | F_m), \quad (5)$$

where $v_{(m+1)j}$ is the j -th issue's value of the new offer we are willing to make. This strategy differs from a similar work [10] where the heuristic is to maximize the expected utility, while here the criteria is to maximize the average increment on expected utility.

The relevance function and the benefit function are used in the shape function to attach a score to each feature of an item. In particular, we could use this information to:

Exploration. Select the features which possess a high benefits B , which means that including those features in the offer would allow for a smart assessing of the opponent's preferences;

Exploitation. Select the features which possess a high relevance g , which means that using those features in the offer would allow the opponent to take a decision using his relevant features, hence minimizing the chance of the offer to be rejected.

In the following we restrict our attention to exploration strategy. We design a recommender engine that exploits the computation of B to support an *Active Offer Shaping* (AOS). AOS recommends features that are ranked at the top by B function.

² As a standard measure of relevance we use the mutual information, but other measures could be used, depending on the context.

4 Empirical Evaluation

As reference example to sketch the experimental framework we refer to a scenario of cooperative rental of a DVD movie. Two users engage an iterative process to select what kind of DVD movie to watch together.

To generate a data set suitable for the simulation of such a kind of bilateral negotiation as described in the previous sections, we start from the *Movielens*.

According to the simulation model adopted in Section 2, we need to model each user by a *pref* function, defined by the rating of *Movielens* users, and the *like* function, defined as the category of movies that discriminate between high and low ratings. These two functions completely describes the “virtual” users that will engage a bilateral negotiation to achieve an agreement on a movie that both of them like. The *shape* function is alternatively implemented as a random choice or as adaptive strategy according to the definition of Section 3.

We designed a simulation using a sample of 1500 user pairs to run negotiations. At each negotiation we have one user which always plays the random strategy, whereas the other one alternatively uses the random strategy or the AOS recommender.

From the simulations we obtained that the rate of success for random strategy is 0.81 (0.01), whereas for the AOS strategy is 0.93 (0.01). There is an increment of 12% (180 negotiations) in the number of successful negotiations when using the AOS recommender. It is interesting to analyze in detail the disaggregated results, as reported in Table 1.

The first row reports the mean number of counts (and the corresponding standard deviation) when adopting the AOS recommender the negotiation fails, whereas, using a random strategy the negotiation succeeds. The second row depicts exactly the opposite, when using the AOS recommender is beneficial and random strategy does not. The difference between these two counts is about 170 which explains the increase in the rate of success when using the AOS strategy. The fourth row depicts the situation where both the strategies produce a successful negotiation, but using AOS a faster agreement is reached. The fifth line is the opposite. Here we can notice that, typically, there are about 700 negotiations where the random strategy produces faster successful negotiations.

Notice that the remarkable performance of the random strategy is due to the fact that the problem, as is, is easy to solve; in fact, among the 18 features

Table 1. Mean number and standard deviation of counts for the different situation that may arise during the negotiations

AOS	RND	%	μ	σ
Unsuccessful	Successful	0.05	85.9	8.1
Successful	Unsuccessful	0.17	257.2	13.1
Unsuccessful	Unsuccessful	0.01	20.4	4.9
Faster	Slower	0.26	400.0	13.6
Slower	Faster	0.46	696.0	20.7
Same speed		0.02	41.3	6.2

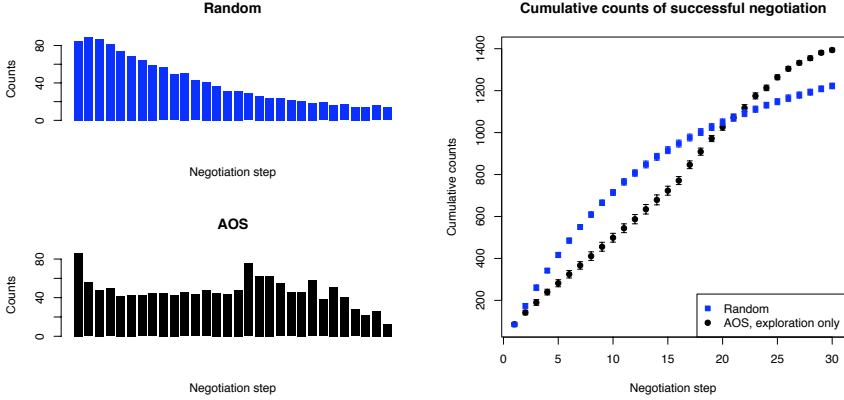


Fig. 1. On the left hand side the differences between negotiations with the AOS recommender and random strategy; on the right hand side the distribution of the average negotiation lengths with the AOS recommender and random strategy

available, you have only to find the right feature (probability $1/18$) and its value, which may assume only two values, 0 or 1 (with probability which depends on the distribution of the genre). The problem can even be solved more efficiently by probing the same feature once one obtains as answer `DISLIKE`. However when the scenario becomes more complicated (for example when the user model implies a combination of right features/values and the cardinality of the features is higher than 2, or possibly continuous) we expect that the random strategy would perform very poorly, whereas, using the AOS recommender and its exploration/exploitation capabilities we expect to obtain far better results. The extremely simple scenario used, nevertheless, is useful for analyzing the behavior of the algorithm used for active learning. Another effect of this simple scenario is that the number of answers of type `DISLIKE` one obtains is very low (once discovered the right feature, being it binary, there is an high probability of ending the negotiation), and the computation of the relevance is deeply affected by it.

Another interesting results is the comparison of the speed in obtaining a successful negotiation.

In Figure 1 we show, in the left part, the histograms of the mean counts of successful negotiations with the random strategy and with the AOS recommender. As you can see, the histogram of random strategy is a decreasing function of the offer number, as expected, whereas, the AOS histogram presents a peak around the 20th step. The first peak in the AOS histogram is due to the fact that the first offer is shaped at random, not having any useful information to use in the active feature sampling procedure.

In the right part of Figure 1 we plot the cumulative counts of successful negotiations (and error bars associated), where we can clearly see that in the first part of the negotiation, the random strategy is better, but after the 20th offer the AOS recommender provides additional successes. This indicates that at

certain point in the negotiation, the “exploration power” of the random strategy stops to be fruitful, whereas the AOS recommender keeps being effective.

5 Conclusions

In this paper we proposed a computational model to support active learning of issues’ relevance. The active sampling approach is particularly indicated because it allows to exploit useful information from ongoing negotiations and to intelligently explore the space of agreements. Furthermore, shortening the estimate of opponent’s features relevance allows to speed up the achievement of an agreement in a bilateral negotiation.

Acknowledgments

This work was partially supported by the project 034744 EU-INFOS-IST ONE.

References

1. Burke, R.D., Hammond, K.J., Young, B.C.: The findme approach to assisted browsing. *IEEE Expert* 12(4), 32–40 (1997)
2. Viappiani, P., Pu, P., Faltings, B.: Conversational recommenders with adaptive suggestions. In: *RecSys 2007: Proceedings of the 2007 ACM conference on Recommender systems*, pp. 89–96. ACM, New York (2007)
3. Carenini, G., Smith, J., Poole, D.: Towards more conversational and collaborative recommender systems. In: *Proc. of IUI 2003*, pp. 12–18. ACM, New York (2003)
4. Cynthia, A., Thompson, M.H., Göker, P.L.: A personalized system for conversational recommender. *JAIR* 21, 393–428 (2004)
5. Winterboer, A., Moore, J.D.: Evaluating information presentation strategies for spoken recommendations. In: *RecSys 2007: Proceedings of the 2007 ACM conference on Recommender systems*, pp. 157–160. ACM, New York (2007)
6. MacKay, D.J.C.: Information-based objective functions for active data selection. *Neural Computation* 4(4), 590–604 (1992)
7. Kersten, G., Lo, G.: Aspire: an integrated negotiation support system and software agents for e-business negotiation. *International Journal of Internet and Enterprise Management* 1(3), 293–315 (2003)
8. Telesca, et al.: Open negotiation environment: An open source self-learning decentralised negotiation framework for digital ecosystems. In: *Digital EcoSystems and Technologies Conference (DEST)*, pp. 186–191 (2007)
9. Olivetti, E., Veeramachaneni, S., Avesani, P.: 5. In: *Computational Methods of Feature Selection*. Chapman & Hall/CRC, Boca Raton (2008)
10. Chajewska, U., Koller, D., Parr, R.: Making rational decisions using adaptive utility elicitation. In: *AAAI/IAAI*, pp. 363–369 (2000)

An MDA Based Environment for Modelling and Executing Business Negotiations

Pierfranco Ferronato¹ and Jason Finnegan²

¹ Chief Architect, Soluta.Net, Italy, EU
pferronato@soluta.net

² Waterford Institute of Technology, Ireland, EU
jfinnegan@tssg.org

Abstract. This paper presents the results of a EU FP6 funded project named “Open Negotiation Environment” (ONE), including its features, the strategies, the use of Model Driven Architecture (MDA), the tools, the process used in the development and the key decisions that drove the entire architecture's specifications.

Keywords: Modelling, negotiations, e-procurements, MDA, p2p, peer to peer, metamodelling, agreement.

1 Introduction

Supporting negotiations in a web based environment is a task that many tools and vendors already offer. Some of these also offer the ability to customize to some extent the tender/auction; for example the price and the duration is customizable, as is the fact that it can be a public or a private negotiation. In general these solutions support a rather small number of negotiation types (auction or direct buy usually) and for each some level of customization is allowed. Supporting new behaviors like for example extending the termination of the negotiation by 5 minutes after the receipt of an offer, or adding complete new models like “reverse auction” or generic tenders, is hard without the need to undergo a rather long set of changes in the application code. The ONE project has been entirely conceived for being able to dynamically support new negotiation models.

The project is constituted of two main environments: the Factory, where models are defined, and the Runtime, where models are instantiated and executed. The latter is further defined in two phases: the Setup where negotiations are configured with data and the Execution where they are started and users actually participate to arrive at an agreement.

2 The MDA Approach

In order to support new negotiation models without undergoing the classical develop, compile, test and software deployment life cycle, the ONE consortium has

decided to leverage the Model Driven Architecture(MDA [1]). ONE provides a tool where negotiation models can be created in a visual environment specifying the negotiation protocol, and the underlying information model: with this tool it is possible to model virtually any kind of eBusiness negotiation.

The models defined in the editor are instances of the negotiation modelling language, called the One Negotiation Metamodel (defined by the consortium [2]) that represents the core expression capabilities of the models.

The ONE consortium has defined a Domain Specific Language (DSL) for describing negotiations (a DSL is a specifically tailored language for describing the structure and the behavior of vertical domains). Adopting this approach, primarily aims at raising the level of abstraction at which the software is produced. The main idea is to capture the domain knowledge by means of models and to develop the supporting applications by instantiating these models and generating code, documentation, and other artifacts.

Models are processed by the run-time environment which allows users to tailor them to their needs such as the products information, the services, the start, the end date, and in general the values of the various variables defined in the model. During this phase, called “Setup”, users can also invite users and specify if the negotiation will be public or private.

3 The Implementation Strategy

The project has been conceived for a community of users and as a such it has adopted the EOA approach (Ecosystem Oriented Architecture) [3].

As defined in the DBE project[4] business ecosystem are based on a peer to peer model, where each participant plays both the role of a server and of a client, with the absence of a central control point as depicted in Figure 1 below. This topological model well reflects the nature of social networks: they are self sustained and are regulated by a mechanism where each participant offers part of its infrastructure as a resource to the community: from the business perspective this avoids the “big brother” syndrome.

The execution environment on the other hand is star based topology: the owner of a negotiation is hosting the negotiations itself, while the participants are using their nodes to access the negotiation. Figure 1 below. There are two dotted circles that represent p2p networks of resources. The first, in the upper part of the figure, are the nodes that can run the negotiation as either an owner or a participant in a negotiation. The second network, in the lower part of the figure, represents the p2p network of the Distributed Knowledge Base. The two other clouds represent two running negotiations: 'A' and 'B'. Each cloud is called a “Negotiation Arena” and their life cycle is regulated by each negotiation: they appear and disappear in sync with the negotiations.

The environment, due to the redundant peer-to-peer nature of the topology, is highly reliable and able to remain effective in case of a random failure of nodes.

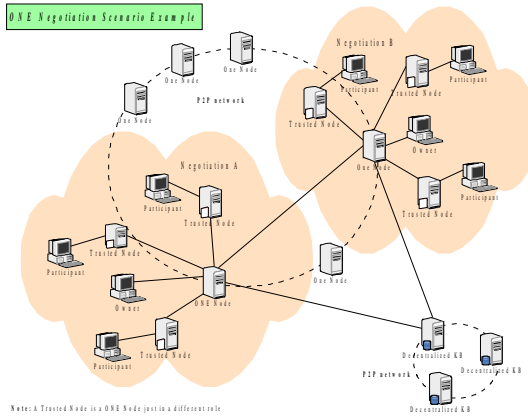


Fig. 1. ONE topological model

4 The Factory Environment

The Factory environment is where models are created, modified and published in the Distributed Knowledge Base. consists essentially of a “boxes & lines” [5] tool that has been generated from the negotiation metamodel itself.

The GEF [6] and the GMF [7] Eclipse framework has been considered as the primary source for MDA related frameworks for the construction of the Negotiation Model Editor. More detailed information can be found in chapter 6, The ONE Negotiation Metamodel..

The generated code has been further customized in order to fit special requirements which can not be expressed in the meta model, for example creating wizards for adding Variables, Protocol Rules, Criterias, Action Language support or how to access to the repository.

In ONE it was decided to draw two state diagrams, one represents the Owner (left side), the other the Participant (right side). The two State Machines communicate via Messages that are sent by Actions in one side and that are captured by Events on the other side.

In order to describe precise behaviors, like Actions, preconditions and in general behavioral logic, it is necessary to support an action language, this means to augment the ONE DSL with an imperative language. Since scripts are to be executed at run time by the Negotiation Engine and since the entire project is based on Java, it was decided to adopt a subset of the Java Language [8] (JSR-274, implemented by Bean Shell [9]) in order to avoid further transformations in the Engine at runtime.

The editor is a rich client application: the models can be stored locally in the file system or in the Distributed Knowledge Base (DKB) as an XMI [10] encoded file. In line with the MDA approach, these models are not specific to an execution engine, they are a Platform Independent Models (PIMs [11]), there is no assumption about the Execution Engine, Jboss, Ilog or others, not even it assumes if the participants of the

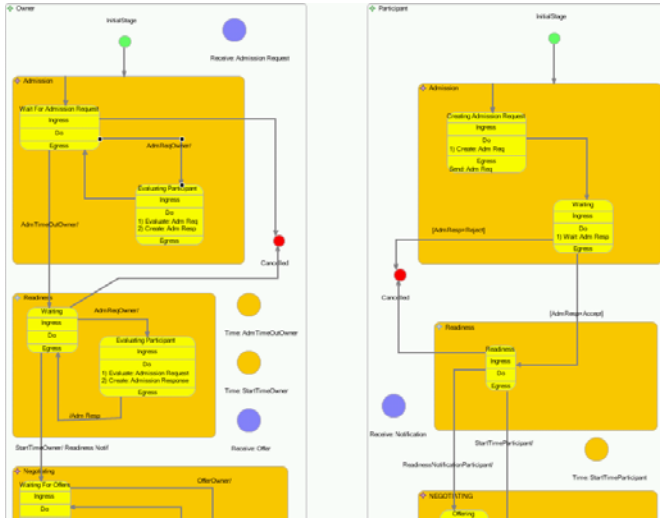


Fig. 2. Negotiation Editor, two state machines, fragment

negotiations are humans using Web based interface or software components using Web Services.

There is no use of an ontology, the projects decided to adopt folksonomy [12] to tag models and the information model.

5 The Runtime Environment

The Runtime is the Web based environment where negotiations' instances are created from models and executed; in addition this environment provides a set of side functionalities[13] like searching, inspecting for negotiations, registration, login and the management of a personal workspace where to keep track of past negotiations, called the My One [14]. The connection point between the Factory and the Runtime is the Distributed Knowledge Base (Distributed Knowledge Base) that receives the models created in the first environment and they are then retrieved, configured and executed from the second.

6 The ONE Negotiation Metamodel

The ONE DSL was created in order to allow users to design any type of negotiation process as simply as possible, while still being precise enough to allow the process to be machine executable: the language has been inspired by FIPA protocol [15] and by the OMG Negotiation Facility [16]. The metamodel uses essentially a subset of UML [17], as no existing UML profile met our requirements.

The metamodel has two sections; the *Information model* which describes *what* is being negotiated, and the *Negotiation Protocol*, which describes *how* an agreement is

negotiated. The Information Model needs to be very flexible to support almost any business domain, so it has a simple structure with attributes and issues that can be negotiated (changed by negotiators). The ONE DSL does not contain any standard agreement or contract types; it is up to users to define these kind of models.

From examining many existing negotiation methods and systems, a number of features common to virtually all negotiation processes were determined. An example of one of these concepts is that to reach an agreement a user must make an offer and that offer must be accepted by another user. These concepts were formalized into the negotiation protocol section of the ONE Negotiation metamodel.

Negotiations in ONE do not require a neutral third party to host or coordinate a negotiation, the parties communicate directly with one another. This brought us to the approach of using independent state diagrams for each user, with the transitions between states being triggered by the sending or receipt of messages, or other stimulus. Each Negotiation State machine have a number of SuperStages that a process would go through on the way to reaching an Agreement. The SuperStages and the basic negotiation protocol were created to help ensure that every negotiation achieved certain legal norms inherent in contract creation [2].

User defined Stages represent specific steps in the negotiation process, such as when a user is creating an offer or other message, waiting to respond to an offer, or waiting after a sealed bid has been made.

As part of the metamodel we defined a number of message types such as: Offers, Agreements, Admission Requests, Information and Notification messages. These Message types have precise meanings but can be extended to support different kind of negotiation models.

In the metamodel there are also Actions that are triggered at certain steps in the negotiation protocol; such as when a user enters or leaves a specific stage. In designing the metamodel a difficult decision to make was the level of granularity for predefined actions. The more functionality that was included in the predefined actions, the less flexible the system becomes. With less functionality the metamodel provides no benefit over existing modelling languages. The decision was to define a few simple common actions, and to define the more complex behavior in action Scripts: small sections of Bean shell code[9], embedded in the models. This approach allows a user to write custom rules like sorting or filtering messages.

6.1 Transformation and Execution

MDA dictates that negotiation models are to be executable. This could have been achieved by developing a state machine engine to step through the ONE stages. However an existing open source state machine engine (JBPM) was found to have all of the core features required for ONE. It was demonstrated that by transforming models from the ONE DSL into JBPM compatible model (JPDL), the development effort could have been greatly reduced.

The process of transforming negotiation models from the ONE metamodel into JPDL was not a straightforward task, since not every feature in ONE metamodel had an equivalent in JPDL. A multiple layer transformation process was developed that supported all the features of the ONE DSL [2]. This process uses a number of ATL rules (Atlas Transformation Language [18]) with XSLT also used for formatting, see Figure 3 below.

The greatest benefit of transforming the models into an existing process model language was that it allowed us to rapidly prototype negotiation processes in a stable state machine environment and use a number of features that would not have been possible to implement with the available resources.

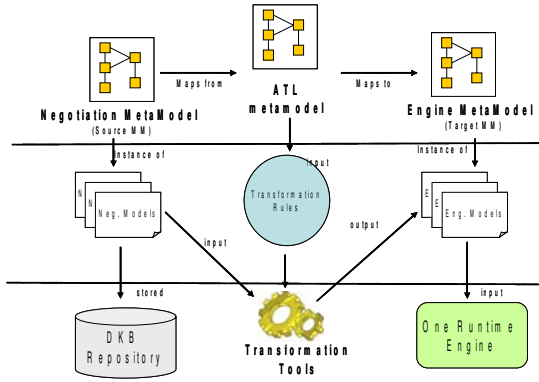


Fig. 3. Models and metamodels

The transformation layer would allow different implementations to use the same negotiation model, with Web Services providing the standard for messaging. As such, it might be possible to replace components of the current ONE platform with new components and remain compatible with the Architecture.

7 The ONE Web Portal and the ONE Node

In the ONE Web Portal participants can create, join and negotiate under the rules specified by the model. There is also a user workspace, where contacts, and past negotiations are stored managed.

There can be several Web Portals in the Internet each with its own different IP address and any of these can be used to log in and to enter the community. The ONE Web Portal is a stateless application server; all the storage of information are managed via the Distributed Knowledge Base only. The One node is the technical infrastructure (see Figure 4), is built on top of the Spring [19] service container. All the functional components are then plugged on top and some of them make also use of the Portal that make use of an Open Source implementation [20] of the Portlet specification [21].

It is assumed that participants have a preference for a trusted node through which they access a ONE node that will be responsible for the traffic and for the activities of their users, as such there is a cross trusted relation that will be properly implemented by the Trust and Reputation component [22]. Users who intend to join a negotiation are to be redirected, together with their trusted/owning node, to the central node where the negotiation will be run. It is important to enforce the concept that a node becomes “central” only for the duration of the negotiation: it is not an fixed property of a node.

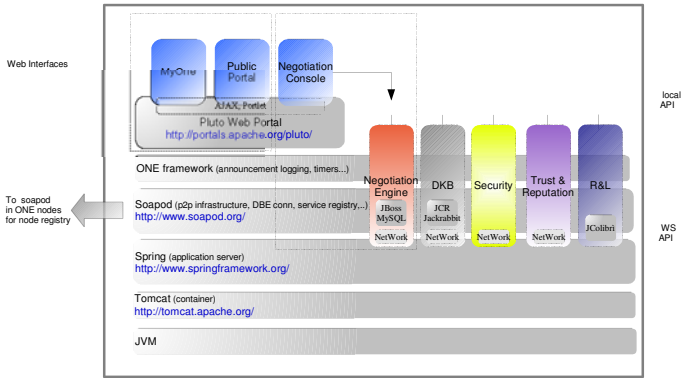


Fig. 4. One node, layers of technologies

Since negotiations are supposed to last for several days or weeks it is hence required to provide a fail-over mechanism: during the negotiation; the Engine will take care of duplicating the negotiation state in a backup node.

8 Recommender and Learner

A node, in addition, could host the connections of many users who could participate to different negotiations and hence the Negotiation Areas do overlap and intersect: a node can support many users in the role of owner and many users in the role of participants in the same or in different nodes (Figure 1 represents an ideal scenario where two separate Arenas do not share any Node).

ONE also provides a recommendation framework and an intelligent learner system that given the negotiation models and the history of past anonymous executions is able to provide feedbacks and infer suggestions to the participants.

The Recommender System [23] that plays a key role in supporting users during the negotiation processes, it recommends how to reply or respond in a negotiation. It is aimed at implementing a learning agent, able to suggest to the user a negotiation strategy. This has been achieved by observing, in the right context, the user behavior and the actions of other partners in past negotiations. The strategy learning algorithms is based on the observation of real interactions that happens in the Execution Environment.

9 Summary and Conclusions

The consortium has successfully adopted the MDA that allowed the solution to functionally scale, to have software components which are not bound to any specific negotiation and are adaptable to custom needs. This provided benefits in the daily organization of the work, easing the interoperability among partners and most of software components that have been generated rather than coded by hand. The project (released of open source [24]) can already be modelled, published and executed.

References

1. Many, Model Driven Architecture
2. Boudjemil, Z., Finnegan, J.: D3.1 – Languages, Models and Agreements (2008)
3. Ferronato Pierfranco, Architecture for Digital Ecosystems: Beyond ServiceOriented Architecture (2007)
4. Many, Digital Business Ecosystem Project (2004)
5. Cavalcance, P., Ferronato, P.: Del 3.3 Set of Software Design Tools (2008)
6. Many, Graphical Editing Framework
7. Many, Eclipse Graphical Modeling Framework
8. Many, Provides classes that are fundamental to the design of the Java programming language
9. Pat Niemeyer, Bean Shell
10. Many, MOF 2.0/XMI Mapping, Version 2.1.1 (2007)
11. Many, OMG MDA Guide (2001)
12. Many, Folksomy
13. Pierfranco Ferronato, D2.1 – Functional Architecture (2008)
14. Stanoevska-Slabeva, K., Faust, R., Hoyer, V.: D1.2 – Software Requirements and Use Cases (2007)
15. Manyù, FIPA Interaction Protocols
16. Many, OMG Negotiation Facility (2002)
17. Many, UML
18. Many, Atlas transformation language
19. Many, The Spring Framework
20. Many, Reference Implementation of the Java Portlet Specification
21. Java Community Process, JSR 168: Portlet Specification
22. Ion, M., Danzi, A.: D5.2 Design of a peertopeer reputation model (2007)
23. Avesani, P., Malossini, A., Penserini, L., Serra, A.: D4.1 Interaction Design (2007)
24. Many, ONE code repository

Analysing Requirements for Virtual Business Alliances – The Case of SPIKE

Christian Broser¹, Christoph Fritsch¹, Oliver Gmelch¹, Günther Pernul¹,
Rolf Schillinger¹, and Sandra Wiesbeck²

¹ Department of Information Systems, University of Regensburg, Germany

² IT-Inkubator Ostbayern GmbH, Germany

{christian.broser, christoph.fritsch, oliver.gmelch,
guenther.pernul, rolf.schillinger}@wiwi.uni-regensburg.de,
sandra.wiesbeck@it-speicher.de

Abstract. More and more companies are realising that business is best carried out in project-based virtual alliances and are intensively networking and collaborating with partner companies. This requires convergent, adaptive, and interoperable IT environments ready to support flexible, loosely coupled networked enterprises, anywhere, anytime - regardless of heterogeneous systems and applications in use. In this paper, a methodology for gathering and analysing user requirements is applied which is specifically tailored for incorporating multi-party views such as necessary for the development of collaboration platforms to support virtual alliances and to prove its significance in a real-world case study. The work presented is performed within the SPIKE project, focusing on the technical, methodological and organisational support of short-term and project-based business alliances between organisations of all types and sizes.

Keywords: Collaboration, networked enterprise, virtual alliance, collecting and engineering requirements.

1 Introduction

Today's organisations face big challenges such as globalisation, shorter innovation cycles, increased competition, and the negative effects of the current financial crises. To continue being successful on the market and to achieve sustainable competitiveness, organisations have to adapt their ways of doing business. Moving from entrenched thinking and practice to flexible and agile business networks in terms of virtual partnerships across organisation's borders with anybody, anytime and anywhere represents an extremely promising way [1]. This is also diagnosed by recent surveys done for example by AT&T [2] or the Gartner group [3]. Both predict a significant increase of virtual business alliances among partner organisations in the near future. In most economies, small- and medium-sized enterprises (SMEs) represent the biggest fraction of enterprises. For instance, in the European Union, SMEs comprise 99% of all companies [4]. Virtual partnerships enable them to open up new business opportunities, to stay competitive and thereby to better manage existing and upcoming challenges. Frequently, SMEs are specialised and extremely innovative in particular areas but are restricted to

offer their services to local customers only. Acting in virtual partnerships – for example, with larger internationally-acting enterprises – may enable them to provide their competencies on a larger global scale and thereby to enter new markets.

Successful virtual collaborations depend on different factors such as skilled employees, integrated processes and extensive use of information and communication technology (ICT) [5]. Although ICT has been identified as an essential aspect for virtual partnerships by several authors (for example [6] or [7]), currently available technologies still lack powerful integration of heterogeneous environments across the partner sites, support of interoperable business process management, the potential to execute inter- and intra-domain workflows, the security and trust of the collaboration platform, and a proper user- and access rights management across domains. To overcome the situation and to provide an integrated solution for collaboration support are hence the goals of the SPIKE project.

The main goal of this paper is to apply a methodology for gathering and analysing user requirements which is specifically tailored for incorporating multi-party views such as necessary for virtual alliances and to prove its significance in a real-world case study. The remainder of the paper is structured as follows: In section 2 we present an overview and the vision of our SPIKE project. This is followed by the applied methodology for requirements gathering (section 3) and a general discussion of the SPIKE requirements (section 4). Section 5 contains the prioritisation of the requirements according to their importance and significance in the SPIKE project. Finally, section 6 concludes this paper with a project status report and an outlook on future work.

2 The Case of SPIKE

The goal of the SPIKE¹ (Secure Process-oriented Integrative Service Infrastructure for Networked Enterprises) project is to deliver a software platform for the easy and fast setup of short-term (up to about 6 months), project-based and loosely coupled business alliances among partner organisations. The SPIKE vision includes a 3-layered architecture of a business alliance (see fig. 1). On a top level, all the organisational and legal prerequisites, which are necessary before the business alliance may start, are negotiated. The second level is the conceptualisation of the business alliance. Usually, a project-based collaboration starts with a common goal of all alliance partners. In order to reach the goal, a high-level strategy is agreed upon among the alliance partners, which in later steps is further refined into a detailed business process specification. The responsibility for the different parts of the business process is shared among alliance partners and the partner having the best qualification for a certain task may be in charge for its accomplishment. Developing the global strategy and further refining it into a detailed business process specification is the second layer of the SPIKE architecture. Layer 3, the SPIKE service layer, consists of mapping instruments responsible for assigning proper (web) services to business process tasks, particularly retrieving, selecting, orchestrating, managing and executing them in a controlled way. Layers 2 and 3 are supported by the SPIKE technology. In particular they are supported by a (web) service specification and enactment infrastructure, ontologies and semantic services, interfaces to business

¹ www.spike-project.eu

process modelling and a workflow management system, a security sub-system, the identity management infrastructure and a SPIKE portal engine.

The SPIKE solution will encompass a semantically enriched service-oriented infrastructure including a virtual semantic service bus for workflow control and handling and transformation of messages. At the enterprise interface level, a collaborative process portal approach will be delivered, capturing the user's working context and seamlessly transmitting it to other applications and services in the alliance. This will also enable integration of legacy systems via tailored portlets and connectors. Special focus is given to the security and identity management issues involved. The solution will include an easy-to-administer security infrastructure for the business alliance which will consist of basic and advanced security services. Additionally, federated identity management and administration of digital user identities and privileges across organisational boundaries is another major issue. A strong focus on the security and identity management aspects is necessary because of the focus on "project-based" collaborations, meaning that alliance partners possibly are competing simultaneously in other projects not part of the alliance. Additionally, further improvements of the state-of-the-art in the fields of workflow management in combination with dynamic and semantically enhanced selection of service instances, integrating enterprise portals and cross-organisational communication infrastructures are expected.

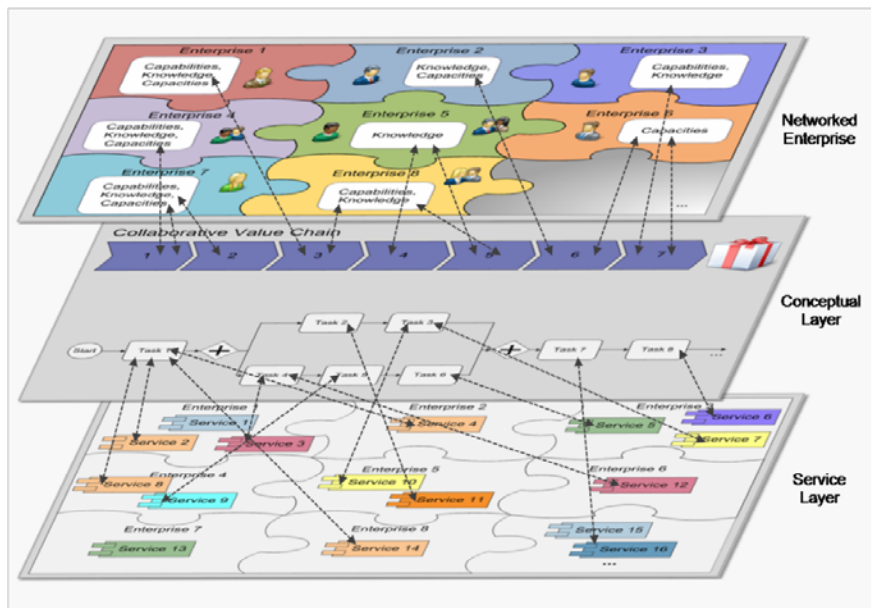


Fig. 1. SPIKE's virtual business alliance

3 Requirements Analysis

It is generally understood that collecting and analysing the requirements is critical to the success of any system. For SPIKE, the general methodology developed by

Maguire and Bevan [8] was applied for the requirements analysis. Based on their findings, the general process for the requirements analysis is performed in four steps which are supported by different analysing techniques accordingly:

1. Information gathering: Background information about stakeholders and processes is gathered. It may be supported by stakeholder analysis, secondary market research, task analysis, or field studies.
2. Identification of user needs: After stakeholder identification, an initial set of user requirements can be identified. User surveys, interviews, scenarios, and use cases as well as workshops are means to collect these basic user needs.
3. Envisioning and evaluation: On the basis of the initial user needs, a prototype (software- or paper-based) is developed. This helps in validating and refining the requirements. It may be further supported by brainstorming, card sorting, and affinity diagramming.
4. Requirements specification: Requirements are specified in compliance to some standardised way. For this, a clear statement of design goals, a prioritisation of the requirements, and measureable benchmarks against which the design can be tested should be documented. Task/function mapping, requirements categorisation, and prioritisation are techniques supporting the requirements specification.

3.1 Strategy and Methods to Identify SPIKE Requirements

For SPIKE, the user requirements analysis was conducted in close cooperation with potential future users and by means of the strategy illustrated in Fig. 2, which is described in more detail below.

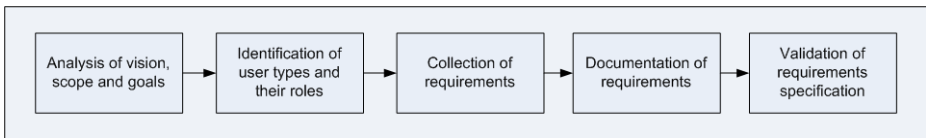


Fig. 2. Strategy for the user requirements analysis

1. *Vision, scope and goals.* A clear understanding of the context in which the SPIKE system will be operated, its purpose and goals need to be determined and formulated. At the end of this phase, all stakeholders must have the same picture of what the project is about. For SPIKE, the result is that the SPIKE system will serve as collaboration platform having its main scope on (1) outsourcing parts of the value chain to cooperation partners and similarly embedding fragments delivered by alliance partners, (2) fast reaction, easy and fast setup of virtual alliances, (3) flexibility, for short-term and project-based virtual alliances mainly, (4) security, as alliance partners may cooperate in some projects but compete in others.

Furthermore, some additional advanced building blocks were defined: *Federated identity management*, with special focus on mapping users and their privileges among different domains; dynamic composition of *workflows across sites* by means of selecting predefined remote services or dynamic service selection during runtime; providing a common user interface, making use of innovative *portal technologies* and overcoming different data and service formats by employing an enterprise *service bus* and *semantic mediation*.

2. *User types and roles.* During an analysis of the potential stakeholders of the SPIKE system, the different user types and their relation to the collaboration platform were identified. The outcome resulted in the following types of stakeholders:

The *technology provider* is responsible for setting up the platform as well as related technologies and ensures the successful design, deployment, operation, maintenance, and decommissioning of SPIKE hardware and software assets. A *portal administrator*'s obligation is to maintain basic data and functions to run the collaboration platform's service portal and to create user accounts for service providers and administer their access rights. The *content manager* is in charge of the functionality of the collaboration platform's portal. For him, close cooperation with alliance partners' business areas is essential in order to transit their requirements to user portlets. The *service provider* offers SPIKE services to other companies. He must provide necessary information such as description or pricing for a specific service via the service catalogue. Furthermore, this stakeholder maintains the legal contracts (access rights for service users of a specific service, duration of the contract) and uses the reporting/auditing functionality of the portal to monitor the use. The *service requestor* makes use of a specific service of a service provider via the collaboration platform. The role of *service executors* can be played by software or human beings. *Service executors* are responsible for the execution of a requested service. The *service locator* is responsible for identifying the respective provider of the requested service.

3. *Collection of requirements.* After all stakeholders and their roles have been identified, the actual gathering of requirements starts. To support this phase in SPIKE, several instruments are applied: A *use case analysis* based on pre-defined application cases turned out to be very effective for gathering requirements. Analysing use cases lead to the description of tasks which have to be carried out by the final system. For the description of the use cases, a template was applied. A *questionnaire*-based survey was conducted by representatives of the stakeholders. Additionally, *interviews*, meaning face-to-face meetings, were conducted with selected participants of the survey in order to clarify inconsistencies as well as to gain additional background information. *Workshops* and *round table* discussions turned out to be very efficient to discuss ideas and exchange opinions between users and developers. Several workshops and round tables with interested companies have taken place to discuss and refine potential requirements.

In addition to the identification of user requirements by the abovementioned methods, further market research was applied to identify user needs and relevant technical literature. Moreover, public information about related projects, such as Crosswork², Ecolead³, Trustcom⁴, and others has been analysed.

4. *Documentation.* After the collection phase, a structured documentation of requirements is essential. In SPIKE, they were assigned to pre-defined types (see section 4). To end up with a standardised documentation, a software requirement specification document, acting as template for the specification has been employed.

5. *Validation.* Validation ensures that the asserted requirements meet the original goals. As a first step, a prioritisation of the identified requirements was performed. Furthermore,

² <http://www.crosswork.info/>

³ <http://ecolead.vtt.fi/>

⁴ <http://www.eu-trustcom.com/>

a check against selected quality criteria such as preciseness, completeness or robustness was conducted and led to the deletion and revision of some requirements.

3.2 Strategy and Method to Analyse SPIKE Requirements

The requirements collected are stated in a requirements document which is subject to further analyses. In SPIKE, requirements validation has been carried out using the requirements analysis strategy illustrated in Fig. 3.

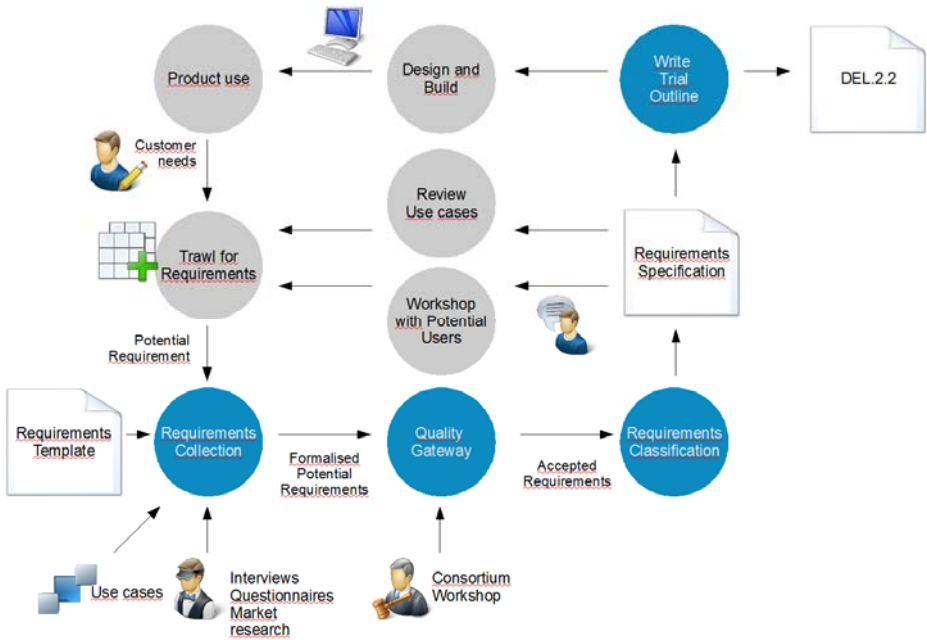


Fig. 3. Requirements Analysis Strategy

The process of analysing user requirements in SPIKE starts with the activity stated in the lower left corner of Fig. 3, passes several phases and finally ends in the requirements specification document (DEL.2.2) shown in the upper right corner. Enabling several iterations to refine and update the achieved results, the requirements analysis is rather a lifecycle than a process. Throughout the lifecycle, use cases and market research are important elements for analysing the requirements. Each single requirement must pass a quality gateway and is reviewed during workshops. Only requirements successfully passing the gateway are still considered and subsequently classified according to predefined requirement types. The analysed requirements serve as initial input for the specification of a SPIKE “functional sketch”, which is passed to the stakeholders for verification.

Relating the SPIKE approach to [8], only minor differences may be determined. We start with an additional phase in order to get a common understanding of the context. Maguire and Bevan’s information gathering and user needs identification phases

are fully included in our strategy and supported by different methods. As the outcome of SPIKE is rather an experimental prototype than a market-ready product, their step “envisioning and evaluation” by an early prototype is of less importance for SPIKE; however, our prototype may serve this purpose well in case it is decided to commercialise the project outcome later. Their final step is also included in our methodology, leading to the final requirements documentation.

4 Categorisation of Requirements for Collaboration Platforms

In order to allow for easier lookup and handling, a categorisation of requirements has been performed along the axes user requirements, usability requirements, and organisational requirements. User requirements merely focus on the functionality aspects of the system; usability requirements include aspects such as level of effectiveness of the system, the efficiency and user satisfaction or the comfort with the system, while organisational requirements target the implication on the individual organisation when introducing a collaboration platform such as SPIKE.

Altogether, almost 100 requirements were collected, whereof sixty were classified as user requirements, thus forming the strongest category. In the following subsections, the categories are discussed on an abstraction level. For a list and a more detailed view, we refer to the SPIKE website⁵.

4.1 User Requirements

This type of requirements resulted from the intended users, including both functional and non-functional requirements. A characterisation of the predominant topics is presented below.

Requirements concerning *workflow management*. Users requested to be able to specify and automate execution of individual activities along a process chain. Consequently, a system supporting inter-domain workflows is expected to consist of different tools, for example a workflow engine, process definition tools, links to applications, workflow enactment services as well as administration and monitoring facilities.

Support for *communication* among the alliance partners. This includes both asynchronous and synchronous means for contacting cooperation partners. For synchronous communication, instant messaging and videoconferencing facilities were mentioned predominantly. Also, team communication facilities were discussed, allowing individual groups to share appointments, invitations to group meetings as well as upcoming tasks and milestones automatically.

Another group of user requirements focuses on *integration aspects*. In this context, the primary concern is the integration of legacy applications within virtual alliances, thus referring to Windows-/X11-based as well as mainframe-based applications. Besides, integration of existing SOA was mentioned as a requirement for collaboration platforms, which can be explained by the huge momentum that was created by SOA in the past [10]. Moreover, integration of ERP software and access management products (e.g. Novell Identity Manager, Firepass, Juniper, Tarantella, etc.) was mentioned as a requirement in order to upload configuration data into partner companies' directories.

⁵ Please find a complete list of requirements at <http://www.spike-project.eu/index.php?id=136>

Another topic addressed is *content management*, again raising the need for team support in terms of joint editing capabilities, enabling users to collaboratively work on content. Content creation has to be kept highly flexible, allowing for usage in different scenarios without any restrictions beforehand. Likewise, semantic capabilities were appreciated, enabling reuse of content generated in different manners.

Security aspects have been identified on different levels: (1) Identification, authentication and authorisation – providing controlled access to the platform and the services offered by third parties. (2) Digital signatures - supporting a fast setup and legal binding of actions in a provable/undeniable fashion. (3) Secure document transfer between individuals within the alliance, supporting confidentiality on different levels (i.e., connection- or data confidentiality). (4) Document encryption – assuring confidentiality and integrity of documents used within SPIKE.

It turned out that users expect to use the collaboration platform from *different devices* (i.e., desktop as well as mobile devices), thus requiring data synchronisation facilities and off-line capabilities in situations where no network connection is available. This also imposes the need for *platform independence* on the client side.

Another important aspect addressed is single-sign-on, inducing the need for *federated identity management* (IdM) of users from partner companies. As collaboration platforms should be open to all types of enterprises, interoperability with existing IdM solutions based on existing standards such as SAML, SPML, and XACML at the individual companies' premises plays a crucial role.

4.2 Usability Requirements

In order to reduce the probability of human errors, an *intuitive user interface* was demanded. At the same time, online help shall be available for key functionality, supporting the user in unclear situations. *Performance* is another issue, meaning that the system should react within predefined response times, taking into account different types of connections to the platform (i.e., ISDN, etc.). Additionally, *fault tolerance* is demanded: Malfunction of a single component may not terminate the whole SPIKE system and should therefore result in a notification of the users only.

4.3 Organisational Requirements

Federated identity management, as already discussed in section 4.1, also requires consideration of the *privacy aspects* of users, allowing individuals within an alliance to control or influence what information related to them may be collected and stored and by whom and to whom this information may be disclosed.

Also, a *structured deployment process*, imposed by the platform administrator, is necessary from an organisational perspective. This implies fallback scenarios during the rollout of new versions of the platform, which is a requirement for high availability operation. This also includes a defined workflow in order to compare changes between individual versions of the platform in case an update of the platform is rolled out. At the same time, the deployment process has to ensure that unintended changes in the runtime environment cannot be performed by unauthorised persons.

Organisational requirements also impose an *alliance-specific administrator* in order to carry out the setup of an alliance. This administrator should not be granted access to the business logic of the alliances (data being worked on in individual workflows, tasks,

external applications). However, administrative personnel from alliance partners may be granted limited administrative privileges.

Time to market was also identified as a key requirement of a collaboration platform. Thus, quick start-up time of an alliance is of utter importance. From an organisational perspective, this means that all legal aspects before the actual start of a business alliance should be able to be carried out rapidly. At the same time, the goal of minimising time to market requires that all potential alliance partners be supported on a technical level, meaning that integration and definition of business processes should be as easy as possible.

5 Requirements Assessment and Prioritisation

The requirements presented above form the basis for the design and further development towards an implementation of SPIKE. Therefore, it is critically important to assess and prioritise them. Our approach to guarantee periodic review of identified requirements has already been laid out in section 3.2. Prioritisation of the requirements depending on their importance is carried out alongside the axes *MUST*, *MEDIUM*, *LOW*, and *FUTURE* carefully as it is generally understood that requirements have a big impact on software functionality, acceptance, and applicability ([11], [12]).

As the requirements originate from several sources and are derived from several independent user groups and usage scenarios, completeness and applicability of the requirements were assessed by the composition of comprehensive use cases, each of them combining several requirements from different sources. By combining independently surveyed requirements from different sources, we were able to detect gaps in the list of requirements and to discover and add missing ones. During this process, varying significances of requirements even within single clusters were observed.

The SPIKE project will put much emphasis on the area of security, both among alliance participants as well as of the platform itself. The start-up time for virtual alliances was also identified as a requirement in the *MUST* priority, along with federated identity management. Also, the topic of workflow-enactment plays a crucial role for SPIKE, as well as all integration aspects.

At the other end of the scale, the issues of deployment as well as the initial configuration process of a collaboration platform or issues of system operation are currently not regarded as a major concern and are therefore treated with low ("*FUTURE*") priority. This is probably because SPIKE is targeting at the realisation of an operable prototype and not a commercial solution.

6 Conclusion and Future Work

In this paper, a methodology for gathering and analysing user requirements is applied which is specifically suited for incorporating multi-party views such as needed for the development of collaboration platforms and to prove its significance in the real-world case study SPIKE. It was argued that it is critically important to elicit requirements from the beginning on and to constantly improve and manage them during the software development process. Based on the requirements found, the SPIKE system architecture, technical building blocks, and components have already been specified and software development is about to start.

The SPIKE platform is going to be developed in an iterative process with several development cycles, each of them leading to a pilot application, giving SPIKE's stakeholders the opportunity to redefine their needs prior to further development. Problems arising during the evaluation of each trial phase are expected to be addressed in the development cycle subsequent to the trial phase. Finally, note that the results presented in this paper are neither conclusive nor exhaustive. They rather mark intermediate results to be subsequently implemented and verified to take another step towards a comprehensive and competitive business alliance support.

Acknowledgement

The research leading to these results is receiving funding from the European Community's Seventh Framework Programme under grant agreement no. 217098. The content of this publication is the sole responsibility of the authors and in no way represents the view of the European Commission or its services.

References

1. Van Heck, E., Vervest, P.: Smart Business Networks – How the Network Wins. *Communications of the ACM* 50(6), 29–37 (2007)
2. AT&T Knowledge Ventures: Collaboration across borders, http://www.corp.att.com/emea/docs/s5_collaboration_eng.pdf
3. Gartner Says Worldwide Web Conference and Team Collaboration Software Markets Will Reach \$2.8 Billion in 2010, <https://www.gartner.com/it/page.jsp?id=507717>
4. Thompson, K.: *The Networked Enterprise. Competing for the Future Through Virtual Enterprise Networks*. Meghan Kiffer, Tampa (2008)
5. Alt, R., Smits, M.: Networkability of organisations and business networks. In: *Proc. of the Fifteenth European Conference on Information Systems (ECIS)*, St. Gallen, Switzerland, pp. 119–130 (2007)
6. Kasper-Fuehrer, E., Ashkanasy, N.: The interorganisational virtual organisation: Defining a Weberian ideal. *International Studies of Management & Organisation* 33(4), 34–64 (2003)
7. Abuelmaatti, A., Rezgui, Y.: Virtual team working – Current issues and directions for the future. In: *Proc. of IFIP Pervasive Collaborative Networks*, vol. 283, pp. 351–360 (2008)
8. Maguire, M., Bevan, N.: User requirements analysis. A review of supporting methods. In: *Proc. of IFIP 17th World Computer Congress*, vol. 226, pp. 133–148. Kluwer Academic Publishers, Dordrecht (2002)
9. ISO 9241: Ergonomics requirements for office work with visual display terminals (VDTs), International Standards Organisation
10. Pettey, C.: Gartner Says SOA Will Be Used in More Than 50 Percent of New Mission-Critical Operational Applications and Business Processes Designed in 2007 (2007), <http://www.gartner.com/it/page.jsp?id=503864>
11. Prause, C.R., Scholten, M., Zimmermann, A., Reiners, R., Eisenhauer, M.: Managing the Iterative Requirements Process in a Multi-National Project using an Issue Tracker. In: *Proc. 2008 IEEE Int. Conference on Global Software Engineering*, pp. 151–159 (2008)
12. Davis, G.B.: Strategies for Information Requirements Determination. *IBM Systems Journal* 21(1), 3–30 (1982)

Business Impact of Online Institutional Recommendation

Mihaela Ion and Regis Saint-Paul

CREATE-NET International Research Center
Via alla Cascata 56D, 38123 Trento, Italy
{Mihaela.Ion, Regis.Saint-Paul}@CREATE-NET.org

Abstract. Institutional recommendation is an important factor for building trust in business or social interactions. Institutional ratings are issued by trusted authorities such as public bodies or reference publications. They assess people's expertise, the quality of products or services or the economic health of companies among other. Institutional ratings are characterized by objective and repeatable evaluations, based on well defined criteria and issued by institutions that build their own reputation on the quality of their assessment. However, on the Internet today, they lack adequate support that would encourage their usage and allow them to play a role as significant as they already have in off-line relationships. This article presents the potential of transposing existing institutional recommendation mechanisms on the Internet and describes new business models and scenarios that could be enabled.

Keywords: Institutional trust, rating agency, rating, business models.

1 Introduction

Any business transaction requires some level of trust between its participants. For instance, a buyer needs to trust a seller that the goods on sell have the advertised quality and that they will be delivered timely. Such a trust relationship might be established by a variety of means including past experience or exchange of documents. The final decision of whether to trust is also influenced by the specific risks involved in the transaction and the participant's own aversion to risk.

On the Internet, most of the times, the parties involved in a transaction have no or little information about each other. Hence, trust building mechanisms became a necessity for the success of online transactions and various mechanisms have been studied and proposed [1]. Usually, trust between participants that do not know each other directly is built through recommendations or ratings from third parties.

Two main approaches can be distinguished here, depending on the source and nature of the recommendation. An approach called "social recommendation", epitome of the Web 2.0 area, consists of relying on fellow customers (peers) to make informed decision about a product (e.g., Amazon) or about a seller (e.g., e-Bay). Another approach, which is the focus of this paper, is to refer to specialized institutions for assessing products, services or their brokers.

For example, investors can refer to financial rating agencies to estimate the risks and future yields of an investment; photography enthusiasts may rely on extensive

camera reviews in photography magazines to make informed buying decisions; and consumers of “bio” products implicitly delegate to specialized agencies the task of verifying that food producers abide to the requirement of “bio” label. Degrees, certificates and awards are yet other forms of evidences that can help establish a trust relationship between parties.

In the remainder of this paper, we will collectively refer to these forms of rating, labeling or certification as “Institutional Rating” or “Institutional Recommendation” (IR for short) and institutions issuing such recommendations will be referred to as “Rating Agencies” (RA for short).

By contrast with social recommendation, IR emphasizes *objectivity* and *repeatability* of the recommendation [2]. A set of well defined criteria, sometimes measurable, are used to sustain the claim of objectivity of the rating. An aggregated recommendation (e.g., a number of “stars”) may also be provided to help digest all these criteria into a global rating intuitively intelligible to the user. For instance, hotels have to provide a minimal quality of service over a range of criteria to obtain a “five stars” label. RAs will therefore issue ratings for each of the reviewed entities. The issued rating is *unique*, meaning that for a given set of criteria and a given entity, objectivity would lead to a same rating (e.g., a hotel cannot be at the same time 3 and 4 star). However, a same RA may provide multiple ratings to a same entity as long as their criteria and semantic are different (e.g., an RA may rate the hotel restaurant in a way and the hotel rooms in another). Moreover, ratings often have limited validity, reflecting the expected variability in time of the corresponding market.

IR is complementary to other forms of recommendation and, in some cases, has no equivalent. For example, someone’s legitimacy to practice medicine cannot (and probably should not) be assessed by ways of social recommendation. Assessing certain criteria may also require access to information not available to the general public (e.g., access to production facility for “bio” labeling).

The above examples show the key role plaid by IR in traditional business. However, on the Internet today, social recommendation is largely predominating and IR is still carried out using “off line” mechanisms (e.g., paper documents).

In this article, we examine social recommendation along several dimensions to highlight its shortcomings (Section 2). We then contrast these shortcomings with the characteristics of IR (Section 3). Our objectives are twofold. First, we wish to motivate IR as a most needed and complementary recommendation form to conduct online business. Second, we intend to characterize thereby a list of desired feature that online IR should present. On this basis, we summarize a possible model of online IR (Section 4) and outline the new business models and activities that could grow on top of such framework.

2 Limitations of Social Recommendation

Is a high definition camcorder better than a low definition one? Arguably, high definition is attractive for its enhanced image quality, in line with the latest television standard and therefore is better. On the other hand, low definition is attractive for it is less demanding in terms of storage space and its format is in line with the requirements of online video sharing. The lower price doesn’t hurt either.

Leave the recommendation to online buyers and expect a heated debate on the pluses and minuses of each model. Depending on the website gathering user recommendations, ranking of the goods may also be mixed with the user's overall satisfaction about the transaction (e.g., delay on the delivery, packaging problems due to transport, etc.) and biased appreciation of one's own choice.

This small example illustrates one of the issues faced by product recommendation based on feedback, also termed the "*word of mouth*" approach [3]. In the following, we briefly review some of the critical issues raised by social recommendation. Our aim is not a comprehensive survey of these issues, which form an active research domain, but rather to outline the main motivation for an alternative approach based on institutional recommendation which, at least partially, answers to these issues.

Scope of application: Social recommendation is inherently limited to the qualities publicly accessible to users, either through the usage of a service or good, or through their existing relationship with a company or person. This excludes complex assessment (e.g., Does Company A abide to EU environmental recommendation?) or assessment based on evidences which are not part of the relation (e.g., Does company A manufacturing process comply with personal ethic?). While such questions are relevant to the user's trust in company A, the user will not be able to provide feedback about them.

Method and quality of measure: The collection of social recommendation may be implicit (e.g., google through site cross-linking) or explicit (e.g., e-Bay). Both approaches have drawbacks: in the implicit approach, feedback gathering can be distributed (any user building a website is at the same time providing a recommendation) however, the feedback is then very poorly qualified (the linked site might be so equally for critic or appraisal). Therefore, implicit gathering mainly measures popularity rather than intrinsic quality. On the other hand, web sites such as eBay ask users to provide feedback by choosing between negative or positive recommendation (or any intermediate appraisal level) and even publish guidelines on how to evaluate. However, the feedback provided is typically subjective and can lead to unfair rating due to a mixing of the criteria (see our introductory example) or malevolent behavior (e.g., discriminatory ranking) [4]. Genuine users might also provide improper rating due to the lack of perspective on the topic (they might know just a single product, the purchased one, and are not able to perform comparisons). Moreover, social recommendation is sensitive to a number of threats such as collusion attacks [5] in which a coalition of actors agree on fake recommendation to artificially raise their ranking or Sybil attacks [6] in which one entity creates multiple identities to increase the reputation of one identity.

Dynamics of the recommendation: Effective recommendation based on feedback necessitates a long term relationship between the market actors for allowing the time to collect feedback. This raises the question of how new entrant can gain initial reputation (cold start problem). Feedback-based recommendation might also be a measure of popularity rather than intrinsic quality (e.g., google pagerank). As a popularity metric, it behave in a dynamic way similar to fashion or vogues. Feedback-based evaluation operates continuously, either aggregating the feedback over some time-window or considering the entire feedback history. A difficulty in this case is to

match the time window considered and the volume of feedback received with the dynamics of the quality of the service or institution being ranked. For instance, if an ISP quality of service is good at some point and leads to positive feedback, it will take a long time to update the reputation of the provider to reflect a change in the quality of service in case of a sudden drop.

Provisioning to third party: Social recommendation is the by-product of the participation of users into a virtual community. Feedback is perceived as a way to serve the community and is typically provided only on the web site that supports the community. For example, e-Bay feedback about a seller or Amazon reviews can only be seen by visiting e-Bay or Amazon respectively. This limits the effectiveness of the recommendation to serve in trust establishment between parties that enter in relationships outside of this recommendation domain.

3 Institutional Recommendations

In this section, we present how traditional mechanisms of institutional recommendation address the issues listed above regarding social recommendation. We note that a number of the benefits of IR are linked to their essentially “off-line” nature. Therefore, we present in Section 4 a generic framework that would allow bringing these beneficial off-line features to the online world.

3.1 Existing Forms of Institutional Recommendation

IR covers a wide range of business and legal activities. It can be defined as “a professional assessment carried out by a clearly identified institution (public or private) or its representatives as the result of an examination process”. The examination process implies that the assessment is not generated in an ad-hoc fashion but relies on some (semi-)formal methodology. Here are few examples of IR: Product certification (e.g., aircraft flightworthiness, MS Windows compatibility), Professional or academic certification (e.g., diploma), Homologation, Professional review (e.g., camera test by a professional publication, edited restaurant guides).

Sometimes, a clear-cut distinction between institutional and social recommendation can be hard to make. For instance, restaurant reviews and ranking can be carried out by paid professional journalists who visit the restaurants. However, any such publication also receives feedback from its readership that might be incorporated, updating the review. We would still qualify a published restaurant guide as an institutional recommendation because the guide puts its credibility at stake when publishing a ranking. Therefore, it is in the guide interest to carefully edit consumer’s feedback prior to their inclusion so as not to undermine the perceived authority of the publication on the matter. From this example, we can see that institutional recommendation can also be characterized by contrast with social recommendation:

Scope of application: Institutional recommendation is carried out as a distinct activity from the product or service consumption. Therefore, it is not bound in its scope to assessing only “visible” qualities generally accessible. When the evaluation is carried as part of a legal process (e.g., homologation, certification), there is virtually no limits

to the range of data accessible to evaluators. When it is carried out by recognized institutions as part of a publication effort, the marketing impact of a good evaluation is often a strong motivation for companies to open their process to journalist's investigation.

Methods and quality of measures: The criteria used in institutional recommendation are at least semi-formalized through a series of guidelines, requirements or protocols. The same criteria are used to assess a large number of entities. The process may range from strictly formal (aerospace certification) to more or less formal (diploma) down to poorly formal (restaurant ranking). However, even in this later case, some measure of repeatability can be achieved by relying on the experience of professional reviewers and by repeating the evaluation. The institution performing the ranking also places its reputation at stake, which is a strong driver of quality. Note that this does not prevent errors or malevolent behavior, with threats including conflict of interest (e.g., review magazine where the reviewed product manufacturer is also an announcer in the magazine) or sheer incompetence (e.g., financial rating agencies who rated AAA US mortgage derivatives).

Dynamics of the recommendation: Several business models are possible for institutional ranking. Often, the ranked institution or person may pay a fee for obtaining the assessment (e.g., professional certification, product homologation, etc.). Most of the models avoid the cold-start issue of recommendation based on feedback. The recommendation life-cycle often presents a clear renewal mechanism, either based on validity limits (e.g., driver license) or on periodic reevaluation (e.g., restaurant guide). The periodic assessment is an integral part of most recommendation institution business models.

Provisioning to third party: The assessment is necessarily issued by an identified institution with authority to this end. This authority might be informal (e.g., built over time, as with restaurant guides) or formal (legal), as is the case of a university issuing diplomas. Once delivered, the assessment can be claimed directly by its recipient (e.g., a restaurant might display the ranking at its door or a student can carry her diploma to an interview). The issuing institution might also allow access to the assessment (e.g., restaurant guide), but this access can be limited to verification purposes (e.g., university).

3.2 Requirements for e-Version of Existing Models

Currently on the Internet, ratings are displayed as logos of the RA on the rated Web site and users can click on the logo to check its validity. When clicking, users are typically redirected to the Web site of the institution that provided the rating or to some report issued by the specialist that evaluated the entity. This approach has several disadvantages:

1. The absence of standardized mechanisms for verifying ratings authenticity. The current approach has many security vulnerabilities related to Web redirection or phishing attacks. Indeed, it is as easier to fake a rating agency's Web site than to fake the rating itself;

2. The absence of a standardized access to rating schema (e.g., which features of the product are evaluated and how) makes difficult understanding some ratings, thereby limiting their usefulness;
3. Ratings cannot automatically be processed by applications. It might be useful to benefit from the ordering they introduce among entities to enhance user experience (e.g., filtering on products that satisfy a given norm);
4. Due to poor interoperability, ratings cannot be easily exchanged between RAs.

These issues highlight to two key requirements that a framework for institutional trust on the Internet would have to fulfill: (1) Propose a standard representation of the documents involved during interaction with institutional rating (that is, the ratings and their metadata); (2) Provide a standard mechanism for the authentication of these documents (including authentication of the RA's legitimacy to issue the certificate, of the entity's legitimacy to claim a rating, and of the criteria that were used for deriving a certain rating).

3.3 A Framework for Online Institutional Recommendation

In the light of above requirements, we have proposed a framework for the representation and exchange of institutional recommendation (and ratings). In this section, we recall its main feature and refer the reader to [7] for further details.

We chose to rely on existing security mechanisms, developed for the purpose of web-site authentication, in order to enable the certification of the documents involved in IR. Besides building on proven technology, an important benefit of such an approach is to speed-up the adoption of the IR framework by lowering the cost of deploying IR specific certification systems. We then introduce a number of documents which are keys to the IR domain:

Identity certificate of the RA: Serves for verifying the identity of the issuing Rating Agency. Several certification standards bind an identity such as a pseudonym, web address or email address to a public key. The correctness of the identity-public key binding is certified by a Certificate Authority trusted by the user.

Attribute certificate of the RA: An attribute certificate binds attribute-value pairs to a public key. Existing standards are flexible enough to support expressing any kind of attribute information. For example, an RA could be issued a certificate by a public body in a country that authorizes it to issue "bio" labels for a particular product in a certain region. Several standards support attribute certification.

Rating schema (or rating metadata): For a rating to be useful to a user, the user needs to understand what the rating means. It is the responsibility of RA to make available all the information needed to interpret its ratings: the type of entity to which the rating applies, the criteria used, and the rating process. The rating schema is meant to capture these information and this document is referenced by a rating to allow its interpretation. The rating schema might be defined by the RA itself or by a public body at a national level (e.g., bio labeling criteria).

Rating certificate: Based on the rating schema, an RA rates an entity and issues a document stating the rating achieved. In order to check the authenticity of the rating, the RA needs to cryptographically sign it with its private key. Because this document carries a signature, we call it a rating certificate. The attribute certificate standards described above are suitable for expressing the rating certificate. In particular, SAML is the most flexible and appropriate standard. SAML allows expressing any pairs of criterion-value because was designed to allow different industries to define attribute names and type of values in specific namespaces. Moreover, being XML based, ratings are easy to process, request and exchange by applications.

Identity certificate of the entity: It is important to correctly link an entity with its rating. This can be achieved through a proper identification of the entity and a linking of the entity's identity with its rating. These are two challenges that need to be addressed in order to transpose IR on the Internet. Identifying any kind of entities like real world objects is an open research problem. Moreover, currently there is no binding of attribute and identity certificates. Public key infrastructure (PKI) standards recommend using a reference in the attribute certificate like the serial number of the identity certificate to which it refers.

4 Impact on Business Models

The availability of an electronic version of traditional IR (as, e.g., outlined in the previous section) is expected to induce modification of existing business models such as B2B and B2C. In particular, electronic and interoperable IR has the potential to create an online marketplace in which:

- Organizations belonging to a domain define their specific rating schema
- RAs rate entities based on well-known and standardized rating schema
- RAs could define schemas tailored for specific domains or categories of users
- Ratings are translated across geographical and organizational boundaries
- Consumers are able to make informed decisions about products and buy those that satisfy their needs better
- Consumers and vendors trust the market and the number of transactions increases

We will illustrate the above points through the example of foreign diploma recognition. Each school evaluates students and issues diplomas based on well-defined criteria, usually established at national level. Inside the same country, an employer can assess a job candidate fitness based on diploma owned because: (i) the employer trusts the school or the national body that accredited the school, and (ii) she is familiar with the rating criteria (number of years or national ranking of school). However, when presented a diploma issued in another country, possibly in a different language, and using different rating schema, the same employer will not be able to rely on the diploma for assessing the postulant. To cope with such situations, the employer might rely on a local body to verify the diploma and issue an equivalent one in the national system. Usually this consists of language translation, rating scale translation (e.g., 1-10 grades, 6-1 grades or A, B, C, D, F) and so on. To facilitate diploma recognition, countries establish agreements and each define a set of rules or mappings from foreign diplomas to their local ones.

With electronic IR, this tedious task can be sped up or avoided altogether. Instead of operating in disconnected fashion, the electronic representation and exchange of IR allows issuing agency to establish direct relationship between each other and to automate the process of translating RA with the following advantages:

- To allow users to consume and interpret ratings issued by unknown RAs.
- To enable trust in unknown RAs since the trustworthiness of the unknown RA is guaranteed by the local trusted RA. For example, a national agency providing labels for Bio product would be able to verify the criteria used by the national agency of another country. If these criteria are largely identical, each national agency can grant its peer a recognition allowing consumer of a country to confidently consume bio products certified in another country.
- To allows an RA to rate entities that are not in its geographical area (through the translation agreements mentioned above).
- To enables business relations between entities in different domains (e.g., for market makers operating on multiple markets).

5 Conclusions

Institutional ratings play an important role in off-line relations. They assess people's expertise, the quality of products or services or the economic health of companies among other. On the Internet, they are not supported properly and we see a need to transpose them to the online world. This paper provided a comparison of social and institutional ratings and showed where IR play an important role and cannot be substituted. We then proposed a framework for IR based on current certification technologies and discussed the new business models enabled by an online representation of IR documents.

References

1. Josang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. *Decision Support Systems* 43(2), 618–644 (2007)
2. Sint, P.P., Sciences, A.A.O.: Institutional rating in everyday life. In: *Proceedings of the Delos Workshop on Collaborative Filtering* (1997)
3. Dellarocas, C.: The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* 49(10) (2003)
4. Dellarocas, C.: Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems. In: *Proceedings of the twenty first international conference on Information systems*, pp. 520–525. Association for Information Systems, Brisbane (2000)
5. Lian, Q., Zhang, Z., Yang, M., Zhao, B., Dai, Y., Li, X.: An empirical study of collusion behavior in the Maze P2P file-sharing system. In: *Proceeding of the 27th International Conference on Distributed Computing Systems (ICDCS 2007)*, Washington, DC, USA. IEEE Computer Society, Los Alamitos (2007)
6. Douceur, J.R.: The Sybil attack. In: Druschel, P., Kaashoek, M.F., Rowstron, A. (eds.) *IPTPS 2002. LNCS*, vol. 2429, pp. 251–260. Springer, Heidelberg (2002)
7. Ion, M., Koshutanski, H., Hoyer, V., Telesca, L.: Rating Agencies Interoperation for Peer-to-Peer Online Transactions. In: *Proceedings of the Second International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2008* (2008)

uCard: A Novel Privacy Friendly Radio Frequency Tag Architecture for Universal Loyalty Card

Krishan Sabaragamu Koralalage and Noriaki Yoshiura

Department of Information and Computer Sciences,
Saitama University, Saitama, 338-8570, Japan
krishjhp@gmail.com, yoshiura@fmx.ics.saitama-u.ac.jp

Abstract. According to some authoritative marketing analysis, a large percentage of daily sales are increased by its frequent customers. Due to the above fact almost all the stores started to implement point card or loyalty card schemes to achieve the eighty to twenty situation for both the customers and the retailers. Currently average customer has got at least five to six loyalty cards on different stores. Some venders have implemented such card system into the mobile phones. In addition to those cards each person at least has to have two or more hospital or access cards too. However subscribing several cards to achieve economical benefits have become a burden due to the difficulty arise when managing them. Therefore in this paper we proposed a novel privacy friendly loyalty card architecture which allows easy management of cards by using all-in-one card called uCard (Universal Card) using Radio Frequency Identification technology. Our universal card architecture will provide an ideal strategy for stores to build new relationships and to harvest the reward of customer loyalty by enhancing the customer experience while eliminating the hassle of managing several cards and decreasing the fear on personal privacy.

Keywords: Loyalty Card, Point Card, RFID.

1 Introduction

A loyalty card program is an incentive plan that allows any business to gather data about its customers. Customers are offered products or service discounts, coupons, points toward merchandise or some other reward in exchange for their voluntary participation in the program. A secondary goal of a loyalty card program is to build repeat business by offering participating customers something that isn't available to non-participating customers [2, 3, 4, 5].

Loyalty cards can be of several types: paper based cards, magnetic cards, smart cards, or software applet card in mobile phones. Typically a loyalty card has a barcode or magnetic stripe that's scanned at the point of sale. The card identifies the customer and sends information about what the customer bought to a database. The information in the database is used to help the retailer understand and influence his customers' buying habits [2, 3, 4, 5].

Nowadays, average customer has to have number of loyalty cards in their possession to use discounts on services or products. In addition to that most of them have to

have several other cards such as hospital cards, insurance cards, credit cards, bus cards, train cards, electronic cash cards, etc too. Though holding several cards on various services or products help customers to receive efficient and discounted services or products, carrying bulky wallets with number of cards and picking the relevant card out of them has become a big burden.

Similarly some customers may have paper based cards, plastic cards or else card software embedded into a mobile phone. However, several forms of loyalty cards increase the complexity and also confuse the customers' day-to-day life. For instance imagine a situation where one customer has to manage four different personal identification numbers to manage four different loyalty cards in addition to bank cards and credit cards. It seems that the customers are already fed up with bunch of cards they have to keep in their wallets and remembering passwords and personal identification numbers as secret.

When investigated, we could find that it is always a unique number do these miracles. All most all the cards are identified uniquely by using a number or unique code. Only when the highly secured transaction is to be carried out, other properties are checked for verification. Otherwise all the data necessary to provide service or product are called from the proprietary database. Therefore the cards are used to locate the relevant record in service provider's database to provide better service.

One best option to solve the above mentioned problem is to make all-in-one card to use on any services or products. However, since the services and products wanted by one customer differ from others and the methods of such offering differ from provider to provider; there is no possibility to make one card with one unique identification number for all the services to use. Additionally some needs to be highly secured and some other needs no security at all [7].

Depending on the above requirements we came up with novel solution called uCard (universal card) to solve above problems. It is all-in-one card option but with secure role based accessing methods. It also enables the service provider to model their own number in uCard to compatible with their system. uCard has got secure communicational protocols to enable communications with any party securely [1, 6].

Rest of the sections in this paper is organized as follows. Section 2 describes the literature review. Section 3 describes the proposed solution uCard including its architecture and functionality. Communicational protocols are described in section 4 while the possibilities of uCard are discussed in Section 5. Finally section 5 concludes the paper with remarks and future works.

2 Related Works

There are many loyalty card schemes implemented in United States of America, Canada, Australia, Europe, Ireland, Germany, Switzerland, Hungary, Spain, Japan, Malaysia, Singapore, Indonesia, India, China, Taiwan and United Kingdom. Out of all the countries to the best of our knowledge there is no solution to relate with uCard other than Nectar in United Kingdom [2, 3]. The loyalty card market in the UK is one of the most significant in the world, with most major chains operating some form of reward system. Nectar loyalty card is one of the famous card schemes in U.K. and it was started by a partnership of suppliers including the supermarket chain Sainsbury's,

the credit card American Express and the petrol distributors BP and initially launched in the autumn of 2002[2]. Sainsbury's, BP, Ford, EDF Energy, Expedia, Hertz, Brewers Fayre, Beefeater, American Express, TalkTalk, Dollond & Aitchison, The AA, Thomson Directories, Gala Bingo, Table Table, TNS mysurvey.com, Homebase are the members currently participating in Nectar loyalty card schemes[3].

Unfortunately, due to the security we are unable to know the architecture of the Nectar cards to compare with our proposal. However we believe that our solution is not similar to the Nectar in all the aspects other than the goal of providing all-in-one card to allow easy management for customers.

3 Proposed Solution: uCard

uCard is a universal card which can be used as a point card, loyalty card, club card, etc. Information stored in uCard is divided into two main parts: public information and private. Public information can be the information that needs no security or have less threat to personal privacy [7]. Private information need high security and can be used to store bank card, cash card, etc. uCard is a passive Radio Frequency (RF) IC chip which is designed according to the special requirements.

uCard can be of two forms. It can be used either as a plastic card or as a mobile phone by embedding a passive RF chip in it. If the same card system is to be used in computer communications, there must be a reader to take input from a plastic uCard or mobile phone uCard. Then uCard can be used as a web-card which will act like a loyalty card for the internet shopping or transactions. For simplicity in this paper we will describe the chip embedded plastic uCard. To be economical and secure, uCard should be made of UHF passive RF chip and restrict only a few millimeters of contactless reading distance. Depending on the demography and the country, expansion of the number of attributes and its division to public or private should be arranged. Normally forty to sixty divisions of public to private may be desired with possibility of expansion.

3.1 Conceptual Design of uCard

Using the object oriented concepts; uCard was designed to be encapsulated and allow access to its data only via its methods. Therefore uCard can be considered as an instance of an object class called Card. It has got public attributes, private attributes and methods to access those data. As shown in figure 1, uCard has got two methods: get method and set method. Figure 1 also illustrates the way uCard instance is generated.

Each uCard should consist of "oName=uCard" and "anonymousID=<random unique number>" at the time of fresh card generation. No private information can be accessed without approval of the card owner. Similarly no information including public attributes can be set without such approval. Following paragraphs describes the conceptual design and logical structure of the uCard. No personal information which can identify one individual person uniquely is included inside the chip. Instead the name and the expiration date may be printed on the card.

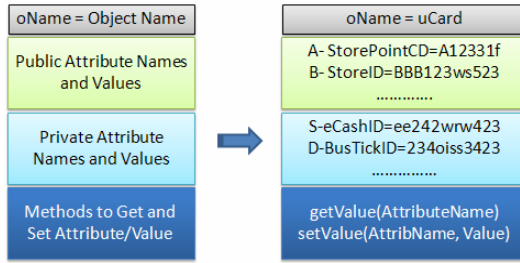


Fig. 1. Illustrate the instance definition of uCard

In uCard, role base accessing methods are implemented using access modifiers, memory types and encryption algorithms. Memory types used here are ROM, EPROM, and EEPROM. uCard controls writing permissions according to the access privileges. ROM is used for read only data and they are created by time of fresh tag manufacturing. EPROM is used for one time writing data and they are all about uCard information. EEPROM is used for all the rewritable data and those data can be changed any time on owner’s request.

3.2 Logical Structure of uCard

Figure 2 illustrates the sequence number, service providers’ code, attribute name, attribute value, access modifiers, occupancy status, user name, PIN and anonymous ID. Data belongs to a single row can be called an entity. Each entity includes sequence number, service providers’ code, attribute name and attribute value. An entity belongs to a one service provider. If someone needs to change an entity of retail shop P to retail shop Q, data belongs to entity P must be overwritten with Q’s entity after successful owner authentication process.

oName	uCARD	Methods	Initial	Grain1 Encryption and Decryption	Processing Module
Nonce Generator, Encryption and Decryption Module, Processing Module and Hard coded Value "Initial"	Nonce (40bits) N_T		Nonce (40bits) N_R		
User Name, Password(PIN) and AnonymousID (random unique number)	User Name(24bits) 1Digit&10Chars		PIN (16bits) 4Digits	AnonymousID(40bits) 10Digits 1234567890	
Occupancy Status 2Bits U-Used N-Not Used B-Blocked	U	U	U	U	U
Access Privilege 2Bits A-Public B-Private	A	A	A	A	A
Service Providers' Attribute Value (Fragment of Identification Code) 40 Bits	45647897	45456463	38968923	XXXXXXXXXX	45344531
Service Providers' Attribute Name 32Bits	RetailShopA	RetailShopB	RetailShopC	XXXXXX	HospitalA
Service Providers' Code (16 Bits) 2 Digits&4 Characters	12werW	34RFgt	RetA37	XXXX11	54sdDS
Sequence 2Bits	01	02	03	04	05
	06	07	08	09	10
	11	12	13	14	15
	16	17	18	19	20
	21	22	23	24	25
	26				

Fig. 2. Illustrate the logical structure of uCard RFID chip

Encryption algorithm used here is Grain1 stream cipher. Therefore encryption key size is fixed to 80 bits and it will be composed with anonymous ID, user name and PIN. Private information is transmitted after successful authentication with encryption. uCard contains methods, hard-coded data “Initial”, Grain1 algorithm and processing module. uCard generates nonce N_I whereas the interrogator (reader/writer) generates three nonce values N_T , ID_I , and ID_T to carry out proper mutual authentication. Service providers’ code is used to protect public readings. In addition to that, user name and PIN are used to ensure the security of both public and private writings.

Public and Private areas are marked as A and B respectively whereas occupancy status are marked as U, N, and B for used, not used and blocked. If three consecutive attempts are made retrieve same attribute value by sending wrong service provider code to the uCard, occupancy status will be changed to blocked status “B”. After that, no data relevant to blocked card information row can be read until changing the occupancy status to “U”. Similarly if wrong PIN is entered consecutively ten times the card will be blocked and have to bring it to the card management centre for recovery.

uCard is an all-in-one card and it is supposed to be filled with array of service providers’ code, attribute name, and identification number. uCard do not store full identification number inside the tag, instead a fragment of the full identification number. The other fragment will be in the service providers’ database system. Thus anybody without both parts of the identification number cannot try for any meaningful information from uCard. Additionally, it is not easy to gain access to service providers’ database even if someone manage to read the fragment of an identification number stored inside the uCard. Hereafter fragment of identification number stored in uCard is referred as attribute value.

As explained in above paragraphs, information stored inside the card is categorized into public and private. In addition to that uCard has a built in authentication method to control read write permission. This has been achieved using a form of user name and password. Password here is considered as a PIN. For proper authentication it is necessary to pass the correct username and PIN to the uCard through the interrogator. Without proper authentication no writings are allowed to both public and private information. Similarly, the readings of private information need owners’ authentication plus service providers’ authentication whereas public attribute value can be read only with service providers’ authentication. However, no service provider should be able to read the other information or other attribute name value pairs written in uCard even if the service provider is authenticated to read their attribute value. This is controlled by using the service provider code which act as a secret ID and stored inside the uCard. The service providers’ code is not even known to the card owner but the service providers’ database system. To read the each attribute value it is necessary to pass the sequence number, service provider’s code and attribute name to the uCard. Then uCard check the validity of those inputs and send the desired results after validation and verification.

When reading public attribute values, there is no necessity to carry out owner’s authentication process. Instead the attribute value will be issued only when the reader could pass the correct sequence number, service providers’ code and attribute name to the uCard. Unlike public reading, in addition to above process; private reading, private writing and public writing need proper owner authentication before releasing or changing any information inside the card. Public modifier allows fast reading of any attribute value by any interrogator who has a common communication interface and

could pass the correct sequence number, service provider code, and attribute name of attribute value to be read. The service providers' code will be different from subscriber to subscriber and even within the same service provider.

Private role is granted only to the owner of the uCard. He or she has the permission to read, write and overwrite any attribute name or value pair against the selected sequence number of the own uCard. However, card owner is restricted to read service providers' code written in own card. On the other hand, the owner has the right to overwrite any row of attribute value pair including service providers' code, if they are no longer wishes to use the some provider's services and needs to subscribe new service instead of current service. Even in new subscription, access to the new service providers' code is also restricted to the card owner. Sequence number in uCard is a fixed value and it is being built using ROM.

4 uCard Protocols

uCard has four main protocols: non-secure reading, secure reading, and secure writing and updating of user name and PIN. Protocol notations are as follows.

<i>I</i>	- Interrogator (Reader/Writer)
<i>T</i>	- Tag (uCard)
<i>K_{prv}</i>	- Private Key [Anonymous ID, PIN, and User name] (80bits)
<i>PIN</i>	- Personal Identification Number [Password] (48bits)
<i>NI</i>	- Nonce generated by Interrogator (40bits)
<i>NT</i>	- Nonce generated by Tag (40bits)
<i>IDI</i>	-Interrogator generated ID (16bits)
<i>IDT</i>	- Interrogator Generated ID (16bits)
<i>Initial</i>	- Publicly defined initial message (16bits)
<i>R</i>	- Response value – Attribute Value, or Successful/Failed [1/0]
<i>{M}K</i>	- Message “M” encrypted by Key “K” using Grain1 stream cipher algorithm

4.1 Non-secured (Public) Reading Protocol

Non-secure reading protocol is used for public reading. Any reader can query the public attribute values by passing the sequence number, service provider code, and attribute name to the uCard as shown in figure 3. For that it is necessary to identify the uCard instance first. To recognize the uCard instance, query is made to get the oName and anonymousID. If oName attribute value is of “uCard” then the service providers' interrogator asks to search its database to find the relevant sequence number, service provider code, and the attribute name to make the query to retrieve the desired attribute value stored in uCard. Once uCard receive a request from interrogator, received data will be matched against the own data and respond with the attribute value if three of them match with received values. In case if there is no match found inside the card, the block counter will be increased and failure will be informed to the interrogator. Likewise each time the mismatch occurs the block counter will be increased. If a proper match is found before the third consecutive wrong attempts, counter will be reset otherwise the occupancy status will be changed to “B” by blocking the readings of that particular entity until it is reset to “U”. Once an entity is

blocked, no access to the attributes of blocked entity can be made until the occupancy status is set to “U” with owner’s authentication by providing the user name and PIN. In the service providers’ side, once the attribute value is read successfully, interrogator composed the absolute attribute value by adding the service providers’ part into the retrieved attribute value and then locates the customer’s record to provide special services or facilities.

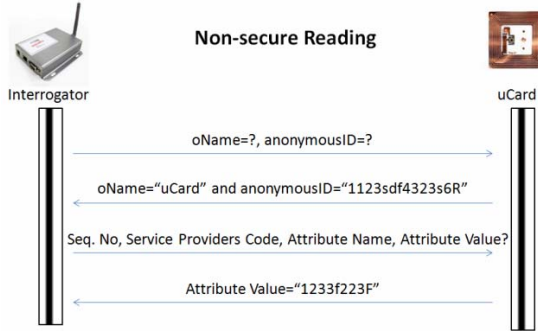


Fig. 3. Illustrate the public reading protocol

4.2 Secured Reading (Private Reading)

Unlike in non-secured reading, each attribute row must be managed securely. Reading is allowed only after successful mutual authentication. Here the KPRV denotes the Encryption key which is a combination of anonymous ID, user name and PIN. Like in non-secured reading, interrogator finds the relevant sequence number, service provider’s code and attribute name before querying the attribute value. Then those three values are passed to uCard, after proper owner authentication. Here the authentication is carried out by taking the user name and PIN as input to the desired reader. Hence only the service provider can read these attributes after receiving owner’s permission. Additionally, to ensure security of each reading messages are encrypted with encryption key and send as shown in Figure 4.

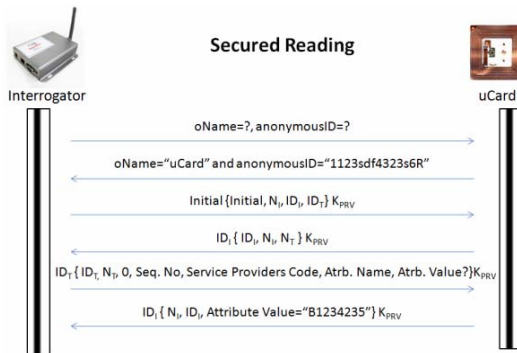


Fig. 4. Illustrate the secure reading protocol with private role key

4.3 Secure Writing

Same as in the secured reading, the secured writing needs successful mutual authentication. Attribute name and attribute value pair should be passed with the one of the service provider code selected from service providers’ database. After successful writing, interrogator will be acknowledged with new sequence number, written service provider code and attribute value to confirm the information and enable the future readings as shown in Figure 5.

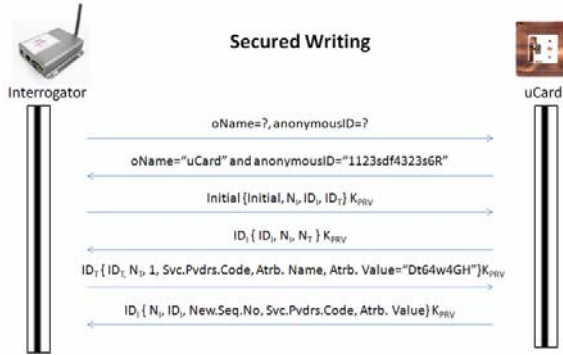


Fig. 5. Illustrate the secured writing protocol with private role key

4.4 Updating User Name and PIN

Same as in the secured reading and writing, updating of user name and PIN can be carry out in uCard. Both of them cannot be updated simultaneously instead one by one. This process needs two more passes than secured writing since it is necessary to confirm before making the actual update. Once the confirmation is issued by the reader the encryption key will be changed to the K'_{PRV} by composing a new encryption key with changed user name or PIN as shown in figure 6.

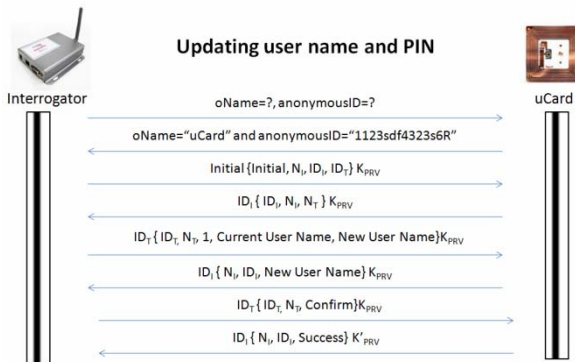


Fig. 6. Illustrate the secured writing protocol with private role key

5 Possibilities of uCard

Figure 7 illustrates the interested actors in communicating with the uCard. uCard can store the entities of several interested service providers to ease the locating of own records. Inside the uCard, within the entity, each service provider can use their raw data, encrypted data or else use the anonymousID which is available from the time of creation.

Private cards may be for hospitals, pharmacies, banks, electronic cash, access control, bus or train, student enrollment, parking permit, prepaid phone cards etc whereas public Cards may be of TV cards, game cards, restaurant cards, sports club cards, retail store loyalty cards, membership cards, amusement cards, vending machine cards, temporary access cards, clothing cards, dry cleaning cards, air ticket cards, gasoline stand cards, etc.



Fig. 7. Illustrate the possible communicators of uCard

Initially the card should be issued by the city governing authority or through several selected retailers. It must be founded by the service providers and should be free of charge to users. Once the uCard is issued to a customer, it can be used for any compatible subscriptions. They may be from loyalty card to bank card.

Once the customer with uCard starts subscribing services, the entities of service providers can be set in the public or private area of the uCard depending on the customer's wish. For instance if one customer make any purchases from retail store P, the retail store P can suggest the customer to use a loyalty card option in their store. If the customer wishes to participate in their loyalty scheme, retail store's entity will be written to the uCard owned by that particular customer. Then the profile related to the entity written to the uCard of customers can be built into the service providers' database by implementing the loyalty scheme on him. Whenever the customer visit the retail store P, relevant entity of the uCard can be read and allow receiving special services provided to the subscribers.

uCard protect its security and privacy by storing absolute attribute value in distributed way. That is one part inside the card and the other in the service providers' database. Similarly the level of desired security can be managed in uCard by storing the information in public or private areas. On the other hand no public reading can be

carry out without service providers' code whereas no private reading, private writing and public writing can be carry out without user name, PIN and service providers' code. Additionally, secure communicational protocols with encryption provides proper mutual authentication and there by guarantee the protection against the man in the middle and replay attacks. Furthermore, no repeated trials of entering wrong codes are allowed and no information belongs to other entities can be seen or retrieved by service providers even with proper authentication.

In case if uCard get corrupted, a new card can be issued from the card managing center and restore the old records after re-registering with service providers' system. When a customer loose the uCard or get it stolen, the card managing center must be informed immediately and their by issue a message to all the subscribers to suspends the transactions on lost or stolen uCard's anonymous ID. Then the risk on lost or stolen card can be alleviated. On the other hand, if attacker reverse engineers own or stolen uCard and receive the containing information, no big threat to service provider or customer occurs since the uniqueness of an entity depends on anonymous ID, sequence number, service providers' code, attribute name and attribute value. However there is a possibility to make several attacks with reverse engineered information though the risk can be reduced.

6 Concluding Remarks

Possibilities of using RF Chip with object oriented design to enhance loyalty programs efficiency are enormous; the only barrier is human imagination. uCard (Universal Card) which can be used to combine several cards to one card without creating any privacy or security issues is proposed. Further it provides the role base accessing and control in readings and writings. Finally the secured and non-secured protocols to communicate with uCard were also proposed. Since this is a feasible option, we are sure this solution will be workout in near future by creating a novel experience for customers and also for service providers by eliminating the hassle of managing several cards and decreasing the fear on personal privacy.

As future works, we are planning to improve this architecture to stand against the rest of the possible attacks, conduct formal verification on protocols, and extend the same system to embedded mobile phone card while enabling the web-card concepts in pervasive computing.

Acknowledgements

We would like to extend our sincere gratitude to Professor Takaomi Shigehara from the Department of Information and Computer Sciences in Saitama University on facilitating the environment for the above research.

References

1. Sabaragamu, K., Koralalage, Selim, M.R., Miura, J., Goto, Y., Cheng, J.: POP Method: An Approach to Enhance the Security and Privacy of RFID Systems Used in Product Lifecycle with an Anonymous Ownership Transferring Mechanism. In: Proc. SAC, pp. 270–275. ACM Press, New York (2007)

2. AIDC: Loyalty (Cards) Can Go Both Ways, Wednesday, February 18, 2009 - AIM Connections, <http://www.aimglobal.org/members/news/>
3. Nectar UK Loyalty Points Card, <http://www.nectar.com/NectarHome.nectar>
4. Loyalty program, http://en.wikipedia.org/wiki/Loyalty_program
5. Loyalty Schemes,
<http://www.moneysavingexpert.com/shopping/loyalty-scheme-tricks>
6. Koralalage, K.S., Yoshiura, N.: OTag: Architecture to Represent Real World Objects in RF Tags to improve future Intelligent Transportation Systems. *Journal of Convergence Information Technology* 4(2) (June 2009)
7. Garfinkel, S., Juels, A., Pappu, R.: RFID Privacy: An Overview of Problems and Proposed Solutions. *IEEE Security & Privacy*, 34–43 (2005)

VERTIGO: Find, Enjoy and Share Media Trails across Physical and Social Contexts

Luca Galli¹, Renata Guarneri¹, and Jukka Huhtamaki²

¹ CREATE-NET, Trento, Italy

² Hypermedia Laboratory, Tampere University of Technology, Tampere, Finland
lg@lgalli.it, renata.guarneri@create-net.org,
jukka.huhtamaki@tut.fi

Abstract. On the grounds of media sharing in social networks, visualisation mash-ups and early hypertext idea of associative “trails”, the opportunity for a new system to explore and enjoy connections among mainly linear media is devised. A novel system called VERTIGO is to be crafted to tackle the current limitations in media consumption across different contexts. On the end-user side, VERTIGO would provide a context-aware, personalised and proactive dashboard with innovative discovery and interaction capabilities; on the enterprise side, media professionals would gain access to a common platform enabling new collaboration models in the value network.

Keywords: Multimedia, Media consumption, Video, Television, Music, Context-awareness, Home, mobile, Geolocation, Social Network, Mashup, Visualisation, Scenarios, Business Model, Value Network.

1 Introduction

The wave of social networks and Web 2.0 services is putting the media industries under an increasing pressure. Long standing models in broadcasting, movies and music sectors are threatened by the overwhelming success of the Internet. Yet there is limited experience of successful endeavours in bringing the best out of the once clearly separated worlds of mass media and personal technologies. Interactive television experiments, for example, have been around for many years, but they have still to deliver on the promise of a big change [1]. Of course, games make for a huge realm, but despite some contaminations they are not to be mixed with mainly linear media. In this respect, novel approaches to create more interactive media experiences are needed as ever before [2]. VERTIGO¹ aims at tackling this challenge by focusing on the opportunity to expand and enrich the users’ capabilities in finding, enjoying and sharing media connections or “trails” across the different physical and social contexts. These media trails are now just part of the user personal experience, at individual

¹ “Vertigo” is one of Alfred Hitchcock’s most powerful, deep, and stunningly beautiful films (in widescreen 70 mm VistaVision); this classic 1958 noir, that functions on multiple levels simultaneously, has been a recurring case in the very initial project discussions and it is actually included in one of the project original scenarios.

level or in the close social sphere of her immediate companionship (e.g. when watching a movie with someone); social networks do have sharing features, but they are scattered among many services with limited interoperability and have little or no integration with the linear media consumption setting. The VERTIGO vision is to transform such currently ephemeral “media trails” into a working technology. The system, currently in the early design stage, envisages providing people with a dashboard that would inherit the old-fashioned TV remotes function by reshaping this rigid tool, tied to the living room environment, into a personalised, context-aware, proactive media-rich interface.

First, the “trail” concept in relation to media and social networks is introduced, as grounded in the historic developments of hypertext and illustrated with a mash-up example. Getting closer to market dynamics, a comparative perspective on multimedia consumption across the living room, desktop and mobile context allows highlighting key phenomena at the intersection of the technology and business dimensions. Then, following a scenario-based approach [3] VERTIGO is presented by means of two original stories then followed by some hints to related technology challenges and concluding remarks.

2 Visualising the Trails in Media and Social Networks

One of the most prominent phenomena of Web 2.0 is the concept of a *mashup*, an application that is “generated by combining content, presentation, or application functionality from disparate Web sources” [4]. Whereas in its simplest form a mashup is a combination of a map or some other visual instrument and a set of data, the recent definition by Yu et al. [4] gives a more elaborate context to mashup development. In fact, mashup development can be seen as a form of information visualisation where the underlying objective is to provide visualisation users with expressive views that serve as an amplifier of the cognition of a user and give insight on a certain phenomena represented by the data [5]. Complemented with information visualisation tools such as Prefuse Flare [6] and others [7], mashup technologies and other Web technologies enable the development of highly interactive information visualisations that, according to Kosara, Hauser and Gresh [8] “allows the user to implicitly form mental models of the correlations and relationships in the data, through recognition of patterns, marking or focusing in on those patterns, forming mental hypotheses and testing them, and so on”.

From the very beginning, the concept of a trail has been present in hypertext. While describing a fictional device memex, Vannevar Bush was the first to present the idea of hyperlink connecting pieces of information by association (instead of indexing) back in 1945, so in a similar fashion than the human mind operates: “The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature.” [9] Indeed, Bush added that “[m]an cannot hope fully to duplicate this mental

process [the intricacy of trails in human mind] artificially, but he certainly ought to be able to learn from it". However, different levels of association (e.g. geo-location, social ties, personal mood, activities) are already made explicit in social networks, even though it is difficult to trace and make use of them in the current scattered multimedia consumption landscape. To reuse McGonigal's [10] excellent metaphor inspired by Lewis Carroll, it would be possible to craft "rabbit holes" through which multimedia consumers are drawn into new multimedia landscapes beyond their imagination to find new, interesting piece of multimedia or a similar-minded fellow consumer [11]. The easy-to-access Web information (Web APIs) and text-based data formats (e.g. Extensible Markup Language) introduces great possibilities to implement such applications.

An example of a possible application is the (still fictional) Multimedia Trails. Collecting pieces of multimedia as hubs composing a set in interrelated multimedia, say a concert by Mondo Cane performing new versions of the Italian pop music of the 50's and 60s's. The setlist of the concert can be retrieved from Setlist.fm, the live recordings are available e.g. at <http://concertlive.co.uk> (in addition to unlicensed sources including BitTorrent), attendees of the concert from Last.fm, related videos can be collected from YouTube, images and trivia of the band from Last.fm API et cetera. In the screen capture (Fig. 1) we can see that the keyword-based filtering of videos expose both performances by Mondo Cane as versions by other actors.



Fig. 1. Example mashup of a Mondo Cane concert

Whereas individual interest hubs of multimedia can be visualised with methods of traditional mash-ups, the connections between the hubs or between individual appreciators of the digitalised culture can be represented e.g. as graphs in the tradition of social network analysis and visualisation [12] [13]. An excerpt of a social network is included in Figure 1: two users of Multimedia trails, Luca and Jukka, are connected to each other. Further, Jukka is interested in Mondo Cane that performs modern versions of Italian pop, one of Luca's favourite genres.

3 A Comparative Perspective on Multimedia Consumption across Three Key Contexts

Today, there are still clear boundaries between different multimedia consumption contexts. Each context has seen the growth of distinct ecosystems, very heterogeneous in

terms of business models, openness of markets and role of the user. Yet at the edges they start to overlap and build interfaces between each other. This is quite evident from the following discussion on three generic contexts corresponding to the main segments in the media business: broadcasting, the Internet and mobile telecommunications.

3.1 The Living Room Context

The living room is still dominated by classical mass media, even though various forms of interactive entertainment are making increasing inroads into it. Its traditional strength lies in creating high-quality laid-back entertainment and information experiences, focused on consumption: going beyond this legacy is an open challenge for those willing to innovate the interactive TV experience [14]. There are a number of multimedia solutions on the market targeting the living room, based on streaming access or personal digital libraries. Apart from large TV screens, set-top-boxes, game consoles or PCs in adapted design [15] have already gained a prominent role there. Proprietary platform solutions include the entire range of “media centre” software, e.g. Microsoft Windows Media Centre, the Apple TV and Nero MediaHome. Certain open solutions are available as well, such as the open source Media Portal [16] or Boxee. While those solutions rely on the Internet for content transmission, the rise of IPTV brought closed network end-to-end solutions like Microsoft Media Room, which offer a better quality of service for the sake of choice. Other products ranging from TiVO to the more recent Slingbox and Vudu that are aimed at either adding some interactivity or device-independence to the conventional audiovisual content consumption are also important. Recently, efforts to introduce Web content to TV devices have emerged with TV widgets (e.g. Yahoo! Connected TV). Moreover, Philips, Panasonic, Samsung and other companies are introducing TV devices that support CE-HTML. Apart from the Philips device, the Internet access will be restricted to a closed bouquet of services, where the manufacturer will act as gatekeeper. To generalise, interfaces are stylish and mostly usable, however limited in functionality.

3.2 The Desktop Context

The desktop, situated usually in an office or private room, is the context with the highest productivity potential. It offers the most efficient human machine interface for lean-forward media production and consumption. As an example, personal music collections are today often managed with desktop software. While e.g. Apple iTunes enables the creation of smart playlists composing music on basis of simple rules, the management and consumption of music is still controlled by hand. Music recommendation services such as Pandora.com or Last.fm Radio enable the editing of playlists and sharpen the user profile by evaluating user interactions, such as banning or adding a song. Web-based multimedia services dominate the desktop context and there is a plethora of them. Beauvisage [17], for example, reports that 63% of the time that a panel of French computer users spent on a computer was used to consume Internet services whereas dedicated multimedia applications were used only a 5% percent of the time. Five user profiles were defined on basis of the usage data: Web-oriented (42% of the users), Instant Messaging (IM) (14%), Multimedia (14%), Gaming (11%), and Serious (18%). Interestingly, multimedia consumption was overrepresented in the IM profile as was the

use of mobile phone communication [17]. Most remarkable recent developments in Web-based multimedia services include the trend to open service APIs to third parties and to create unified APIs, such as the OpenSocial API. This allows for creative service mash-ups, reinforcing the position of the World Wide Web as prime innovation source for media services. Songkick, for example, collects their tour information from MySpace, Last.fm and other sources. Big players such as Facebook are developing a platform strategy by integrating third party services. In Facebook alone, one is able to manage her musical interests with the Facebook default profiles and several applications including Music iLike, Living Social: Albums, and YouTube Video Box. But even though they all rely on the Facebook platform, their recommendation data is build up from scratch and they do not share it with other services. While standards like APML [18] will hopefully lead to better interoperability, the free flow of profile information is eventually dependent on business-level decisions made by the service providers. The state-of-the-art in video-centred Web applications offers another perspective. Major efforts are concentrated on the streaming quality and the social aspects are considered less. Services as Zattoo or Livestation concentrate almost only on the aggregation of channels on a dedicated video player; Hulu.com offers content from a group of TV broadcasters; the BBC iPlayer includes also a content download feature. Some other applications, instead, let users produce videos, mix, tag, and share them to communicate with other users. The most famous case is obviously YouTube and its many interesting features including video recommender, geographical video positions and the recent collaborative annotations tool. Overall, these services are pushing ahead a television consumption model in which the online desktop brings in more interactivity, even though they somewhat diminish the traditional social aspects of TV communal watching [19].

3.3 The Mobile Context

The mobile context traditional strength lies in communication and contextual services [20] [21]. Mobile media consumption has been moving from mass-market download offers to streaming services most suitable for high-end 3G and Wi-Fi enabled smart phones (see e.g. YouTube on the Apple iPhone). In parallel, broadcasters have been promoting 3G and DVB-H TV channels, apparently with limited success, at least in Europe. Moreover, users are provided also with alternative media aggregators such as Joost and Babelgum. Modern mobile devices can figure out their geo-location, but this is generally not used for current media applications. At the moment, multimedia management is still done with a desktop where smart playlists, multimedia stores and multimedia-ripping tools can be used to connect to different sources. Of course, mobile music is another area of huge interest. As Apple has brought a virtual iPod on their iPhone, other vendors are trying to catch up in this space; Nokia, for example, is pushing the offer of music packages with selected phones, integrated with subscription offers. Conventional broadcasting radio is also commonly available on mobile phones, with standard antennas or via streaming, with the addition of interactive extensions. Beside music listening, some more original features are emerging e.g. with the Last.fm applications for the iPhone and Google Android; they allows users to tap into some of the Last.fm social networking aspects of music consumption. Other recent socially-aware mobile applications include the recent iTunes DJ feature, where

people in e.g. a party can use the iPhone Remote application to vote and choose upcoming songs in the party playlist.

3.4 Device-Service Convergence and Platform Strategies

Drawing from the previous comparison, two general trends can be identified: the convergence of both devices and service provision on the one hand, and the raising of distinctive platform strategies on the other. Handheld devices, television sets and computers are constantly adding or combining features that were previously reserved for specific devices. As a consequence, the lines between the core functionalities of devices blur: any of them can meet some multimedia consumption (and creation) user' needs. This evolution goes hand in hand with the convergence of service provision as actors move in into new areas to explore new business models. The convergence of service provision is closely linked to platform strategies. This concept refers to actors in a value network vying for dominance by attempting to attract third-party services, content and end-users to their platform. This form of two- or sometimes multi-sided markets can take different forms and in the review above shows many diverging examples of organisations implementing this strategy.

4 VERTIGO Scenarios

4.1 Trails as Memory Helpers

Ann is a very sociable person and well educated in using online services (although she is not a techie). Outside working hours you will find her sitting at some fashion bar drinking an exotic cocktail with a friend, or posting on Facebook, or answering questions on LinkedIn. At the same time she really loves a quiet evening at home enjoying one of her favourite movies, just by herself, or with a few selected friends sharing similar interests.

Today

Ann really loves movies and anything that is in any way movie related, from music soundtracks to actor biographies and the like. She has a collection of them on DVDs, either bought or recorded from the TV or downloaded over the Internet. The funny thing is that when you open the case of one of her DVDs, there are a bunch of notes and half glued Post-ITs that fall out. These are things like: "This movie is a remake, I should find out what the original was and who was playing in it. Lisa would know. ASK!!!", or "There was a kind of controversial book about the film director. GET IT and READ IT", or "The scene of the girl walking alone on the Champs Elysées, really moving. Julia would have loved it. Next time invite her and enjoy together", or "This movie does not seem that funny anymore, probably last time I watched it, I was in a better company". It is always something personal: a reminder to check the sound track or call that friend who knows everything about that actor.

Tomorrow

Ann is even fonder of movies but she is now able to personalise every media experience in an online inter-linked stream of memory flashbacks and plans for the future.

She has a different way of keeping all of this together. Ann’s media dashboard interaction features drive her to expand beyond simple playing: she can post a comment, store it for future memory, share it with friends or possibly let it be public for everyone; get the answer to her questions from Lisa or Julia and also order that controversial book from a partner online bookstore. When she decides to watch again a movie, her history of watching that movie can be retrieved and the new experience will be added. All of this experience is visualised as a media living “add-on”, a starting point for further experiences. When she decides she wants to watch a new “noir”, she can get opinions from the “noir-lovers 24/24 7/7” group in the movies’ social network and make a selection based on this advise. She can see and visualise the “not-to-miss” movie of the week – on the basis of trusted opinions – and plan to watch it on TV. She can also check if her friends are watching the same program, or invite some of them to do so and share the experience.



Fig. 2. On the left, Ann is overwhelmed with stick notes as memory helpers; on the right, Markus is struck by the connections or “trails” between the Golden Gate and the Vertigo movie

4.2 Linking Places and Media

Markus is a long time independent traveller, quite a heavy music listener (having been himself active in playing with a college group) and avid reader of books and magazines. Now in his late forties, he just feels a bit frustrated by the complexity of pursuing his interests online, not being in the habit of moving from one service or social network to another, building profiles, installing and updating new software and the like. He is a bit more confident with cell phones though: he likes to fiddle around with them and has enough money to buy the good ones, even if he is not so much into the technical details.

Today

Markus is on vacation in the US, wandering around the Golden Gate Bridge in San Francisco. A wave of memories wells up in him: a key scene from a Hitchcock-classic, VERTIGO, was shot at this particular location and, of course, he forgot to include the soundtrack in his iPod playlist when he left the hotel in the morning (actually, he only has a CD-version of the soundtrack in his CD shelf). It occurs to him that that the bridge has starred in quite a few movies. He tries to figure out where to look for some more information now that he is has a brand new smart phone, with data access even in the US thanks to that roaming option he has subscribed before leaving. The difficulty is where to look for the information. Wikipedia entries are

difficult to read on the small screen; moreover, it should only be possible to listen to that track. This way, he would be sure to remember the movie!

Tomorrow

*A few years later Markus is back on the West Coast. As he is on location, the map on his smart phone entices him with icons pointing to multimedia “trails” found in this location. A dynamically composed playlist appears in his player including the main theme of VERTIGO as well as versions of other Hitchcock themes by Laika & the Cosmonauts, a long time favourite of Markus, and by Los Angeles Philharmonic conducted by Esa-Pekka Salonen, a fellow Finn. Also images and movie scenes appear in the playlist. He shuffles it until he notes one piece that reminds him of that famous movie he was pondering about the other day: the bridge is also featuring in the opening scene of *The Maltese Falcon*! He is able to post an extract to the media social network he is finally able to master, with an impromptu comment. What a better place to comment on a movie than its real stage! Media outdoors can be even more fun than at home – but certainly he is going to enjoy the full movie on his HDTV set when he returns home (perhaps with some more comments from friends).*

5 Towards the VERTIGO Contextual Media Dashboard

The scenarios presented above flesh out the VERTIGO vision from the end-user point of view; at the same time, they try to convey the idea of how to translate multimedia “trails” into working mechanisms.

From this latter technology perspective, VERTIGO is an extensible context-aware media delivery, exploration and management platform featuring a rich-media end-user control dashboard. The system taps the power of social networks and ubiquitous computing to change the ways in which media such as movies, video, music and all types of TV content are delivered, explored, managed, and enjoyed. To make this real, a number of challenges have to be faced in the areas of multi-context content and media service provision, social network-aware media consumption and sharing, personalisation and recommendation systems, adaptive, context-sensitive visualisations for media consumption, among others.

It is also worth to recall that the early ideas presented in this paper are the result of an initial common effort in which have taken part players from different industry sides, including a public broadcaster, a leading mobile operator, a major player in consumer electronics and set-top box manufacturing, in cooperation with research and innovation specialists. True openness and full interoperability with Web-based services was unanimously sought for, well beyond the addition of feeds or other limited layers of interactive content, as shown by the already available market offers. Current limitations are not only the result of established proprietary approaches, but in some cases they also stem from the conflict between competing platform strategies, be they based in the Internet or IT domains, or in the traditional broadcasting and media sectors. Yet there is an increasing awareness that the flexible and personalised interactions made popular by the Internet have to expand to the media and all across the various contexts in which media are consumed, shared and extended. In order to make this, users and professionals have to be provided with new interfaces, new services

and new expressive means. The discovery and enjoyment of “media trails” appears as a promising path, both from the user’ interest and the relevance in technology and business terms. In this respect, VERTIGO strives to be an experiment in which a new media value network might take shape.

Acknowledgments

The authors wish to acknowledge the contributions from all the VERTIGO project proposal consortium members (CREATE-NET, RBB, TXT Polymedia, Telecom Italia, Vestel Electronics, Optibase, Entropy, Almende, Tampere University of Technology, Fraunhofer FOKUS, IBBT, Aristotle University of Thessaloniki, Budapest University of Technology and Economics, Fondazione Ugo Bordoni, Lancaster University, CEDEO, University of Klagenfurt, LodgON), in particular Nils Walravens-IBBT (nils.walravens@vub.ac.be), Christian Raeck-Fraunhofer FOKUS (christian.raeck@fokus.fraunhofer.de), Björn Stockleben-RBB (Bjoern.Stockleben@rbb-online.de), Andrea Bernardini-FUB (a.berna@fub.it) and Leonardo Chiariglione (leonardo@chiariglione.org).

References

1. Cooper, W.: The interactive television user experience so far. In: Proceeding of the 1st international Conference on Designing interactive User Experiences For TV and Video (2008)
2. Cesar, P., Chorianopoulos, K., Jensen, J.F.: Social television and user interaction. *Comput. Entertain.* 6, 1 (2008)
3. Carroll, J.M.: Making use: scenarios and scenario-based design. In: Proceedings of the 3rd Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques (2000)
4. Yu, J., Benatallah, B., Casati, F., Daniel, F.: Understanding Mashup Development. *IEEE Internet Computing* 12(5), 44–52 (2008)
5. Ware, C.: *Information Visualization: Perception for Design*, 2nd edn. Elsevier, San Francisco (2004)
6. Prefuse Flare, <http://flare.prefuse.org/>
7. Visual Complexity, <http://www.visualcomplexity.com/>
8. Kosara, R., Hauser, H., Gresh, D.: An interaction view on information visualization. In: Proceedings of EUROGRAPHICS, pp. 123–138 (2003)
9. Bush, V.: As We May Think. *Interactions* 3(2), 35–46 (1996); Reprinted from *The Atlantic Monthly* 176(1) (July 1945)
10. McGonigal, J.: Down the rabbit hole. Presentation at Berkeley, CA, USA (May 5, 2005)
11. Heer, J.: Socializing Visualization. In: *Social Visualization Workshop, CHI 2006*, Montreal, Canada (2006)
12. Freeman, L.C.: Visualizing Social Networks. *Journal of Social Structure* 1(1), [np] (2000)
13. Heer, J., Boyd, d.: Vizster: Visualizing Online Social Networks. In: 2005 IEEE Symposium on Information Visualization (INFOVIS 2005), pp. 33–40. IEEE Computer Society, Washington (2005)
14. Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., Tschelegi, M.: Trends in the living room and beyond: results from ethnographic studies using creative and playful probing. *ACM Comput. Entertain.* 6(1), Article 5 (2008)

15. HP TouchSmart IQ500 Series PC,
<http://www.hp.com/united-states/campaigns/touchsmart/>
16. Media Portal, <http://www.team-mediaportal.com/>
17. Beauvisage, T.: Computer usage in daily life. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 575–584. ACM, New York (2009)
18. Attention Profiling Markup Language (APML), <http://www.apml.org>
19. Barkhuus, L.: Television on the internet: new practices, new viewers. In: Proceedings of the 27th international Conference on Human Factors in Computing Systems, CHI EA 2009, pp. 2479–2488. ACM, New York (2009)
20. Galli, L.: Context as content. The MobiLife project perspective, ITU/German Federal Network agency workshop on the regulatory environment for future mobile multimedia services, Mainz (2006)
21. Häyrynen, A. (ed.), Aftelak, A., Galli, L., Killström, U., Kurvinen, E., Lehmuskallio, H., Lähtenmaki, M., Mercer, K., Salovaara, A.: Users, Applications and Services, and User Centricity. In: Klemettinen, M. (ed.) Enabling Technologies for Mobile Services. The MobiLife Book, Wiley, London (2007)

VOICE: A Decentralized System for Ubiquitous VoIP Services

Costas Kalogiros¹, Costas Courcoubetis¹, and Panayotis Antoniadis²

¹ Athens University of Economics and Business*
Department of Computer Science
47A Evelpidon Str, Athens 11363, GR
`ckalog@aueb.gr`,
`courcou@aueb.gr`

² Université Pierre et Marie Curie, Paris6/LIP6
Paris, France
`panayotis.antoniadis@lip6.fr`

Abstract. We propose VOICE (Wireless vOICE), a fully distributed architecture capable of supporting and offering ubiquitous voice services over wireless IP-based data networks. Existing WiFi infrastructure - deployed by enterprises or even individuals - will be supplemented with software modules based on open standards in order to support VoIP. Furthermore, a new breed of VoIP providers by aggregating these currently isolated access networks will be able to offer alternative and advanced services to reachable end-users. These services would in certain cases completely bypass traditional telephony and cellular providers and in others smoothly cooperate with them. We will describe the overall architecture and the necessary modules that should be in place for realizing a system that creates a marketplace for communication services.

1 Introduction

VoIP (or IP Telephony) refers to the usage of IP technology for realizing telephone calls, as well as other more advanced communication services. It is widely believed that it has a major role in the evolution of the telecommunications industry from a centralized terrain to a competitive value chain of highly specialized providers. Voice nowadays is considered 'just another application', not a service totally controlled by the user's provider. This setting implies increased control to end-users, as they can deploy their own telephony servers and use their Internet connectivity to route calls to non-local destinations.

We believe the main technical reasons for this change in telecommunications market is a) the decoupling of end users from a single **identifier** and b) the ability of recent devices to use multiple access networks. These two developments acted as an antidote for monopoly termination; a situation that can arise when only a single provider can terminate calls to that user.

* This research project is co-financed by E.U.-European Social Fund (80%) and the Greek Ministry of Development-GSRT (20%).

Traditionally the user identifier was a phone number that was tightly related to user's location in the provider's network (like the local telephone switch of PSTN). Today, SIP and H.323 addresses can be used for identifying the destination party, and as a mean to resolve the ultimate network address (**locator**) where packets will be directed to. But, a mapping system must be in place in order for a provider to **search for alternative identifiers** and select the most appropriate (if more than one exist). Today a number of such systems are operational, including (Public) ENUM, Carrier ENUM and .tel domain.

The next step after collecting one or more identifiers for the call destination is to **search for network locators** (usually an IP address) of the subsequent signaling node where the call will be routed. There are cases when the next node (another signaling server, a telephony gateway or the destination) is not unique. This may be the case of users with multi-mode end devices, for example a 3G mobile phone with WLAN interface. Furthermore, imagine an end user that has activated multiple devices at the same time (i.e. home and office) and thus the incoming call notification must be forked to all available network addresses. Perhaps the more usual case is when the destination is a PSTN subscriber and there are multiple gateway providers that can terminate the call, each one with different cost and quality attributes (i.e. call blocking probability). Access to this information is achieved by specific routing protocols available today, like TRIP, H.501 and DUNDi. We propose a system, called VOICE, that has the potential to give providers another call routing alternative. Within the proposed network, currently isolated existing WLANs (e.g. hotspots deployed by cafes and restaurants) will form the access network upon which new attractive services will emerge and offered by interested parties. WiFi providers will be able to exploit their infrastructure in a flexible way by generating new business streams, while ITSPs (Internet Telephony Service Providers) will have the chance to provide innovative services to their customers and explore new business possibilities. We are interested mainly in Wireless LANs due to their widespread adoption by dedicated providers, municipalities and even individuals.

Calls from and to users within its coverage will be efficiently routed and in certain cases it will be possible even to completely bypass traditional fixed telephony and cellular network providers. Note that VOICE provides a mapping mechanism that allows incoming VoIP calls to end users, not only outgoing calls that is the usual case with traditional VoWLAN (Voice over WLAN) services. But, since it is unrealistic to assume that the proposed network will be sufficient enough to route every call by using exclusively its own pool of resources, trustworthy interfaces with traditional providers will be negotiated and put in place, once the network has recognized size and value.

The proposed system is characterised by its open and dynamic nature, since it facilitates interested parties to join and leave it by following flexible although well-defined procedures. On top of that, a solid management structure that will set the necessary rules for ensuring the operational integrity of the entire network will be established, covering a wide range of topics such as authentication of users, pricing, contract negotiation, accreditation of participants etc.

Flexible contract negotiation can have significant impact in VoIP value chain. Providers want to be compensated for their participation in call setup, either by direct payments or through a reciprocative mechanism. But this requires an established arrangement before hand and although VoIP peering is desirable for providers of the same size, currently it is very limited due to the complexity and time consuming process of negotiating interconnection terms. So they currently find it beneficial to enter in an agreement with wholesale ITSPs, also known as IXCs. This resulted in most ITSPs rejecting to process direct incoming calls from non-trusted providers.

The core functionality of WOICE is actually a mechanism for mapping between user identifiers like the different flavours of ENUM do. Actually it has many similarities with Carrier ENUM systems where routing information is shared between interconnected ITSPs, but in our case end users have given their consensus. However the dynamic nature of the routing information (mobile phone number to a SIP address for example) requires an automated procedure for finding if such mapping exists, either on demand or by protocols like TRIP and H.501.

We should note that WOICE is along the same line as other collaborative VoIP systems that are becoming widely popular (Skype, FWDout). Furthermore, Anagnostakis [3] has proposed a reciprocative mechanism for dealing with the 'free-riding' issue between VoIP peers that terminate calls to the PSTN. Similarly, Efstathiou and Polyzos [1], deal with the same issue in the Wireless LAN roaming context by using a token-exchange based accounting mechanism that incentivizes participating providers to offer network access to each other's customers.

This paper is organized as follows; Section 2 describes the general architecture and gives a simple usage scenario. The functionality of WOICE Aggregators and the 'middleware' functionality that is distributed in the WOICE network is summarized in Section 3 and 4 respectively. Finally, we conclude in Section 5.

2 The WOICE Architecture

We propose a two-layer, hierarchical system where Hotspot owners (WOICE WISPs) and WOICE ITSPs form the lower and upper layer respectively. From a functional point of view, the basic entity of the proposed system is the WOICE Aggregator. That is, an ITSP that can form her artificially owned access network by setting up bilateral agreements with WOICE WISPs and thus being able to terminate calls with low cost from/to the mobile telephones of users that enter one of the associated hotspots. In order to increase the value offered to its customers she can also set up agreements with traditional telecommunication providers (PSTN, cellular, ITSPs) in order to accept or place outgoing calls.

There exist two main options for WOICE Aggregators. In the first case, they could choose to act independently trying to attract the maximum possible number of hotspot owners. The other option is the creation of a collaborative VoIP network, a peer-to-peer community, of 'small' WOICE Aggregators which will

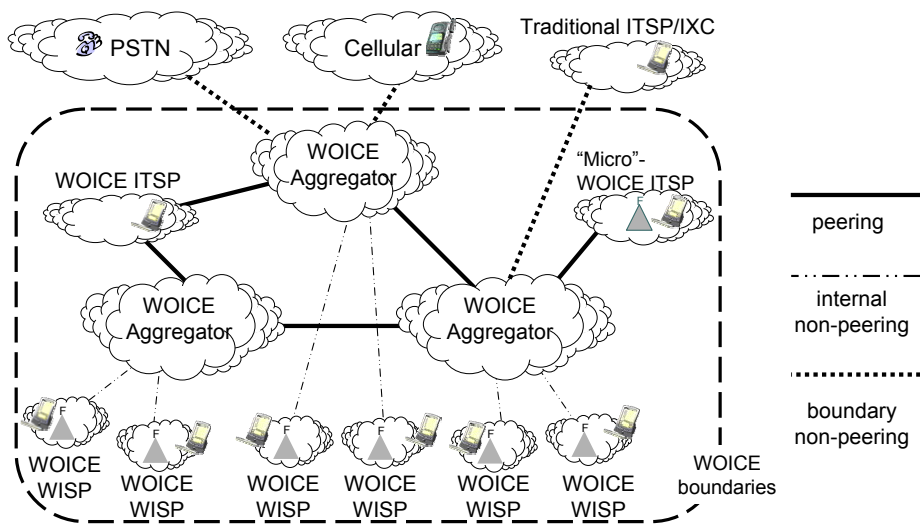


Fig. 1. The General VOICE architecture

exchange internally the call routing information of their networks in order to present aggregated call termination tables to external entities and be able to make profitable agreements with them. Note that in this case there is no restriction on the size of the enterprises that could participate in this VOICE network; even single VOICE WISPs with VoIP capabilities could play this role. Of course, besides the above extreme options there are many intermediate situations which one could think, such as a hierarchy of VOICE aggregators where some of them act as super-peers in the VOICE community.

Figure 1 shows a possible scenario of a VOICE network and its business relationships with traditional providers. The participants are of different types and size, showing the systems ability to support a large variety of business models. Each VOICE Aggregator has partnerships with a number of VOICE WISPs and may be interconnected with traditional providers. Other members of this VOICE community include a VOICE ITSP and a 'micro' VOICE ITSP. The former is a traditional ITSP with existing customer base and partnerships while the latter is a WISP with VoIP capabilities.

Figure 2 shows a usage scenario of VOICE. Lets assume that users A, B have dual-mode mobile phones that support the SIP signaling protocol. When, for example, 'User A' enters HotSpot, the VOICE WISP collects and verifies her mobile number. At the next step, the VOICE Aggregator is being informed about the new reachable number and subsequently advertises reachability to the VOICE ITSP so that calls can take place.

When 'User B' calls 'User A' on her traditional mobile phone number, the VOICE ITSP of 'User B' will search her routing table and find that the call can be terminated using end-to-end VoIP through VOICE Aggregator. VOICE

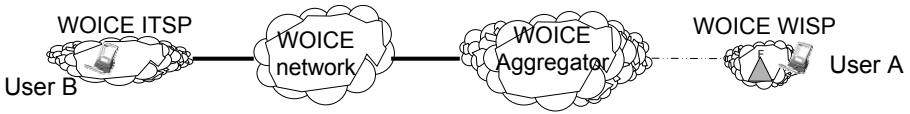


Fig. 2. A call placement scenario through WOICE system

ITSP sends a call setup notification (for example using SIP) to the WOICE Aggregator and exchange accounting information. Finally, WOICE ITSP and WOICE Aggregator rate each other and this reputation information is stored inside the WOICE network in a distributed way for enforcing rules in the future.

3 The WOICE Aggregator

We will present the functionality of WOICE incrementally. We will first describe the functionality for an independent WOICE Aggregator (no peer partners) relative to the interfaces that should be supported with his WOICE WISPs. Then we will analyse how this functionality may be implemented in a distributed fashion, so as to enable the creation of a large WOICE network and thus a variety of business models particularly attractive to a wide range of participants.

Figure 3 shows a high-level view of the core WOICE software modules and their interfaces that allow the interoperation of a standalone WOICE Aggregator with his partners. This resembles the case of a WOICE network with one big WOICE Aggregator.

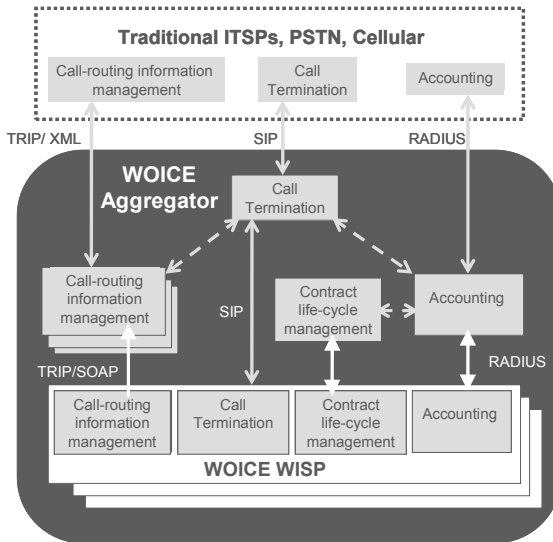


Fig. 3. Non-peer Interoperation

3.1 Contract Life-Cycle Management Module

VOICE Aggregators may have many bilateral contracts with VOICE WISPs. The most significant part of these agreements is usually the general contract terms, as well as the pricing scheme being used and each party's obligations (SLAs). For example, the two parties could agree that a Hot Spot visitor can place VoIP calls for free from his WiFi-mobile phone and the venue owner will be charged. In parallel, the same visitor can receive VoIP calls for free and the venue owner will receive a payment for using his Internet connection. This module automates the negotiation procedure by offering templates and guiding the transition from one contract term to another, until the agreement is either established or abandoned. The messages exchanged between the two providers represent their initial requests or bargaining responses.

3.2 Call-Routing Information Management Module

During the valid period of an agreement and every time a mobile-user enters the Hot Spot of a VOICE WISP the new routing information is propagated throughout the system and the external partnered entities. These actions are performed by the interoperation of each partner's Call-Routing Information Management module. Due to the importance of the VOICE Aggregator for the system's operation his module has interfaces for both intra-domain and inter-domain communication. The messages sent from each VOICE WISP to the VOICE Aggregator can be performed with TRIP or by web services technology (SOAP messages). The messages sent from the VOICE Aggregator to external partners can be performed with a call-routing protocol like TRIP or by XML files. In this way, the VOICE Aggregator can terminate calls from traditional providers to mobile users that reside in associated Hot Spots.

In [2] it has been concluded that in many cases providers can have the ability and incentive to ask for more than one providers to try and connect the same call (a technique called forking). This gives an incentive to providers to route calls to low-cost systems even when the call blocking probability is very high. Thus, even if providers don't have access to this routing information, they could still attempt to route the call through VOICE and a traditional provider.

3.3 Call Termination Module

Every time a VOICE Aggregator participates in a VoIP call (either initiated by or terminated to a mobile-user) the Call termination module of a number of entities must interoperate. These modules can reside on the users mobile phone, the VOICE Aggregators / external ITSPs signalling server or on a PSTN/cellular telephony gateway. The messages exchanged refer to a signalling protocol and perform call management (establishment, modification and termination).

3.4 Accounting Module

During the valid period of a bilateral agreement and every time a partner requests a service that is supported by the contracts terms of service, the two parties must

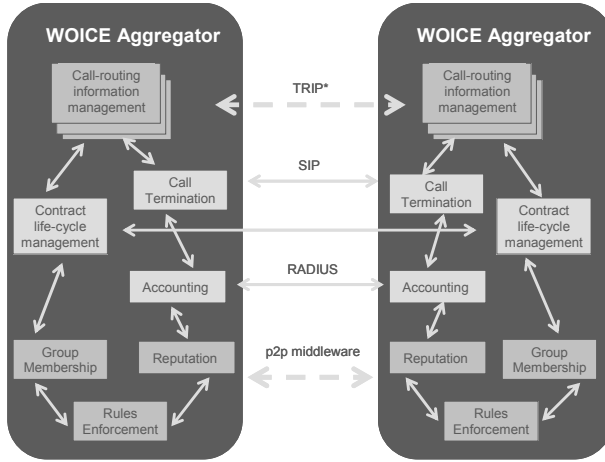


Fig. 4. Peer Interoperation

keep information regarding this transaction. For example, a WOICE WISPs request to WOICE Aggregator for terminating a mobile users call would drive the two partners to track the acceptance (or not) of request, duration, charges (or revenue shares) among others. This task is performed by the Accounting module of each provider that is based on the RADIUS accounting protocol for example.

4 The WOICE Network

We would like to enable the creation of a decentralized network of WOICE aggregators. For this to happen, one needs to implement specific 'middleware' modules that would support a) the distributed, efficient and accurate call routing information management, and b) group management, decision making, and enforcement of specific community rules that are needed to discourage selfish behaviour and ensure certain levels of QoS. In this way, the formation of a community is more likely to succeed. In the following, we summarize the functionality of each module.

4.1 Group Membership and Search

Group membership includes the fundamental procedures of joining and leaving the community, the management of identities and certificates, and the ability to exclude members. Furthermore, we will support the ability to search for WOICE aggregators with specific characteristics for the establishment of agreements. Examples of such attributes are reputation score and location.

4.2 Reputation

Reputation can be considered as a distributed accounting mechanism since its value for a specific peer expresses his 'aggregated' behaviour in the past. Reputation values are constructed through ratings of individual peers after a completion of a transaction or through monitoring of the behaviour of certain peers in the network. Issues such as 'cheap pseudonym' could be mitigated due to the ability to recognize domain names.

4.3 Rules Configuration and Enforcement

The existence of community rules which will ensure the desired behaviour of the participants will play a major role to the success of VOICE. However, we would like to offer substantial freedom to communities to configure these rules according to their needs (e.g. through voting or off-line configuration by the system designer, or dynamic adaptation according to some predefined measures). Examples of such parameters of community rules could be the constraints on the reputation threshold for a peer to be able to act as a super peer and the pricing policy of the community.

5 Conclusions

We described a collaborative system that allows offering of ubiquitous VoIP services to end users and flexible business models for its participants. In order for these targets to be met a) routing information is exchanged between providers that maps traditional user identifiers to VoIP identifiers and b) supplementary functionality is introduced that aligns participants' incentives.

References

- [1] Efstathiou, E., Polyzos, G.: A peer-to-peer approach to wireless LAN roaming. In: WMASH 2003, pp. 10–18 (2003)
- [2] Courcoubetis, C., Kalogiros, C., Weber, R.: Optimal Call Routing in VoIP. In: 21st International Teletraffic Congress (to appear, 2009)
- [3] Anagnostakis, K.: Exchange Mechanisms And Cooperative Distributed System Design. PhD thesis (2005)

Monitoring Ecosystem Characteristics on Negotiation Environments

Nicolás Hormazábal, Josep Lluís de la Rosa, and Gabriel Lopardo

Agents Research Lab, Universitat de Girona, Girona, Spain
{nicolash,pepluis,glopardo}@eia.udg.edu
<http://arlab.udg.edu>

Abstract. In this paper we suggest monitoring ecosystems' characteristics in negotiation environments, such as the diversity of the organizations' working areas, which can be used for recommendation for the whole ecosystem to promote the entrance of new organizations or the recommendation of adapting the organizations for improving their performance.

We will see the connection of performance and diversity and an illustrative experiment of how inviting diverse companies after an alarming descendent diversity of companies in a open negotiation environment impact in improved fitness of companies.

1 Introduction

In Digital Business Ecosystems (DBE), where networked organizations cooperate, compete and share knowledge, most of the agreements happen through negotiation processes [8]. In real ecosystems, species grow their population or disappear depending on the available resources they have, as well on their ability to adapt themselves to environmental changes. The organizations, the living species in the DBEs, must adapt themselves, evolve or disappear depending on the environment's and organizations' characteristics. For example, agreements not always can be made if organizations cannot find potential partners in a field.

The ONE project aims to develop a business oriented application for small and medium enterprises, through the creation of a distributed and peer to peer negotiation platform, flexible enough for allowing its adaptation to different industry sectors. It will support automated negotiation execution based on knowledge sharing and self-learning

Being analogous to real life ecosystems, their features such as heterogeneity, openness, stability, security and fitness of their population (Small and Medium Enterprises, SMEs), that impact on the distributed learning process can be characterised in the DBEs. Monitoring of ecosystem characteristics will guarantee equal opportunity, efficiency, effectiveness, trust and worth, and privacy.

The ONE (Open Negotiation Environment) project¹ aims to develop a business oriented application for small and medium enterprises, through the creation

¹ <http://one-project.eu>

of a distributed and peer to peer negotiation platform, flexible enough for allowing its adaptation to different industry sectors. It will support automated negotiation execution based on knowledge sharing and self-learning. Monitoring ecosystem characteristics in a negotiation environment such as ONE, will serve to have hints on the policies to take related to the whole environment such as improving its openness for increasing the heterogeneity.

This paper is organized as follows: Section 2 will mention some ecosystem characteristics and why they should be monitored. Section 3 will explain how the measure of the diversity can be used in a business environment. Section 4 will present a simple example on how the monitoring of the ecosystem characteristics can affect on the overall performance. Finally Section 5 will present conclusions and future work.

2 Ecosystem Characteristics

An ecosystem is a self-sustaining association of plants, animals, and the physical environment in which they live. It is composed by two main components: Community, which refers to the organisms (plants and animals) in an ecosystem, and Habitat, which refers to the type of environment that the organisms have adapted to and live in. In a DBE the community is composed by the SMEs, and the habitat is the negotiation environment.

The DBE concept is concerned with the development of an ecosystem inspired approach to the design of agent systems [4]. In this context an ecosystem can be viewed as an entity composed of one or more communities of living organisms, in which organisms conduct frequent, flexible local interactions with each other and with the environment that they inhabit [5]. Although the capability of each organism itself may be very simple, the collective behaviours and the overall functionality arising from their interactions exceed the capacities of any individual organism, as it happens from the resulting negotiations and subsequent coalitions between SMEs.

The introduction of new elements into an ecosystem can have a disruptive effect. In some cases, it can lead to ecological collapse or "trophic cascading" and the death of many species belonging to the ecosystem in question. Under this deterministic vision, the abstract notion of ecological health attempts to measure the robustness and recovery capacity of an ecosystem; that is, how far the ecosystem is can go away from its steady state ([1], [2]). Ecosystems can achieve and be maintained in states far from its equilibrium or steady state, through its openness (or non-isolation) property, which has as a consequence in the ecosystem's diversity [3].

Diversity has a significant role in Business environments as it can help providing equal opportunities for each different activity performed in the environment. The measure of heterogeneity or diversity is key to stabilize an ecosystem of companies with dynamic deals. The main causes of instability that nowadays are foreseen are: imperfect data (lack of data, bad data, wrong data, delayed data, distorted data, etc), dynamic market (chaotic changing demand, and increasing competition), and finally the lack of an appropriate supporting environment and knowledge.

The most common diversity measure, the Shannon-Wiener index, is entropy, giving the uncertainty in the outcome of a sampling process [6]. When it is calculated using logarithms to the base two, it is the minimum number of yes/no questions required, on the average, to determine the identity of a sampled species; it is the mean depth of a maximally efficient dichotomous key.

3 SME Activity Diversity Measure

Here we pretend to explain the significance of the diversity measure and how it can be used for the overall performance of the DBE.

In the DBE, SMEs can offer complex services and create new market opportunities, combining and sharing knowledge, products, services and infrastructures by joining together. Based on the Business Process Simulation (BPS) model for using it as example for this section, a business process has among other elements, resources and activities [7]. In this example, SMEs can run business processes where each related activity is developed by the different SMEs (being the resource element in the model) whose working field is related to it.

Each negotiation is focused on one or more items, often related to the partners specific business activity. In the case of the ONE platform, there will be different items being negotiated, and the diversity will depend on how many negotiations are being made containing each one of them. The diversity is important because a high level of diversity here, represents an even competition for each partner's interests or activities, what should result on good agreements as there should be enough options to choose from.

Let us take for example a football (or soccer) team. In a football team, there is the need of some kind of diversity level, not everybody has to be a defender, or a goalkeeper. The team needs every position covered, what gives as a result a high diversity level, but also not every position needs to have an equal number of elements, the optimal level of diversity depends on the environment's scope. In the case of the ONE platform, many offers will be available for different activities, then every activity should have enough partners to attend these offers, and compete between them. Let's take for example that there are 8 different activities represented in the ONE platform by different partners over a population of 1000 partners (table II).

Table 1. Activities distribution with high diversity

Service	Population
Cleaning	125
Security	125
Catering	125
Transport	125
Logistics	125
Porterage	125
Reception	125
Green Management	125

Using the Shannon-Wiener diversity index:

$$H' = - \sum_{i=1}^S p_i \ln(p_i) \quad (1)$$

$$p_i = \frac{n_i}{N} \quad (2)$$

Where:

- H' Is the Shannon-Wiener Index.
- p_i Is the relative abundance of each species. Calculated by the number of elements for each species (n_i) divided by the total number of elements (N).
- S Is the number of species.

It has a diversity index of $H' = 2,0794$, which is the maximum possible value. Normalizing the value, we have $H' = 1$. In this case, with high entropy, we can

Table 2. Activities distribution with high diversity

Service	Population
Cleaning	438
Security	88
Catering	124
Transport	36
Logistics	5
Porterage	48
Reception	163
Green Management	98

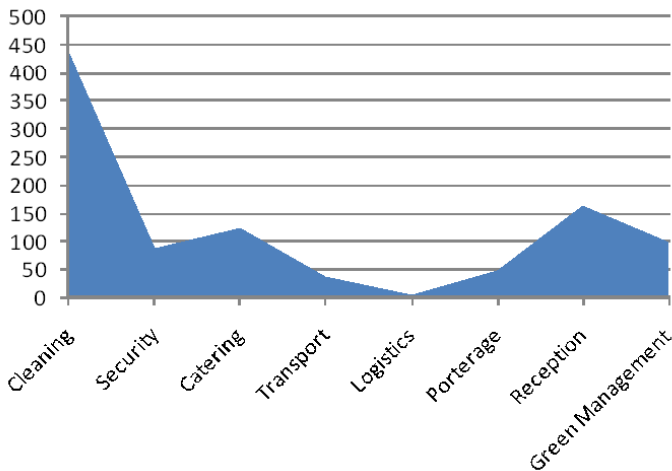


Fig. 1. Example of a low diverse distribution of the activities

see that there are a lot of competition on every activity, and also, every activity has enough people when an offer related to it arrives.

If the population is not so diverse as in the previous case (table 2).

The distribution is as shown in the figure 11. The diversity normalized (using the maximum index value, $H_{max} = \ln(S)$) in this case is $H' = 0,7933$. We have too many in the Cleaning, and a lack in Logistics, this is because the diversity index has been affected, and probably will affect on the result on the negotiations.

4 Diversity Measure Example

SMEs can look for other enterprises which perform complimentary activities, and together perform complex business processes by the aggregation of their services. We developed a simple simulation in Repast². Repast is a free and open source agent-based modelling toolkit developed in Java. The simulation represents an environment where a population of different resources (the species), where each one of them can perform only one activity, have to look for another resource which can perform another different activity, if they do not find it it means that they fail to create a business process. For keeping the simulation as simple as possible, at each timestep the SMEs look for another with a random activity different from the activity they can perform.

We will track the activities' diversity at each timestep, and try to find a relationship between diversity and business processes creation failure. It is supposed that at a higher diversity value, there will be a lower number of failed business processes creation. This is because at higher diversity values, the resources will have more options to choose from for each activity. The equation used for calculating the diversity will be the Shannon-Wiener Index explained in the previous section.

For this case, the species will be recognized by the resources activities. So each species will be represented by an activity. We will run the experiment starting with a low diversity inserting the 75% of the elements belonging to the same species (this means, 75% of SMEs that perform the same activity) and the 25% remaining will belong to a random species (there will be a total of 100 elements and 8 different species). At each timestep, the half of the population will try to create a business deal, if they do not succeed because they did not achieve to find another resource capable to perform a different activity, it will be recorded a *business deal creation failure*.

From the timestep 5000 onward depending on the diversity value, the environment will allow the entrance of new resources which will replace the current ones, keeping the population's number. The system will try to keep the diversity as high as possible (above 0.98 for this experiment), the system will remove a random resource and add a new one that performs a random activity. This should make the diversity raise from timestep 5000, and reduce the number of *business deal creation failure* at each timestep (figure 2).

² <http://repast.sourceforge.net>

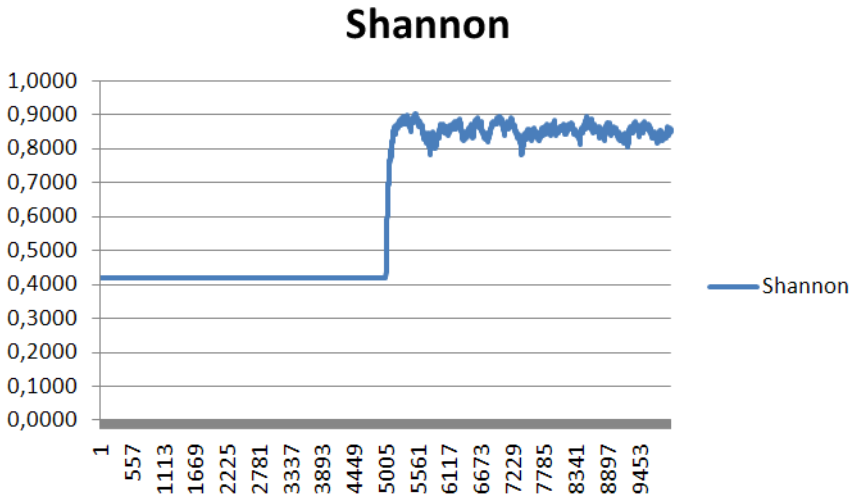


Fig. 2. Average Business Deal Creation Failure and Diversity

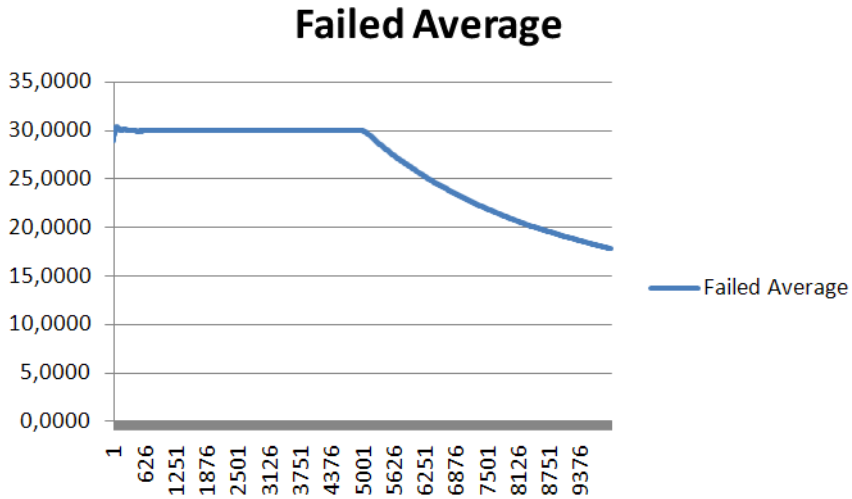


Fig. 3. Average Business Deal Creation Failure and Diversity

As expected, the average business deal creation failure per time-step falls once the environment "opens" itself allowing the entrance and departure of the resources. This can be seen on the figure 3, where after the time step 5000 the average business deal creation failure drops significantly. This could seem obvious, but the real application here is the use of the diversity monitor to encourage the DBE members to enter or exit for the environment.

From an administration point of view, an ecosystem characteristics monitor can give significant information to promote actions on the whole system which could improve the overall performance.

5 Conclusions and Future Work

An essential element of ecosystems is the negotiation of alliances, which enable companies to join competencies as well as services and products into a complex offering. Given this, DBEs should be empowered with a tool to support tactical negotiation and agreement processes among participants. This environment should support the creation of Virtual Organizations with a common business goal, offering complex services by the aggregation of its member's services, and should facilitate the building, stabilization, and improvement of the ecosystem performance on a shorter time frame.

The measure of ecology's properties in business environments can have different applications. In the field of recommender systems for example, it can be used for recommending on the negotiation style for each user in a negotiation environment [9]. In the current paper we present another application on the ecosystem monitoring on a business environment, but the application is not for the individual user; is for the environment as a whole. The data extracted from a ecosystem monitoring application can be used for improving the whole system, helping to take decisions like promoting the entrance of new members in the system (which can be translated on improving the ecosystem's openness).

In this paper we presented an example, in a simplified environment on how the diversity measure could affect on it, and how could be improved. Normally, real DBEs are more complex and not always the highest diversity index is better. Each environment has to be studied and evaluated, for example in business environments not every service has the same demand and then, not every service should have the same offer, this means that not necessarily the services' diversity index should be at the highest value.

Currently we have developed a monitoring tool for the ONE platform, and in the future it is expected to use its output to make recommendations for improving the system, but first a relation between the retrieved data and the system's results must be defined, in the meantime it is being used for monitoring the overall activity of the system within different metrics. As said before, it is necessary to find which are the best values on the indexes for the best overall performance, and not always look for maximizing them (like the case of the diversity index), and that can be done only observing the activity on the running platform.

Acknowledgments. This research was partially funded by the European Union project Num. N 34744 ONE: Open Negotiation Environment, FP6-2005-IST-5, 216746 PReservation Organizations using Tools in AAgent Environments (PRO-TAGE), FP7-2007- ICT Challenge 4: Digital libraries and content, SIE 2007 Motor intelligent de recuperació i processament semantic dinformació en base a criteris de personalització, a grant from the ACC10 Catalan Government, and the PR2007-0432 Salvador de Madariaga Grant.

References

1. Christopherson, R.W.: *Geosystems: An Introduction to Physical Geography*, 3rd edn. Prentice Hall Inc., Upper Saddle River (1997)
2. Wilson, E.O.: *The Diversity of Life*. W.W. Norton & Company Inc., New York (1992)
3. Jorgensen, S.E., Patten, B.C., Straskraba, M.: Ecosystems emerging: 3. Openness, *Ecological Modelling* 117(1), 41–64 (1999)
4. Nachira, F., Le Louarn, M., Rivera Lèon, L. (eds.): *Building the Foundations of Digital Ecosystems: FP6 Results and Perspectives*. European Commission, Brussels (2007)
5. Waring, P.H.: *Ecosystems: Fluxes of Matter and Energy*. In: Cherrett, J.M. (ed.) *Ecological Concepts*. Blackwell, Oxford (1989)
6. Magurran, A.E.: *Measuring Biological Diversity*. Blackwell Publishing, Oxford (2004)
7. Tumay, K.: Business process simulation. In: *Proceedings of the 28th conference on Winter simulation*, Coronado, California, United States, pp. 93–98 (1996)
8. Digital Ecosystems, <http://www.digital-ecosystems.org/>
9. de la Rosa, J.L., Lopardo, G., Hormazábal, N., Montaner, M.: A negotiation styles recommender Approach Based on Computational Ecology. In: *XI International Congress of the Catalan Association of Artificial Intelligence (CCIA 2008)*, *Frontiers in Artificial Intelligence and Applications*, vol. 184, pp. 293–302 (2008)

Incentives, Business Model and Required Technical Infrastructure for the Economic Aware Data Grid

Dang Minh Quan^{1,*}, Orestis Terzidis², and Sonja Klingert¹

¹ School of Information Technology, International University in Germany, Germany
² SAP research, Germany

Abstract. The data Grids appear mostly in the scientific community but it is also necessary to have the data Grid in the industry sector. Applying the way of scientific data Grid to the industry brings many disadvantages such as expensive investment, not efficient usage and unfair sharing. This paper presents an alternative approach, an economic aware data Grid. The incentives of having an economic aware data Grid, the business model and the working mechanism are analyzed in depth.

1 Introduction

The data Grid appears from the requirement of sharing, accessing, transporting, processing, and managing large data collections distributed worldwide. At the present, the data Grids appear mostly in the scientific community [1]. A data Grid usually includes many High Performance Computing Centers (HPCCs) with enormous storage capacity and computing power. It is obvious that investing a data Grid need a lot of money. The financial sources for building such data Grid are from governments and scientific funding foundations. For that reason, researchers within the data Grid can use those resources freely.

In the industry sector, the data Grid is also necessary. We can look at the scenario of investment banks. The investment bank has many branches distributed over the geography. Each branch has its own business. Each branch usually runs data mining applications over the set of collected financial data. With the time, this activity becomes important and need to be extended. The computing tasks need more computing power, storage. The data source is not just from the branch but also from other branches. Because all the branches belong to the investment bank, the data can be shared among branches with a suitable authorization policy. Thus, it is necessary to have a data Grid in the investment bank. We can see many similar scenarios in the real world such as an auto producer with many distributed branches and so on.

To build such a data Grid, one solution is applying the way of scientific data Grid. Each branch invests to build its own computer center. Those computer centers are then connected together to form a data Grid. The users in each branch

* Corresponding author.

can use the Grid freely. However, such approach brings several disadvantages. It cost a great amount of money for hardware/software investment, electric power, personal and so on. The resource usage is not efficient. Usually the data mining applications are executed when all the financial data are collected. This point is usually at the end of month, end of quarter or end of year. At those time periods, all computing resources are employed and the workload is 100%. In normal time period, the workload is lower and thus many computers run wastefully or unused. There may have unfair resource usage on the Grid. Some branches may contribute little resource to the Grid but use a lot.

An alternative approach is outsourcing. This means each branch does not invest to build a computer center it self but hiring from resource providers and pay per use. In other word, the investment bank will build a data Grid over the cloud environment. This approach brings many benefits to the investment bank and its branches. Some of them are flexibility, efficiency and fair sharing.

However, up to now, there is no business model and technical solution to realize this approach. The work in this paper is the first attempt to solve this issue. In particular, the contribution of the paper includes:

- The incentive of having an economic aware data Grid. The paper will discuss the advantage of the economic aware data Grid.
- The business model. We will analyze the business relation among many business components joining the data Grid.
- The high level system architecture to realize the economic aware data Grid.

The paper is organized as follows. Section 2 describes the related works. Section 3 presents the incentives of having the economic aware data Grid and Section 4 describes the business model. The system architecture is described in section 5. Section 6 concludes the paper with a short summary.

2 Related Works

Current large-scale data grid projects include the Biomedical Informatics Research Network (BIRN) [2], the Southern California Earthquake Center (SCEC) [3], and the Real-time Observatories, Applications, and Data management Network (ROADNet) [4], all of which make use of the SDSC Storage Resource Broker as the underlying data grid technology. These applications require widely distributed access to data by many people in many places. The data grid creates virtual collaborative environments that support distributed but coordinated scientific and engineering research. The economic aspects are not considered in those projects.

The literature recorded many works related to utility computing in the Computational Grid [5,6,7,8]. Many economic principles are analyzed in the context of the utility Grid which allow resource owners, users, and other stakeholders to make value and policy statements. However all of them considered only the computation resources. The combination of many data sources to form a data Grid as described in our scenario has not been mentioned.

Related to the data Grid, in [9] a cost model for distributed and replicated data over a wide area network is presented. Cost factors for the model are the network, data server and application specific costs. Furthermore, the problem of job execution is discussed under the viewpoint of sending the job to the required data (code mobility) or sending data to a local site and executing the job locally (data mobility). However, in the model, the cost is not money but the time to perform the job execution. With this assumption, the system is pseudo economic-aware. More over, the infrastructure works with the best effort mechanism. The QoS and resource reservation are not considered. Thus, it does not suit with the business environment.

Heiser et. al. [10] proposed a commodity market of storage space within the Mungi operating system. In which, the system focuses on the extra accounting system used for banking store management. The model is designed such that all accounting can be done asynchronously to operations on storage objects, and hence without slowing down such operations. It is based on bank accounts from which rent is collected for the storage occupied by objects. Rent automatically increases as available storage runs low, forcing users to release unneeded storage. Bank accounts receive income, with a taxation system being used to prevent excessive build up of funds on underutilised accounts. However, the system considers only the storage resource and the scope of the system is just inside an organization.

Buyya in [11] discusses the possible use of economy in a scientific data Grid environment, in particular in the DataGrid project. Specifically, a token-exchange approach is proposed to regulate demand for data access from the servers of the data grid. For example, a token may correspond to 10KB of data volume. By default, a single user may only access as much data as he has tokens. This gives other users a chance to access data. The amount of data that they access for a given token needs to be based on parameters such as demand, system load, QoS requested, etc. The users can trade-off between QoS and tokens. However, the negotiation/redistribution of tokens after their expiration, their mapping to real money and the pricing policies of storage servers are not discussed in such an approach. More over, this work focuses on the resource providers level while in our work we focus on the system built up over the commercial resource providers.

3 The Incentive of Having the Economic Aware Data Grid

The economic aware data Grid brings many advantages. First of all, it is flexibility. Because of employing the cloud computing environment, each branch could have Grid resources on demand. In the critical period, they could hire more resources from the cloud. When they do not need Grid resources any more, they could return them to the cloud.

It is economic efficiency. Assume that each month, a branch needs 10.000 GB storage, 10 computers for the daily workload in 28 days, and needs 100 computers for the critical workload in 2 days. The investment of such a cluster is about 1

million Euro. Assume that the life cycle of the cluster is 5 years. Assume that the branch spends 20.000 each month including the personal cost, electricity and so on for running the cluster. Thus, the branch must pay 2,2 million Euro in 5 years. If the branch uses Cloud services, according Amazon pricing schema, the cost in 5 years is approximately 200.000.

The economic aware data Grid provides a fair resource usage. Each branch independently pays for its Grid resource usage. Thus, if a branch uses many Grid resources, it has to pay more and vice versa.

4 The Business Model Proposal

Figure 1 presents the business relationship among many components joining the system. The business model includes three main components.

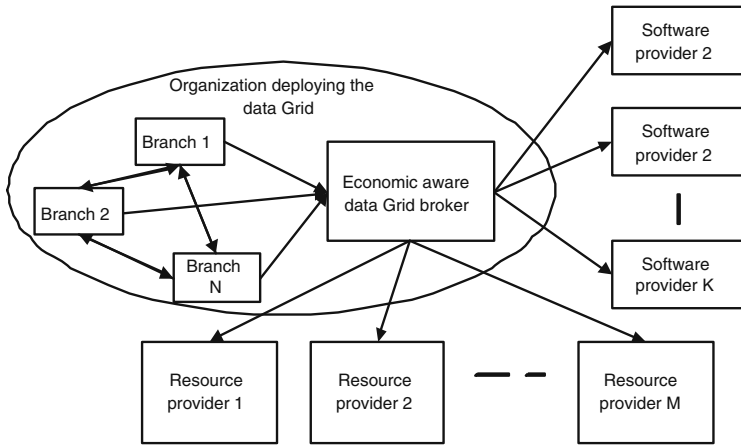


Fig. 1. The business model of the system

4.1 Resource Provider

For the utility computing business model, the resource provider provides storage capacity and computing power service. Thus, the computing power, storage, and network bandwidth are main resources to be traded. In the real environment, for example the Amazon pricing schema, the cost of network bandwidth, which is visualized through the data transfer in/out, is located inside the cost of the storage service and the computing power service. The provider should have its own accounting, charging, and billing module and metering, job deployment and so on. We assume that the price of using resource is published. To ensure the QoS, the resource provider should have advance resource reservation.

The resource provider charges the user using storage service following fees: fee of using storage, fee of having data transfer in/out, fee of having request to manipulate the data. The resource provider charges the user using computation service following fees: fee of using computing nodes, fee of having data transfer in/out.

4.2 Software Provider

Software provider is a business entity having software. The software provider provides software service. In particular, he provides software and its license so the software can work under negotiated condition. The income of the software provider is from selling software license.

4.3 Organization Deploying the Data Grid

Each branch should have its own separate data storage on the Grid because the data is very important with each branch. In particular, each branch hires basic storage service from a resource provider. The data are then shared with many authorization levels to other branches. To coordinate the working of many data sources and data exploitation activities, it is necessary to have a Grid middle ware called economic aware data Grid broker.

The economic aware data Grid broker should belong to the organization deploying the data Grid because of following reasons.

- It saves the organization many cost of using broker service.
- It is easier for organization to apply cost optimization policy when having its own control system.
- Giving data management task for a third party is not as trustable as by its self.

The economic aware data Grid broker does two main tasks. The first is the data management service. It includes data transfer service, replication service, authorization and so on. The second task is the job execution service. It receives the requirement from user, gets software, locates data, reserves computing resource, does deployment, runs the software and return the result. In fact, two tasks relate closely with each other. When the job produces output data, it must also be stored/replicated some where on the data Grid. The contract with software provider and resource provider are realized with Service Level Agreement (SLA).

The user in each branch puts/gets/finds data and runs jobs on the data Grid. Thus the branch has to pay the cost of using storage and computing power for resource provider, the cost of using software for the software providers. From the cost of using storage service, we can see that the cost includes data transfer in/out cost. Thus, if user of branch 2 does many copies from the data storage of branch 1, letting branch 1 pay for the transfer cost is unfair. Thus, it is necessary to have payment among branches to ensure that every branch is fine. They use much they have to pay much. They use less then they have to pay less.

The goal of the economic aware data Grid broker is managing the data Grid in away that minimize the cost of the organization.

5 High Level System Architecture

Figure 2 presents the high level system architecture of the economic aware data Grid. In this section, we will describe the working mechanism of the system.

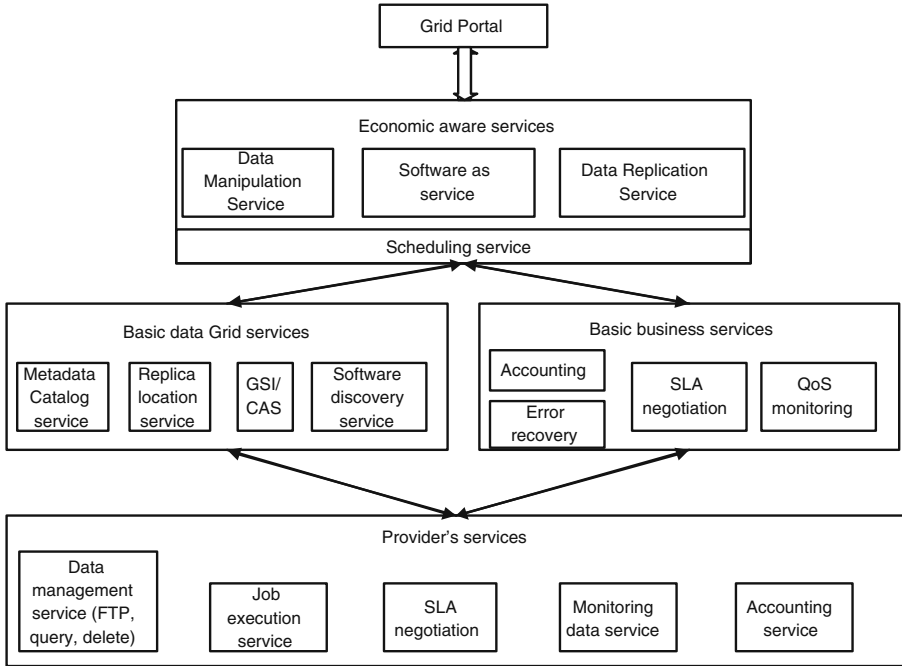


Fig. 2. High level system components

5.1 Data Manipulation

The user of a branch puts/finds/downloads/deletes a file to the Grid. As each branch has a separate storage on the Grid, the file should be put to that storage. In this case, the system will do following works:

- The Grid receives the requirement with the Grid Portal. The Grid Portal invokes the Metadata Catalog Service to find the appropriate information. If the request is put, the system returns the data storage. If the request is find/query/delete, the system returns the data location.
- The Grid Portal invokes services provided by the service provider to handle the request.
- When the request completes or fails, the Grid Portal notifies to the user. If success, the Grid Portal stores data information to Metadata Catalog Service, Replica Location service, stores the accounting information to the accounting service.

5.2 Replication

The request for replication appears in many cases: The system needs to replicate a file on Grid following a defined replication plan defined; The user demands a file

replication; A job has a request for a particular file in the Grid; The replication can also be triggered by analyzing the patterns of previous file requests. Thus, there is the migration of files toward sites that show a corresponding increased frequency of file access requests. When having the replication request, the system will do following works.

- The Data Replication service receives request, read and interpret it.
- The Data Replication service invokes scheduling service to find out a suitable replication location.
- The scheduling service discovers candidate resources, matches the user's requirements and the candidate resources in an optimal way and then returns to the Data Replication service.
- The Data Replication service reserves bandwidth with resource provider by an SLA.
- The Data Replication service invokes the FTP service of the determined resource provider for transferring data.
- The Data Replication service invokes monitoring module to monitor the QoS.
- If success, the Data Replication service stores data information to Metadata Catalog Service, Replica Location service, stores the accounting information to the accounting service.

5.3 Job Execution

When the user wants to run a job, he provides the information of the job such as software name, name of the input/output data, resource requirement, deadline and so on. The system will do following works.

- The Grid Portal invokes the SaaS service.
- The SaaS invokes the Software Discovery service to find the location of the software provider.
- The SaaS invokes the Metadata Catalog Service, Replica Location service to find the location of data file.
- The SaaS invokes Scheduling service to find the suitable resource provider.
- The SaaS signs SLA of hiring software, computing resource, bandwidth with software provider, resource provider.
- The SaaS download the software and data to the execution site and execute the job.
- During the execution, monitoring module is invokes to observe the QoS.
- During the execution, if having some error, SaaS will revoke the error recovery.
- When the execution finishes, the SaaS move the output data to the defined placed and update Metadata Catalog Service, Replica Location service.
- The SaaS store accounting information to the account service.

It is noted that unlike the general SaaS, the number of software used in the system is not so big. Thus we need only a simple Software discovery service.

6 Conclusion

This paper has presented the incentive, the business model and the required technical infrastructure of the economic aware data Grid. The economic data aware Grid has many benefits comparing to the traditional form of the data Grid such as flexibility, efficiency and fair sharing. The business relation among business entities is clearly defined to gain the fairness. The high level architecture and is described to have an overview of the working mechanism of the system. This is the first step in realizing the whole system.

References

1. Groep, D., Templon, J., Loomis, C.: Crunching real data on the Grid: practice and experience with the European DataGrid. *Concurrency Computat.: Pract. Exp.* 18, 925–940 (2006)
2. Peltier, S.T., Ellisman, M.H.: *The Biomedical Informatics Research Network, The Grid, Blueprint for a New Computing Infrastructure*, 2nd edn. Elsevier, Amsterdam (2003)
3. Southern California Earthquake Data Center team: *Seismogram Transfer Program (STP) Reference Manual*, http://www.data.scec.org/STP/STP_Manual_v1.01.pdf
4. Fricke, T., Ludaescher, B., Altintas, I., Lindquist, K., Hansen, T., Rajasekar, A., Vernon, F., Orcutt, J.: Integration of Kepler with ROADNet: Visual Dataflow Design with Real-time Geophysical Data. Poster in American Geophysical Union Fall Meeting (2004)
5. Treaster, M., Kiyancilar, N., Koenig, G.A., Yurcik, W.: A Distributed Economics-based Infrastructure for Utility Computing, *ArXiv Computer Science e-prints* (2004)
6. Brunelle, J., Hurst, P., Huth, J., Kang, L., Ng, C., Parkes, D.C., Seltzer, M., Shank, J., Youssef, S.: EGG: an extensible and economics-inspired open grid computing platform. In: *Proceeding of the 3rd International Workshop on Grid Economics and Business Models, GECON 2006*, pp. 140–150 (2006)
7. Wolski, R., Brevik, J., Plank, J.S., Bryan, T.: *Grid Resource Allocation and Control Using Computational Economies*. In: Berman (ed.) *Grid Computing: Making The Global Infrastructure a Reality*. John Wiley & Sons, Chichester (2003)
8. Buyya, R.: *Economic-based Distributed Resource Management and Scheduling for Grid Computing*, PhD thesis at School of Computer Science and Software Engineering Monash University, Melbourne, Australia (2002)
9. Stockinger, H., Stockinger, K., Schikuta, E., Willers, I.: Towards a Cost Model for Distributed and Replicated Data Stores. In: *Proceedings of the 9th Euromicro Workshop on Parallel and Distributed Processing (PDP 2001)*, pp. 461–467 (2001)
10. Heiser, G., Elphinstone, K., Vochtelloo, J., Russell, S., Liedtke, J.: The Mungi single-address-space operating system. *Software—Practice & Experience* 29(9), 901–928 (1998)
11. Buyya, R., Abramson, D., Giddy, J., Stockinger, H.: Economic models for resource management and scheduling in Grid computing. *Concurrency Computat.: Pract. Exper.* 14, 1507–1542 (2002)

The Incoming Trends of End-User Driven Service Creation

Zhenzhen Zhao¹, Nassim Laga², and Noel Crespi¹

¹ Institut TELECOM, TELECOM SudParis, Rue Charles Fourier. 9,
91000 Evry, France

{zhenzhen.zhao,noel.crespi}@it-sudparis.eu

² Orange Labs, France Telecom R&D, Rue des Coutures. 42,
14000 Caen, France

nassim.laga@orange-ftgroup.com

Abstract. Service creation is the next big end-user driven hype in the technology arena. As the first Web 2.0 services are already enabling tech-inclined users to build their own composite services, this trend will lead to a huge set of new features and services will be provided by all stakeholders. This paper illustrates the incoming trends of the end-user driven service creation with extensive new features, and provides five main research topics from both technical and business perspective. The underlying key technologies, as well as the current platforms/tools are also detailed.

Keywords: Web 2.0, Service creation, Semantic, Context-aware, Community.

1 Introduction

Web 2.0 puts people into the center of the vision. Communication, collaboration, collective intelligence, and social interaction between people are what Web 2.0 is about much more than the technology pieces [1]. The past few years have witnessed end-users' increasing involvement in the content creation process (e.g. Wikipedia, Blogs and YouTube), which is of great benefit from collective user-generated content [2]. Currently, a growing user community is also trying to profit from existing services to generate its own services, which triggers the trends of the end-user driven service creation.

Towards this new trend, generally, three questions are raised.

- Why should end-users create services?
- What are the ongoing works on end-user driven service creation?
- What are the challenges of enabling end-users to do future service creation?

Like different people have different personalities, the best way to achieve user personalization is giving way to end-users selves to develop services satisfying their specific needs. Telecom operators and service providers offer an increasing number of services to their customers, but it's impossible to foresee all the conditions that a customer can experience in his daily life. From the business perspective, the main idea behind the

end-user service creation is to shorten the analysis/development of the marketing chain, harness the collective intelligence and promote innovation in software development. Of course the services that end-users create won't be as sophisticated as professional services, but it will give end-users the possibility to differentiate themselves.

The service creation process is based on the concept of service composition, which allows creating/assembling a new service from existing services or a set of basic piece of services (called service enablers in Open Mobile Alliance (OMA) [3]). The guideline for designing and developing a service creation platform is "Do half work. Let users do the other half." Since the services are the results of a composition of a set of pre-defined services, the "half work" should be done is to provide the intelligent sub-services that allow for the creation and orchestration of enhanced services in a personal, dynamic and multi-domain environment.

Designing such a platform is not easy due to the complexity of current service oriented architecture (SOA) based technologies [4]. Currently programmers create easily Mashups [5], which is a customized combination of multiple functionalities of heterogeneous web sources or web services APIs. They do easily using either SOA technologies (such as REST [6] and WSDL [7]/SOAP [8]), or invoking and reading RSS feeds, or finally by invoking and extracting data from web sites through screen scraping [9]. These methods are accessible for each novice in the web development, but they are not understandable by non-developer people. Therefore, new mashup creation tools have emerged such as YahooPipes [10], MARMITE [11] and Microsoft Popfly [12] which are more designed for advanced users (tech-inclined users), but they remain not intuitive enough as ordinary users still find them difficult to use.

To reach an effective end-user service creation we have to go through several challenges. The first one consists in how to provide the intuitive and semantic tools to enable ordinary users to do creation. Following the design philosophy to be easy-to-use and to achieve a perceived usefulness, current research is making great efforts on semantic based mashups and natural language (NL) processing methods [13]. Semantic based mashups aim to enable automatic service composition, automatic service advertising (according to the current activity and context of the user), and the enhanced service discovery. NL processing methods, in conjunction with the semantic technologies, aim to provide tools that generate services dynamically according to a user NL request.

The second challenge concerns about how to construct a coherent service creation environment (SCE). It requires a certain level of harmonization between the increasing Internet/Telecom/3rd party services. It also needs to be investigated, what's needed to support and help the end-users for an easy and intuitive way to create services not only within the internet, but for all environments, they may use. This trend will lead to a huge set of new features, like context-aware service creation and social service co-creation. Effective incentives like value-added service with revenue sharing could also be used to encourage end-users to do creation. And concerning that, the trust and privacy issues should be taken into account.

In the remainder of this paper we present five future research domains of the end-user driven service creation: semantic based service creation, context-aware service creation, service creation with trust and privacy, social service co-creation and finally, service creation with revenue sharing. Fig. 1 shows the relationships between these topics. The remainder of the paper is organized by these five topics, respectively.

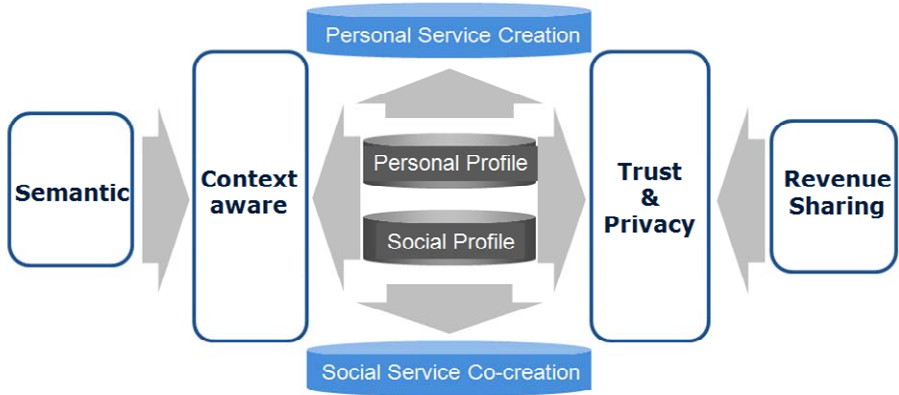


Fig. 1. Future trends of end-user driven service creation

2 Semantic Based Service Creation

Semantics allows end-user to do goal-driven service creation in a technology-agnostic way. Separating technology of a service from its meaning could improve the process of assembling new services. High performance SCE has to provide an extensible and interoperable framework able to integrate semantic field such as semantic service discovery, assembling, and publishing.

From end-users perspective, service creation process must be intuitive and self-explaining: no software development skills are required to develop a new service. End-users cannot understand current service technologies such as WSDL [7] and SOAP [8]. We witness an exponentially growing number of services on the web, which makes the end-user confused about which service to use in his SCE [10-12]. These factors make the intuitiveness almost a key for the success of service creation platform. One proposed solution is to develop a semantic service creation assistant: e.g. whenever the user picks an existing service, the system should be able to suggest a set of syntactically or semantically related services that can be connected to the existing service (Inter-service communication); or user can locate a specific service based on a description of all (or part of) its functionality with a near natural language request (NL composition [13]).

All the above processes require semantic matching, which further requires the semantic annotation of each service. Semantic annotations of the services are semantic tags/vocabularies that are integrated into service functionality. These annotations should define the interfaces of the service and the generated data. However, to enable semantic matching, the service creation platforms should either define a shared semantic vocabulary which will be used by all service providers, or to understand each semantic vocabulary of each service provider and makes translation between each other. The first solution is not dynamic as it forces each service provider to follow the platform specific vocabulary; the second one is not feasible especially for the current target where end-users become parts of service providers. In addition, both approaches suppose the creation of a vocabulary which covers all concepts of all potential services, which is not realistic for a wide range of services.

The second challenge of semantic composition process resides in how to automatically choose (possibly discover) the specific services from the large amount of tagged services. Regarding to this, the platform should be designed taking into account the user preferences and context. Based on the contextual information, business rules or user profile, the platform should intelligently select the appropriate services for a composite service. This could be cross investigated with the context-aware issue.

In the following we will illustrate the features and future trends of NL composition, which was partly investigated by SPICE project [13]. In order to build a customized service directly from a user request, the faced challenges are:

- How to transform the user request: from natural language to formal request?
- What are the required services that respond to that request? (Service discovery)
- What is the execution sequence of these services that responds to the user request?
- And finally, does the created composite service match the user request?

By introducing Composition Factory component, SPICE composition environment summarizes perfectly the listed challenges.

Composition Factory receives a user request as an input in a formal language in which end-users can easily retrieve the requested services. This component is in charge of which composite service should be provided. First, it retrieves request-related services from the service repository and their non-functional properties from the non-functional properties component. Non-functional properties include information like efficiency, reliability and security that do not define the functions of the system but still remain important parameters in service usage. Using the user goals extracted from the request, the component builds the causal link matrix (CLM) [14] or the extended CLM (CLM+ [15]). CLM is a matrix that represents all the matching possibilities between inputs and outputs of the services, and the extended one takes into account the non-functional properties. The matching between two parameters is based on their logical relations and is statically quantified according to an ontology database, e.g. web ontology language (OWL) [16]. The quantification of semantic matching between parameters enables the quantification of the whole composite service quality. Therefore, it allows end-users and developers to select the best composite service between them.

NL composition is definitely the simplest and most intuitive way to create composite services for the end-user as it is just required to enter a NL request. However, since the web services should follow the same semantic dictionary, the semantic reasoning is a heavy task. Moreover, the created service should be validated by the end-user due to the possible mistakes of NL processing. Therefore, the NL composition should combine different technologies and skills like artificial intelligence and self-learning. The further refinement could be done in the following aspects:

User Friendly GUI Design: What graphical user interface (GUI) should be designed to help end-users to formulate their requests?

Automatic Service Composition: The platform should be capable of initializing a service composition definition within the service creation tool. Moreover, the run time service composition should automatically create the user interface from the description of a requested service.

Support of Telecom Services: Telecom services such as messaging, and localization need to be integrated in the NL composition platform, so that the end-user can formulate services using these functionalities such as “Send information to me by SMS according to my current position”.

Support of Multi-Language (Linguistic): The platform should support of multiple languages. Besides, adding voice can be an alternative to the NL request writing.

Support of Self-Learning: Disambiguation through recommendations and self-learning should be investigated in the NL composition platform to enhance robustness.

3 Context-Aware Service Creation

Time of the old-fashioned "same static content for all users" paradigm is gone. As services are becoming more intelligent and advanced, end-users further expect the services to be discovered, created and delivered dependent on some basic information like who they are, where they are, what they have done, what they are doing and when they are doing it. This falls into the initial concern of the context-aware.

The question of how to define the context varies from different research works [17][18]. In this paper, we define context as the changing state of the user, its environment, and the interaction between them. Examples are time, place (location), user's terminal, belonging network, history of interaction and current presence. Besides, we separate the static profile from the definition of the context, examples are user identity/role and user preference.

The main features of the context-aware service creation are:

Services Selection for Composition: Intelligently choose the appropriate services adapted to the user identity/preferences/location/previous interaction.

Adaptive Interface Creation: Adapt the user interface to their current device (PC, mobile phone, PDA, TV...), browser and usage pattern.

Adaptive Notifications: Notify end-users via different channels depending on users' presence status and time preference.

Automatic Recommendations: Take into account the dynamic context information as well as the static profile of the end-user. Trace end-users' behaviors and help them to organize and filter information in order to provide personalized service/product.

Context awareness starts from capturing the users' dynamic context information. This requires the use of heterogeneous sensors distributed throughout the entire space. However, due to the fact that sensor nodes are not always available, the data aggregation mapping can change dynamically over time. Concerning that, a flexible and resilient architecture for sensor data aggregation in a heterogeneous (and dynamic) environment should be investigated. Thereafter, the context information should be interpreted and thus extract the decision to adapt or generate a service for the current situation of the user. This requires a common semantic, and decision rules between the different entities (sensors, decision maker, and the services).

The relationships and features of the semantic and context-aware service creation are summarized in Fig. 2.

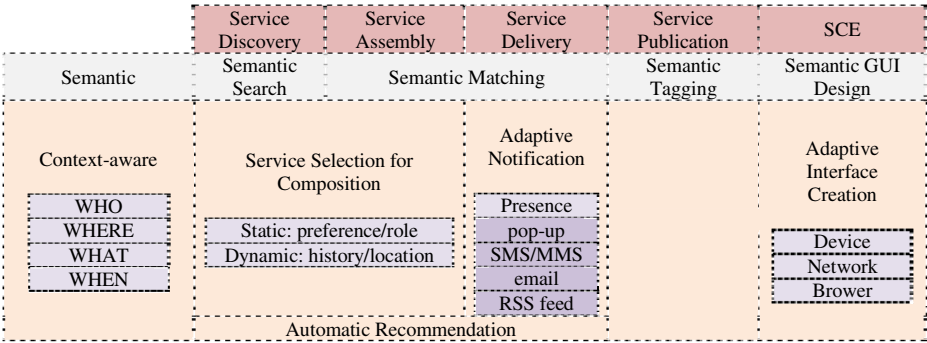


Fig. 2. Features of semantic and context-aware service creation

4 Service Creation with Trust and Privacy

Future service architecture has to proof their trust-worthiness and security awareness. Since this trend is more openness, the future services are required to be created and personalized in a privacy-sensitive manner.

Privacy protection is a cross-cutting issue since the related functionalities appear everywhere (e.g. personal data processing, shared content delivery, financial transactions). The challenges mainly concern the ability for end-users to control, maintain and implement privacy policies to protect their sensitive data in a non-intrusive manner. Therefore, an easy-to-use and intuitive user interface for manipulating their privacy policies should be considered and investigated.

Here we address the trust & privacy management issue from the following aspects:

Identity Management

The identity management is responsible for the management of the user and service attributes (logical and physical) related to the identity.

Service Identification: Any service application should be identified and authenticated. This is essential for service collaboration security.

User Authentication: One identity key should be defined as the end-user’s ultimate control tool over their identities. The identity key may support single sign-on (SSO) to all the services.

Profile Management

The profile management implies the user profile to be stored in a common representation format, and to be shared in a secure manner.

Personal Data Privacy Control: Allow end-users to have full control of their personal data. Their private data will be protected and not revealed to 3rd parties (only the virtual profile will be exposed to the public).

Content Management

Personal Content Privacy Control: Allow end-users to have the full control (includes removal) of their personal contents (media files, blog items, etc).

Shared Content Access Control: Verify the identity for accessing the shared content.

Privacy Preferences of Shared Content: Allow end-users to set different privacy preferences with respect to sharing their content.

Session Management

Session Access Control: Apply proper authorization policy (such as access, modify) to the identity who requests the access to the ongoing sessions.

Session Transfer Control: Allow ongoing sessions interacting/exchanging between different authorized devices.

Advertisement Management

Search and Receive Advert: End-users can search and receive relevant advert. The advert would take into account privacy policy settings specified in the user profile.

Retrieve and Send Advert: User profile helps advertiser to target the relevant end-users. The advert is send to end-users based on their privacy and profile settings.

Service Publication Management

Service publishing control: Allows end-users to publish the created service with a set of sharing and business policies that keep the service private for their own. Users are also allowed to change the sharing policies to make the service publicly available afterwards.

5 Social Service Co-creation

The above sections mainly discuss the “Do-it-yourself” personal service creation. In this section, we will focus on “Do-it-together” social service co-creation.

One main feature of Web 2.0 is collective intelligence. The customer relations will be the most important factors in the future networks for their survival and growth. Therefore, giving end-users the opportunity to create services together is the essential step to harness the social interaction and strengthen their relationships. The features and challenges of the future social community are:

The future social community should harmonize service creation approaches. Looking into the different vendor/project specific approaches it becomes clear, that no common approaches, even not the same kind of definitions or enabling frameworks are used. Therefore interoperability or even cross-technology service creation is not possible. In order to make service creation tools more widely useable though social communities, a certain level of harmonization between different tools needs to be achieved. This can start by merging the same kind of enabling frameworks (like OMA enablers [3] and Parlay (X) API [19]). This kind of integration can lead up to the reuse of common semantic definitions.

The future social community should provide different levels of service creation tools. Community consists of people with various levels of technical background. The service creation tools should therefore be able to operate on different technical levels. The technical minded people will be a valuable resource in developing solutions beneficial for all members of the community. Proper tools for these advanced users will be more demanding, and allowing for more complex functionality.

The future social community should be expanding to social physical environments. Physical environment like hotels, clinics, sports bar and travel agency can be easily adapted to the physical desires of people in everyday life. On the other hand, web based social communities are well satisfied people through expressing and sharing their emotions and thoughts. Since the current social networks are proved to be a big success of strengthen user relationships, this relationship should be more exploited to meet users' necessary requirements by linking the physical environment to the social networks.

Create a Community – Contact and Social Profile Management

Our definition of community is to enable end-users to easily gather a set of group members based on their shared attributes (e.g. common interests, families, co-workers, buddies, etc). The members of community could be from Telco as well as Web world.

Contact List Control: Allow end-users to be able to manage several contact lists, e.g. create a community by choosing their friends and family members from different contact lists.

Identity Lookup: Search end-users based on their personal attributes, e.g. look up the user community ID based on their phone number.

Unified Social Profile: Provide a unified social profile shared by different networks. Moreover, end-users should be allowed to set different privacy levels to different data of his profile.

Profile Synchronization: Allow end-users to define and update their profiles in a timely manner. When a user modifies his personal profiles, the registered services and related contacts should be notified and changed accordingly.

Interaction with the Community

Social-aware Service Creation: Allow end-users to search and retrieve relevant information that takes into consideration the attributes of all community members (such as interest, presence and location). This will contribute in a created service that best suited to all the members of the community.

Autonomic Recommendation of Social Communities: Select social community for the end-users according to their static profile and dynamic context.

Interaction with Current Social Networks: Allow end-users to interact and share data with contacts belong to the well-known social networking sites.

An example of the interaction with the social community is though intelligent calendar widgets. Since our daily lives are organized by calendar events, it's very beneficial if the participants of a calendar event can be handled as communities. Personal

content could be linkable to the calendar events. Through this linking, they automatically get a context associated. All possible community-specific operations such as giving access rights to content, distributing notifications on new content will be handled automatically.

Fig. 3 further illustrates the future features and relationships between the four research topics discussed above.

		Trust & Privacy	Social Service
		Identity	Identity Lookup
		Profile	Unified Social Profile
		Content	Profile Synchronization
		Session	Personal/Shared Content
		Advertisement	Privacy Control
		Created Service	Advertisement Management
Semantic	Semantic Search	Service Identification and Collaboration	Semantic Search
	Semantic Matching		Semantic Matching
	Semantic Tagging		Semantic Tagging
Context-aware	Service Selection for Composition	Personal Data Privacy Control	Social-aware Service Creation
	Automatic Recommendation	Advertisement Management	Automatic Recommendation of Social Communication

Fig. 3. Features and relationships between service creation topics

6 Service Creation with Revenue Sharing

The fifth topic is much more related to the business concern. Indeed, enabling end-users to create their own services promotes software innovation and reduces the time to market of new services. It also raises additional concern such as business models. The open business related issues are essentially focused on charging (billing), revenue sharing and QoS guarantee.

Generating revenue for the participants in a fair share could be useful to encourage end-users to do creation. However, in such multi-vendor supply network, revenue sharing model and reseller model should be defined to implement a fair division of investments, business risks, costs and revenue sharing among the different actors. Apart from this, the charging mechanisms will be revolutionary modified since the end-user (becomes one of the stakeholders) profit and risk should also be dealt with. Note that the advertisement revenue is another source of revenue, so that end-users may share the revenue through giving an opportunity of inserting an advert.

More detailed questions related to revenue sharing and charging are:

- What is the relation between the generated service creator and the composed services providers?
- How to bill end-users using a user generated service? Should service providers calculate the price or the service creator define the price by himself? In the second case, who is going to pay for the usage of the basic existing services?
- Are the used basic services billed at the creation time or at the execution time?
- How to bill a user generated service especially in the case where the used basic services have different payment method (prepaid, postpaid)?

The second raised question is how to guarantee and maintain the QoS of the user generated services. Concerning this issue, Service level agreements (SLAs) should be established between end-users, service providers and 3rd parties. Moreover the QoS features should be utilized to ensure the SLA as agreed. The exploitation of the QoS and SLA in characterization of the services will be investigated to ensure the fulfillment of the agreements in a service creation environment. In addition, the trust and privacy issues should be taken into account.

7 Conclusion

Web 2.0 is envisioned as an open garden for services. Enable end-users to do service creation can reduce the time to market of new services, harness the collective intelligence and promote innovation in software development. In this paper, we have presented an overall view of future end-user driven service creation, from both technical and business perspectives. Future trends with extensive new features have been illustrated as well as the current platforms/tools. It has been shown that the future service creation tools must be intuitive and self-explaining, and the services should be created in a context-aware way. It also becomes clear that the “Do it together” social service co-creation with revenue sharing is the next step that will happen. Moreover, in order to achieve a pervasive usefulness, a certain level of harmonization among different service creation approaches needs to be achieved.

References

1. O'Reilly, T.: What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies* (1), 17 (2007)
2. Braga, D., Ceri, S., Martinenghi, D., Daniel, F.: Mashing Up Search Services. In: IEEE Computer Society, *Internet Computing*, pp. 16–23. IEEE Press, Los Alamitos (2008)
3. Open Mobile Alliance, <http://www.openmobilealliance.org/>
4. Newcomer, E.: *Understanding Web Services: XML, WSDL, SOAP, and UDDI*. Addison-Wesley Longman Publishing, Wesley (2002)
5. Jin, Y., Boualem, B., Casati, F., Daniel, F.: Understanding Mashup Development. In: IEEE Computer Society, *Internet Computing*, pp. 44–52. IEEE Press, Los Alamitos (2008)
6. Fielding, R.T.: *Architectural Styles and the Design of Network-based Software Architectures*. Thesis dissertation, University of California, Irvine (2000)
7. Web Services Description Language, <http://www.w3.org/TR/wsdl>
8. Simple Object Access Protocol, <http://www.w3.org/TR/soap>
9. Chow, S.W.: *PHP Web 2.0 Mashup Projects*. PACKT Publishing (2007)
10. Yahoo Pipes, <http://pipes.yahoo.com/pipes/>
11. Wong, J., Hong, J.I.: Making Mashups with Marmite: Towards end-user programming for the web. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, pp. 1435–1444 (2007)
12. Microsoft Popfly, <http://www.popfly.com/>
13. SPICE project, <http://www.ist-spice.org/>

14. Lécué, F., Léger, A.: Semantic Web Service Composition Based on a Closed World Assumption. In: 4th European Conference on Web Services (ECOWS 2006), pp. 233–242 (2006)
15. Lécué, F., Silva, E., Pires, L.F.: A Framework for Dynamic Web Services Composition. In: 2nd ECOWS Workshop on Emerging Web Services Technology (WEWST 2007), Halle, Germany (2007)
16. Web Ontology Language, <http://www.w3.org/TR/owl-features/>
17. Lieberman, H., Selker, T.: Out of Context: Computer Systems that Adapt to, and Learn From, Context. *IBM Systems Journal*, 617–632 (2000)
18. Dey, A.K.: Understanding and Using Context. *Personal and Ubiquitous Computing* 5(1), 4–7 (2001)
19. Parlay, X.: <http://www.parlay.org/en/specifications/>

The New Way to Think Television on the Web

Riccardo Mangiaracina, Giovanni Toletti, and Luca Turba

Politecnico di Milano, Department of Management, Economics and Industrial Engineering
Piazza L. da Vinci 32 - 20133 Milano, Italy

{riccardo.mangiaracina,giovanni.toletti,luca.turba}@polimi.it

Abstract. Television is attracting an enormous amount of attention from both researchers and managers, due to profound changes that are taking place thanks to the diffusion of digital technology. The Internet and the Web, affecting the overall landscape, the consumer habits, the strategies of the players, are playing a key role in the TV revolution. This paper, based on 26 case studies, a survey involving over 200 players and a complete census of all the channels provided via Web in Italy, aims at defining the current state of the Web TV (we call it Desktop-TV) in Italy and identifying the main business models and the strategies adopted by the players, thus allowing us to give a few predictions on the possible future changes in the industry. Although the analyses have been carried out in Italy, the results will have a general applicability, since the Web is not country specific by definition and the majority of players included in this research have international relevance.

Keywords: Digital TV, Desktop-TV, Web, Strategy, Business model.

1 Introduction

The television industry, for decades quite static, is currently undergoing important changes [1]: such business is in fact being shaped by a number of driving forces, such as the digitalization of the TV signal, the diffusion of new alternative access technologies, the development of broadband and streaming video technologies, the introduction of Web TV and new generation access terminals, and the progress of interactive and personal television solutions [2-4]. Past changes, as well as those to come, are so radical that some authors prefer to talk about TV (r)evolution rather than evolution, asserting that the very meaning of the term “television” needs to be revised [5].

The TV business is no longer restricted to few actors, but is becoming territory to be conquered by new entrants from industries like Internet and telecommunications. The market offerings are reaching new access terminals (such as PC and handheld devices), meeting and integrating with the Internet and with other online multimedia services [6,7], and changing by creating new television formats, such as the personal and the interactive television [8].

The technological evolution has shifted the boundaries of the TV industry. Television platforms are in competition, not only among themselves, but also with “New Media” offering contents and services not specifically related to the television sector [9,10]: i) Internet and ii) all the new ways of offline use of digital contents (e.g. podcasting,

downloading of entertainment contents on the PC or mobile phone). There is strong competition between traditional media and new media for the *share of time* of users and the *share of advertising* of investors.

Six digital platforms can be identified on digital networks: Sat TV, DTT, IpTV, Web TV, Mobile TV on DVB-H network and Mobile TV on cellular network. However, given the differences and similarities between the six platforms, a further clustering that would achieve a subdivision into only three TV macro categories, could be a possible scenario [11]:

1. *Sofa-TV*, including all digital television typically viewed through the “traditional” television screen. The “new” Sofa-TVs are based on three digital platforms: Sat TV, DTT and IpTV. The use of the expression, “Sofa-TV” aims to clearly describe the viewing opportunities and modalities of these televisions (standard Tv set watched from the Sofa).
2. *Desktop-TV*, including all the video channels that are viewable through the Web (and Internet in general). In this case, the distinguishing element is the proactive viewing (“elbows on the desk”) of the contents.
3. *Hand-TV*, including all the TV and video offerings available on the Mobile platform, based both on DVB-H networks and cellular networks. The use of the expression “Hand-TV” aims to focus on the concept of TV viewable in the palm of the hand, which frees this type of television from the underlying technologies, both at the network level (DVB-H, cellular networks, and – in the future – WiFi and its evolutionary products), as well as at the terminal level (not only cell phones, but possibly other small devices, like portable music readers, Mobile game consoles, etc.).

Despite its market value is not clear yet, the Desktop-TV has the potential to change consumer habits and, generally speaking, the way the users watch TV. This process of change has already begun. For this reason, this paper will focus on the area of Desktop TVs. The paper is divided into four sections. Section 2 will present the empirical study on which our considerations are based. Section 3 will describe the offerings of the Desktop-TV sector. Section 4 will deal with players of the Desktop-TV industry and their strategies. Finally, section 5 will offer the main trends in the Desktop-TV fields. The analysis will have a general applicability, as the considerations made are not particularly country-specific, although performed within the Italian context, one of the most advanced in the development of digital television platforms.

2 The Methodology

The information needed for the analysis and evaluation of the Desktop-TV sector in Italy was mainly collected through a case study methodology and a survey involving 200 operators. An exhaustive census of the channels provided via the Web was also taken in order to complete the information base.

The empirical study has been conducted on a sample of 26 companies operating at different stages of the new digital television value chain and possessing different characteristics (e.g. size, revenue, business model).

The panel comprises:

- 12 cases from Desktop-TV operators (i.e. C6TV - www.c6TV.it, Elemedia - www.repubblica.it, Gruppo Mediaset - www.mediasetrivideo.it, Google - www.youtube.it, Gruppo Telecom Italia - www.yalp.com, Infostrada - www.libero.it, Microsoft - www.MSN.com, Rai - www.rai.tv, RCS - www.corriere.it and www.gazzetta.it, SKY Italia - www.sky.it, Streamit - www.streamit.tv, Tiscali - www.tiscali.it);
- 6 cases from content providers (i.e. Digicast, Einstein Multimedia, Endemol, Turner Broadcasting System Italia, Walt Disney, Yam112003);
- 3 cases from service providers (i.e. Glomera, Ksoft, Sharemedia);
- 5 cases from media/advertising centres (i.e. Carat/Isobar, Digitalia '08, Media-Com Italia, Niumidia Adv, Sipra);

To help interviewers and ensure comparability, an investigative framework was created to cover the most important issues to be investigated in order to achieve the research results. In particular, four main areas of investigation were identified and analyzed during the interviews:

- Overall company strategy (e.g. television platforms portfolio, launch of new platforms);
- Value chain analysis (e.g. make or buy decision, partnership relations);
- Organizational structure (e.g. number, roles and structure of the people involved in the Desktop-TV organizational functions);
- Television offerings (e.g. channel offerings on the specific platform, differences of own vs. competitor strategy).

In order to extend the empirical base we carried out a survey that involved over 200 operators. The questionnaire used for the survey was focused on the same key points investigated through the interviews.

The information obtained via the case studies and the survey was integrated through an exhaustive census of the Desktop-TV offerings aimed at mapping all the channels transmitted in Italy today. More than 1.000 channels were individuated and analyzed both in 2007 and 2008, and for each, 23 important variables were investigated.

The interviews and the survey were carried out between April 2007 and October 2008. The census was first taken in October 2007 and then in October 2008, in order to highlight the changes. All the research on the New TV sector began in January 2007 and is still in progress.

3 The Desktop-TV Offerings in Italy

In this section, we will analyze in detail the television channel offerings in Italy relative to the Desktop-TV. A census was performed on over 1 000 channels distributed by over 500 different operators. Compared to 2007, in which there were over 800 channels and 435 operators, for channels and operators we see a growth of 25% and 17% respectively.

Filtering the offer on the basis of “quality” and eliminating all channels that do not go over a minimum level of “dignity,” the number of channels drops to 633. The

“dignity” of a channel has been assessed according to a multi-criteria approach based on different parameters depending on the specific type of channel (content range and update frequency for the on demand channels, variety of the programs for the linear programmed channels). From this point on, all our evaluations will be based on this rectified universe.

First and foremost, it is interesting to evaluate the turnover both of channels as well as operators which have come about in the 2008 year. There were 57 channels present in 2007 that are not longer active in 2008 and a whopping 214 new channels born this year alone. The overall effect is an increase of over 150 channels for a percent growth of 33. The med-high channels of “dignity” grow therefore at a higher rate compared to the one registered by all channels (equal to, as already down, 25%). There are 40 operators who are no longer active in 2008 and 81 new entrants to the market of Web TV with a net growth of little less than 15%.

Analysing the different typologies of channels, we see an increase in On Demand channels that, growing at a higher rate than linear programmed channels, surpass 70% of the overall offer.

4 The Business Models and the Strategies

This section wants to shed some light on two of the most important issues in the Desktop-TV area:

- the business models adopted by the operators;
- the main strategies (that depend of course on the business models).

4.1 The Business Models

In order to have an accurate understanding of the strategic scenario of Desktop-TVs in Italy, it is useful to provide an overview of the business models, which emerged from Research and which we have been classified in four main categories.

B2c advertising-based model. This is the most diffused model among Desktop-TVs since the large part of the operators (i.e. more than 80%) adopted this model. The alternatives to insert or associate advertising to Video/TV contents are different. The most traditional ways are advertising outside video and advertising included in video. In the first case, the most diffused format is banners, whereas in the second, we see both spot in pre-roll (both for On demand channels and some linear, even today with the click to play format that leads the user directly to the site of the “advertiser”) as well as included in the stream (for linear programmed channels). In the choice between advertising included in the stream and advertising outside of video, one must first evaluate wisely the trade-off between effectiveness and invasiveness (a spot in pre-roll attracts more attention compared to a banner alongside content, but is undoubtedly more invasive for the user who could potential quit watching). Amongst the most innovative models, we point out overlay, a good compromise in regards to trade-off between effectiveness and invasiveness, even if up until today, it is still little diffused, due to the technological complexity (not all the platforms are able to distribute it) and at times even for the limited creativity. Alongside these different options

whose purpose is to associate commercials to video contents, “ad-funded” contents are pervading the Web and represent a new concept in advertising: based on the production of video with “value” (and not mere spots) for the user (as in entertainment, instruction, etc.) that promote the product/service or company only indirectly and that are diffused on the Web, either using the virality of the contents or through thematic channels.

B2c premium model. These are those models which envisage either the payment of a subscription allowing access to contents/programming for a specific period of time, or a fee for viewing a single content (pay per view). The results obtained by the few players who have decided to play in this game are not clear yet. It seems to us that the model finds its most natural area of application for contents that tend to more of a niche and above all exclusive (that is not available on other platforms). The diffusion of this model is still very limited with just a minority (i.e. less than 5%) of the operators that have adopted it (e.g. Telecom with YALP, RCS Group with Gazzetta.it, ShowTime which is the Tv of the AC Milan, Mediaset with Mediaset Rivideo).

CRM-based model. For operators following this business model, the distribution of Video/TV channels fits into a rather precise strategy of Marketing and CRM, geared to beef up the communication and interaction with their own clients/users. The awareness of the noteworthy capabilities offered by such a rich media as the television is spreading even among companies and public entities, as is demonstrated by the numerous cases identified this year by Desktop-TVs included in this category. Measuring the returns of this type of initiative is surely complicated, but certain drivers (like the viewed contents, the number of users, the average time per user on each site) demonstrate the effectiveness of many projects. Since this model have been adopted by the Public Administration and the companies who decided to have a Desktop TV, its diffusion is quite limited (only 12% of the operators belongs to the category of “P.A. and companies”).

B2b model. In an increasingly greater number, are the companies - and often pure players (but still less than 10% of the players) - who abandon the B2c business model in order to become service, solution and content providers, and at times even for other players, like Web publishers, traditional publishers, companies, and public entities, etc.

It should be made clear, however, that in many cases, the model adopted by these single initiatives is the result of a combination of the different business models described above. It is not rare to find “hybrid” models in which, for example, advertising based models flank premium models, or B2c models flank those B2b.

4.2 The Strategies

We can schematically read the different “strategic” clusters that characterize the Italian Desktop-TV according to three different axis as shown in figure 1:

- category of players (Traditional publishers, i.e. publishers already present on traditional channels that can be split into press and Tv publishers; Web publishers, i.e. publishers already present on the web that have developed specific competencies and abilities on the online channel; Pure Players, i.e. companies focused on the

Desktop Tv only; Public entities and companies, i.e. players that are not included in the publishing sector);

- types of channels (Transposed linear, i.e. channel moved from the traditional Tv with no or minor changes; Native linear, i.e. linear channel conceived for the Web; On demand published, i.e. on demand channels with contents provided by publishers; On demand UGC, i.e. on demand channels with contents provided by the users);
- business models (B2c advertising-based model, B2c premium model, CRM-based model, B2b model, as described in the above paragraph).

The driving factor in the cluster identification is of course the category of players, since each kind of player has its own background, competencies and objectives with situations that can be really very different from each others. We can identify five main clusters:

- The cluster of Web Publishers, that offer mainly On demand editorial channels, and less frequently, the other types of channels. The strategy is mainly oriented toward increasing the overall portfolio of the offer, intended as the sum of the contents, services and features (for this we have called the cluster, *Portfolio Strategies*). The Web Publishers are proceeding with an increasingly greater integration of video within their portals, thanks as well to the “embedded” feature able to easily position videos within different sections of the site. It is therefore increasingly more complicated, and maybe even little significant, to isolate the video offer from the overall offer. In this cluster, business models based on advertising, even if some Web Publishers with premium channels also play a part, clearly prevail.
- The cluster of Traditional TV Publishers, that mainly offer transposed linear programmed channels and On demand editorial channels as a complement to the offer on other platforms (we refer to this cluster as *Multichannel strategies*). The presence of native linear programmed channels along with the increase in On demand editorial channels is an indication of a growing attention towards the specificity of Desktop-TV and a greater awareness of the capabilities of the Web for this category of Publishers. Different business models adopted: certain players (like for example Rai and Sky), have leaned on advertising (both in pre-roll as well as banner), whereas others (like for example Mediaset with Rivideo initiatives) have chosen to supply even pay contents.
- The cluster of traditional Publishers (non television), for whom the Video and TV offer is an enabling factor both the presence on the Web channel as well as the extension of the portfolio of offer (*Multichannel and portfolio strategy*). Among the most interesting cases in this cluster are the initiatives of the big names in newspapers (like RCS and “la Repubblica”) and certain important national radio stations (like Radio Dee Jay) The adopted business models are mainly advertising based and are only in a few isolated cases, premium.
- The Pure Player cluster, conceived to operate in this specific sphere, offers On demand channels, both editorial as well as UGC, and native linear channels created ad hoc for the Web (we have called this the *Pure Strategies* cluster). The initiatives present in this cluster are very heterogeneous. From Online video

companies (like YouTube) and editorial initiatives supported by big investors (like Current), to start-ups that, although small in some cases, are developing extremely interesting entrepreneurial projects for adopted technological solutions (StreamIt, Sharemedia, etc.), programming and ways of producing content (C6.TV, N3TV). The majority of these operators have started their activity with advertising-based business models, quickly discovering however, the noteworthy difficulties in making it financially profitable, based on advertising alone. Exactly for these reasons, many companies are gearing their strategies towards B2b models as well.

- The public entities and companies cluster, that having as its main objective the activation and/or consolidation of the client relationship and the affirmation of the brand, has been therefore named *CRM Strategies*. This is the cluster which in 2008 marked the greatest growth (in 2007 it was in fact rather insignificant). Particularly enterprising were the companies (automotive) and soccer teams (both A and B series) who in 2008 launched channels with business models and supply means that were different even amongst themselves (from On demand channels in pay-per-view to free linear programmed channels).

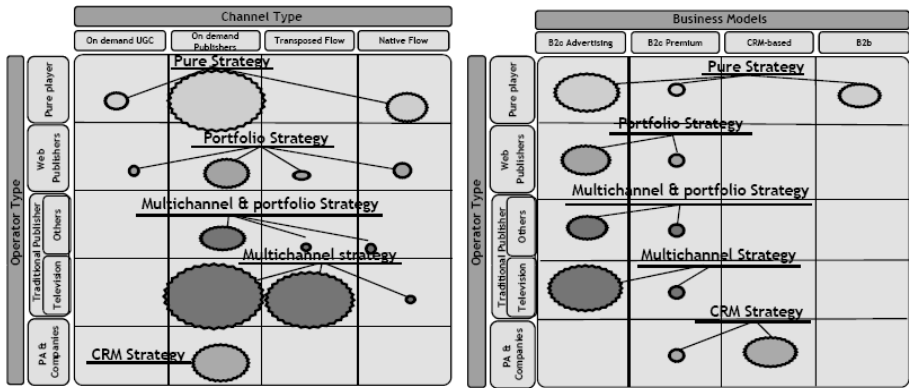


Fig. 1. The main strategic clusters

The diffusion of the different clusters and the trend observed in 2008 with respect to 2007 can be summarized in figure 2. We see these principal findings:

- The weight of the Traditional Tv publishers drops by five percentage points, from 44 to 39%, mainly due to the exit from the scene of certain Tv stations present in 2007 and due to the increase in other categories of operators like Web publishers and companies (the latter practically absent last year).
- The percentage weight of Public Administrations and companies has jumped from 7 to 12%, mostly thanks to the companies that make increasingly greater use of Desktop-TV as an instrument of communication and interaction with their own users/client.

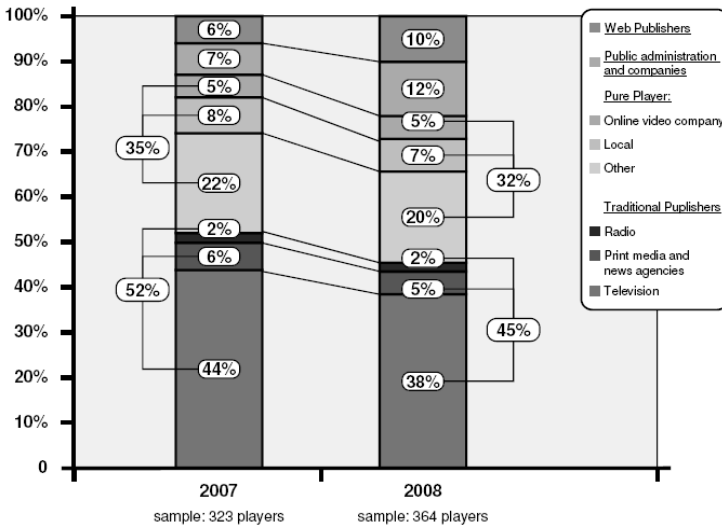


Fig. 2. The distribution of operators by typology

- An increase from 6 to 10% in the percentage weight of Web publishers who consider the Video and TV offer increasingly important and a complement to their online offer.
- The other categories of operators maintain a more or less unaltered percentage weight, with a growth, from the numerical point of view, in line with the number of overall operators.

5 Conclusions: Current Trends and Future Scenarios

There are two principal trends on which we would like to focus our attention.

The first concerns the concept of Desktop-TV itself and the possibility to consider it as a “standalone” platform: the current phenomena are clearly demonstrating that Web TV is not to be considered a specific market, able to be isolated from the overall online world, but rather as part of the offer - particularly rich and attractive for the user - of the overall, composite world of the Web. This translates into less space for “pure” operators, those start-ups conceived specifically to operate in this market (which, in fact, as already mentioned, are having not few difficulties making it financially worthwhile, and exactly for this reason, are often gearing more towards B2b offers) and more space, on the other hand, for those publishers, - both traditional as well as Web - who will be able to interpret the Video/TV offer in an overall strategy that in its totality is more multichannel and with a more expanded portfolio of offerings. There is more space even for those companies and public entities that want to put their bets on Video to build up the communication and interaction with their own users.

The second trend concerns the noteworthy mass of experimental activity currently taking place in this area: the Web is proving to be a particularly favourable setting in which to conduct experiments in the world of Video and TV, in order to

tweak and test innovations at the level of ways of structuring and defining Video contents (format/scheduling), but also of production and programming of contents and supply of contents and management of user interface. More in detail we have identified three dimensions of possible innovation:

- *Ways of structuring and articulating Video contents (format/programming).* The Web allows an infinite number of possibilities to bring together, structure, articulate and compare contents. The concept of scheduling is radically changed by all these possibilities and is replaced by the most generic concept of format.
- *Ways of producing and programming contents.* On the Web, the production of contents can be entrusted, besides to the “usual” professional publishers, to other categories of actors (from the users themselves, who according to the Web 2.0 rationale, can directly contribute to programming to semi-professional authors, e.g. theatre schools, cinema schools, etc.). Included as well in this category of innovations is the timeliness made possible by Internet in broadcasting contents, intended both as speed in gathering and distributing information.
- *Ways of using contents and managing user interface.* The Web (and Internet in general) allows the user to take advantage of contents that the broadcaster makes available to him in innovative ways. There is the possibility of using linear programming even in an On-demand fashion (the user “moves” in the programming to select contents that he wants to see) or the possibility of interacting, leaving feedback/scores and comments, starting discussions, participating firsthand in the transmissions, supplying content, etc.

We like, therefore to see the entrepreneurial revolution and innovative vivacity that are in this moment characterizing the Video offer on the Web, as an actual “laboratory” in which new TV concepts are being tested which will be able in time not only to enrich the offer at the Desk-top TV level, but also the offer of other television platforms, Sofa-TV and Hand-TV.

We in fact think that from this privileged “laboratory” many important implications can be derived, even for a rethinking and a major innovation of traditional television formats that up until today have been dominating Sofa-TVs and also Hand-TVs markets.

These phenomena will also be pushed by an increasingly greater transposition of Web onto the big screen of the Sofa-TVs as well as the little screen of the Hand-TVs, a transposition “technically” made possible by the respective diffusion of universal set top boxes for Sofa-TV, and of cell phones, which allow an effective Mobile Internet for Hand-TVs. For instance Current TV and Qoob TV, both born on the Web, have just been transposed to the Sofa-TV (Sat TV and DTT).

References

1. Meisel, J.: The emergence of the internet to deliver video programming: economic and regulatory issues. *Info*. 9(1), 52–64 (2007)
2. Bria, A., Kärberg, P., Andersson, P.: TV in the mobile or TV for the mobile: challenges and changing value chains. In: *The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Athens, Greece (2007)

3. Hu, A.W.: An empirical test of a use-diffusion model for Taiwan mobile digital TV. *International Journal of Technology Management* 39(3/4), 248–263 (2007)
4. Waldfogel, J.: Lost' on the Web: Does Web Distribution Stimulate or Depress Television Viewing? NBER Working Papers. National Bureau of Economic Research, Cambridge (2007)
5. Rangone, A., Turconi, A.: The television (r)evolution within the multimedia convergence: a strategic reference framework. *Management Decision* 41(1), 48–71 (2003)
6. Ha, L.: Enhanced television strategy models: a study of TV Web sites. *Internet Research: Electronic networking Applications and Policy* 12(3), 235–247 (2002)
7. Weisz, J.D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R.E., Konstan, J.A.: Watching Together: Integrating Text Chat with Video. In: CHI 2007, San Jose, California, USA (2007)
8. Van Asschea, S., Hendrickxb, F., Oortsa, N., Nachtergaele, L.: Multi-channel publishing of interactive multimedia presentations. *Computers & Graphics* 28, 193–206 (2003)
9. Poel, M., Renda, A., Ballon, P.: Business model analysis as a new tool for policy evaluation: policies for digital content. *Info* 9(5), 86–100 (2007)
10. Furregoni, E., Rangone, A., Renga, F.M., Valsecchi, M.: The mobile digital contents distribution scenario. In: Sixth International Conference on the Management of Mobile Business, Toronto, Canada (2007)
11. Toletti, G., Turba, L.: Digital Televisions: the new Sofa-TV...and not only. In: International Conference on Business and Economy, Constanta, Romania (2008)

The Wisdom of Sustainable Communities in the Digital Era: The Case of Efficient Energy Management

Alessandro Tavoni¹ and Luigi Telesca²

¹ Advanced School of Economics at the University of Venice, Cannaregio 873,
30121 Venice, Italy

² CREATE-NET, Via alla Cascata 56/D, 38100 Trento, Italy
atavoni@princeton.edu,
luigi.telesca@create-net.org

Abstract. Wise management of natural resources has become a fundamental challenge nowadays. Cooperative approaches are needed in order to provide a sustainable framework for stimulating the individual awareness and participation towards common environmental objectives. In order to reach mutually beneficial outcomes, public administrations, companies, individuals and communities will need to organize for managing their aggregated consumption, reducing unnecessary use of scarce resources. This is especially true in the field of energy management where the bounded nature of the resources and the behavioural impact of the stakeholders on the overall system both play a fundamental role. The proposed research aims at contributing to what remains a challenging issue across disciplines: explaining the evolution of cooperation among unrelated individuals in human societies. We utilize an evolutionary game theoretic modelling framework to identify conditions under which collaboration among energy end-users mediated by information and communication technologies (ICT) will support its sustainable exploitation.

Keywords: Energy management, social contagion, bottom-up innovation, community, Web 2.0, smart electricity grids, replicator dynamics, evolutionary game theory.

1 Introduction

Local and global ecosystems are under growing pressure worldwide. Their sustainable management requires not only an understanding of the environmental factors that affect them, but also a knowledge of the interactions and feedback cycles that operate between the resource dynamics and the socio-economic dynamics attributable to human intervention. History teaches us that the livelihood of our species has been inextricably related to our ability to cooperate in the sense of restraining use of ecosystem services to sustainable levels rather than giving in to individual short-sighted resource over-exploitation. Maintaining such cooperation against the myopic self interest of the individual users and despite growing environmental pressure is a challenging task that often depends on a multitude of factors, as both successful and unsuccessful environmental management has shown. As an example, take the objectives of CO₂ emission

reduction set by the European community, which necessitate a dramatic behaviour change at the individual level.¹

Today, household consumers perceive electricity as a commodity which can be consumed at will and in unbounded quantity. In most of the cases, the only feedback they receive takes the form of a monthly or bi-monthly bill. On the bill, many parameters contributing to the global cost are detailed such as actual or estimate consumption, corrections from earlier estimates, fixed costs, and an often complex pricing model which, depending on the country, location or utility, may include considerations such as variable pricing depending on the time or day of the week and the consumption threshold reached over a period. As a result, the effect of individual behaviour is lost. Individual monitoring solutions, if not linked with community profiling, local data conditions and energy production information, will have a limited impact due to lack of visibility and understanding on the user part.

We think that community well-being and social engagement have a strong potential for motivating behavioural change, but have not yet been fully leveraged. Therefore a holistic approach, considering all possible (hardware) elements, but also including end-users, is needed to address the energy challenges of the coming years. To achieve this goal, substantial help may come from a technology that empowers small consumers such as households as well as small and medium enterprises (SME) to actively engage them in the effort, providing the incentives and means for achieving a significant contribution to the reduction of the environmental footprint of human activities and the improvement of energy efficiency. If empowered with the right tools, energy users will strive and collaborate to define their own (community) energy saving measures and consumption models. This bottom up approach has the potential to reach faster concrete results and favour the behavioural changes. The unprecedented innovation pace happening on the Internet and characterised by the "Web 2.0" wave, demonstrates the potential of bottom-up innovation in complex socio-technical systems.

The proposed research aims at contributing to what remains a challenging issue across disciplines: explaining the evolution of cooperation among unrelated individuals in human societies. In the absence of enforcement mechanisms, conventional evolutionary game theory predicts that the temptation to act in one's own interest leads individuals to inefficient use of resources. A growing body of knowledge, however, demonstrates that in many situations there is more to human behaviour than selfish behaviour: experimental results from behavioural economics, evolutionary game theory and neuroscience have inexorably made the case that human choice is a social experience. Recent empirical studies show that social norms significantly influence behaviour, often effectively enhancing the efficiency in resource use.

In our research we endow communities with a multi-layered technological framework that will provide them with an innovative P2P infrastructure linked with smart metering systems allowing an open, distributed monitoring and control plan for local energy grids. These, in turn, enable community-level organization of users and new services and business models. As shown by the internet, "peoples' solutions" have the potential to reach faster and far beyond the limits imposed by political constraints and

¹ The targets for 2020 are 20% reduction in emissions compared to 1990 levels, 20% share of renewable energies in EU energy consumption; and 20% savings in EU energy consumption compared to projections.

slow societal changes. The successful implementation of this approach is conditional on the success of a collaboration model favouring the overall sustainability of the system via the uptake of the implemented solution by a substantial proportion of society members. That is to say: will such a bottom-up radical reform of the energy system develop and survive in a world divided between socially-inclined individuals as well as selfish ones? We investigate this question by taking an evolutionary game theory perspective.

More specifically, we examine the role of prosocial behaviour² as a mechanism for the establishment and maintenance of cooperation in resource use under variable social and environmental conditions. Using evolutionary game theory, we investigate the stability of cooperation in a population of resource users (e.g. energy users) that are concerned about their status with respect to other community members and therefore are able to compensate (in a frequency-dependent fashion, i.e. to a degree related to the frequency of prosumers relative to traditional users) the losses incurred by shifting energy practices to more sustainable ones with benefits that are exclusive to the community of prosumers. That is, individuals of this idealized economy, which for the sake of concreteness may be thought of as a region or a municipality, face a trade-off: on the one hand they can continue with their business as usual, i.e. continuing to consume energy at will without incurring the costs of planning ahead of time and coordinating with other community member, or on the other hand they can shift paradigm and constraint themselves to a socially agreed upon acceptable level. By doing so, their conventional materialistic pay-off is assumed to be below that of the non-cooperating agents, to account for the above mentioned costs of coordination, but such disadvantage may be offset by a term capturing the benefits that accrue to prosumers only by belonging to the sustainable community. Such benefits have a magnitude that depends on the relative size of the community, since, as will be further explained in the next section, to take full advantage of the new paradigm, it is assumed that many participants should be involved; in other words, at low frequencies of prosumer types, the defectors, i.e. those that opt out of the technological and behavioural agreement (to be described below), will be better off, but at high enough frequencies of cooperators (prosumers), the latter will enjoy benefits that will make it advantageous to be part of the community.

To recapitulate, we model two types of agents: cooperators (prosumers) that restrain their energy usage level to a coordinated and socially accepted amount that is closer to the efficient level for the community, and defectors who base their consumption decisions solely on their immediate needs in a own-regarding way. The prosumers' pay-off from resource use is augmented contingently upon both the pay-off difference with respect to the defectors (a larger one signals greater divide between the groups and consequently a greater cohesion among cooperators) and the benefit function, which is increasing in the frequency of cooperators (as the advantage of belonging to the group can be reasonably assumed to be higher the larger the group is in relative terms).

When sufficient levels of social and technological benefits accrue to those enforcing the social norm (innovative and well coordinated emission-reducing energy practices), cooperation obtains, leading to a sustainable resource management. We analyse whether coexistence of both norm-observing and norm-violating types is possible, and

² See Gowdy (2008), Tavoni (2009) and the references contained therein for an exploration of social value orientations and other-regarding preferences.

to what extent it depends on the social and ecological environments. This approach allows us to better explore the interactions between the environmental and the socio-economic dynamics, and their response to endogenous (i.e. behavioural) as well as exogenous (i.e. climatic) pressures exacerbating the resource unpredictability.

This study is meant to contribute to the literature by providing a systematic exploration of the resilience of coupled social-ecological systems to local and global change, by allowing the evaluation of the impact of shocks in energy dynamics on resource use and, more generally, on the capacity of the users to adapt to uncertainty in resource availability. At the behavioural level, this framework is well versed for analysing the role of information and communication technologies (ICT) in changing the energy consumption behaviour of people and local communities towards more sustainable practices. The question we have in mind is: Under which conditions will a community of energy-aware individuals thrive? In order to address it we will present an example of how we picture such a sustainable community in section 2, while section 3 will present the evolutionary game-theoretic model based on replicator dynamics, followed by a section containing conclusive remarks.

2 Enerweb as an Example of Efficient Energy Management

In our research, we aim at leveraging the concept of community to create a sustainable solution for the energy grid. In fact, in order to successfully address global warming, there is a need to devise radical innovations in the manner energy is managed. The European Commission has identified two important conditions and challenges that need to be solved in order to provide systemic results. The first one is to foster energy efficiency and savings by engaging users in participating actively to the reduction of the global carbon footprint. The second is to increase the penetration of renewable energy sources in the market.

We propose an innovative approach (from now on labelled enerweb) that enables the raise of bottom up energy wise communities and the operation of prosumers (users who can also act as micro-producers) and virtual operators that obtain access to the electricity grid and enter the market to trade the power they generate. These new producers and community traders would be working at relatively low voltage and be distributed in several locations, in the vicinity or within urban or suburban areas. Several collaboration and partnership models with the local distributors would arise, with the final result of lowering the entry barriers, thus letting prosumers become part of a global socio-ICT ecosystem, where they can negotiate the energy they produce and consume.

Enerweb is therefore a socio-technical environment engineered as an open, secure, resilient and scalable service oriented infrastructure built on the P2P paradigm for the management of the local smart grid and its customization and extensibility towards multi-utilities management and control. It offers an Open Standards and Open APIs infrastructure that allow all the energy value-chain stakeholders to manage energy-related data and communication. It acts as the information aggregator between all elements of the network for their monitoring and control and will constitute the technology backbone which will enable the enerweb community concept to work properly. Enerweb offers self-optimization mechanisms to help organising large-scale unstructured networks

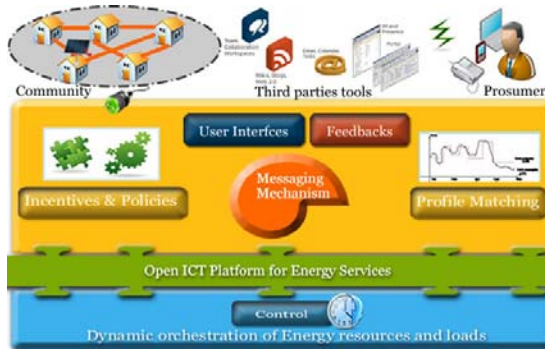


Fig. 1. Sketch of the proposed framework

of small devices into well-defined geographical groups that optimise resource allocation to save time and energy while preserving the network stability in case on changing energy conditions. It also offers an intuitive user interface that will provide users with all the necessary community information aspects and options for actions.

In this framework incentive policies play an important role in shaping user behaviour and ultimately achieving performance gains. We intend to study several possible incentive policies that will help the prosumers organize in communities and more generally promote cooperation. Current incentive strategies are summarized in low cost energy offers for overnight use. An example of more complex incentive strategy would be to offer cheaper prices for electricity provided in low variation levels. This would motivate prosumers to organize in communities in order to coordinate and normalize the collective power profile.

Enerweb would allow reducing the lack of involvement of consumers in the energy saving efforts. We believe that pure monetary considerations are insufficient to induce a behavioural change among consumers. Several propositions have been made to address this problem through, e.g., market mechanisms with real time pricing or using real time monitoring to improve the accuracy of the feedback. However, while such initiatives contribute to making monetary incentive more effective, they do not address the root of the problem, namely, that other forms of incentives are needed. Considering Europe, where the average income is relatively high and electricity represents a small fraction of household expenses (significantly below 10%), monetary incentive will remain insufficient to motivate dramatic behavioural changes until energy prices dramatically increase – in other words, until it will be too late.

In fact, in contrast with modelling efforts highlighting the role of agents' selfishness and sole concern for economic incentives, we consider user motivation in all its social relevance (as captured by the prosumers' benefit function). Users are given the option to self-organize into groups (communities) of various sizes, with broad freedom in determining how far they are willing to commit and reduce green-house gas emission. Instead of relying on selfish interest (e.g., through monetary incentive) or top-down infrastructure development, the enerweb will promote collaboration among users (forming dynamic communities), the design of innovative solution and the fulfilment of greater common objectives. For achieving their objectives, these communities express concrete and measurable targets regarding, in particular, their joint electrical consumption.

A mix of selfish and altruistic interests such as environmental concerns or cost reduction, possibly further motivated by specific governmental incentives is driving the behaviour of the prosumers. Communities provide many potential benefits to the prosumers themselves as well as to the European and global power system. These benefits include greenhouse gas emission reduction, energy saving, promotion of renewable energies and direct economic gains to the energy users. Prosumers, while incurring higher costs relative to the non-cooperators due to the adoption of new technology and constraints imposed by the goal of reducing emissions and coordinating efforts, are also able to gain financial benefits (potentially greater than the incurred costs of cooperation) as they participate in energy transactions within the group at more favourable conditions than in the outside energy markets, given a relatively high level of participation. Through self-organisation, for example, prosumers are able to actively affect the group energy balance and be actively involved in day-ahead energy planning. Socially-oriented behaviour among individuals with common motivations and interests (e.g., environmentalists) is also expected to emerge, positively contributing to the benefits enjoyed by the community members.

3 Lessons from a Game Theoretic Model of Bounded Rationality

Agents are considered as productive units (one can think of an agent as an individual or a family). Their source of revenue is assumed to positively depend on two factors: the availability of an indispensable resource for both productivity and livelihood, such as energy broadly conceived, and the amount of effort agents put in their productive (income-generating) actions, which itself is an increasing function of the energy abundance. That is, both the energetic resource and the (energy-absorbing) work effort enter in the agents' (twice-continuously differentiable) "income production function" $f(E, R)$, where E represents the community effort (e.g. the aggregate energy consumption) resulting from the actions of the n agents, and R is the resource available to the community (e.g. the available energy level, which may either be entirely consumed in a given time period, or saved in part for future consumption). Formally, letting e_i be the individual effort (loosely speaking his/her carbon footprint), which can either take value e_c for a prosumer or e_d for a traditional energy consumer ($e_c < e_d$ due to the more sustainable practices of the former), the following inequalities are assumed to hold:

$$\frac{\partial e_i}{\partial R} > 0, \frac{\partial f(E,R)}{\partial E} > 0, \frac{\partial f(E,R)}{\partial R} > 0, \frac{\partial^2 f(E,R)}{\partial E^2} \leq 0, \frac{\partial^2 f(E,R)}{\partial E \partial R} \geq 0, \frac{\partial (f(E,R)/E)}{\partial R} \geq 0.^3$$

³ These assumptions are generally employed in the literature concerning resource exploitation in a common pool resource, such as, for example, a fishery where a community of fishermen have access to it and each can decide on the individual level of exploitation (jointly affecting the sustainability of the resource utilization). Given the similarities we have drawn between this type of community operating in a traditional ecosystem with the one considered here and its relationship to the environment and the socio-ICT ecosystem, we believe the assumptions to also be plausible in the context of this paper.

Let's go in further detail about the model. It is useful to consider again the joint level of effort E resulting from the actions of the n agents choosing their level of effort e_i ; letting $f_c \in [0,1]$ be the proportion of cooperators (prosumers), we have $E(f_c, R) = f_c * n * e_c + (1 - f_c) * n * e_d$. We assume that f_c is continuous and non negative and note that for positive levels of e_c and e_d , the total level of effort is a decreasing function of the frequency of cooperators.

The two effort levels, that are here assumed to sum up the behavioural inclinations of all agents in the community, are bound below by the efficient resource use level and above by the static Nash equilibrium level. This amounts to require that both agent types follow practices that are above those that would maximize collective utility, but to a different extent: the traditional energy users ignore the emergent social norm prescribing the socially agreed-upon effort e_c by choosing a greater level e_d (up to the individually rational but inefficient Nash equilibrium level resulting in resource overuse), while prosumers stick to e_c , which, as a special case, may coincide with the level that efficiently trades off the individual incentive towards high or uncoordinated energy utilization with the social need to impose constraints to guarantee a sustainable resource use (which ultimately benefits the individuals). Letting E_{eff} be the community efficient level, and $e_{eff} = \frac{E_{eff}}{n}$ the corresponding individual efficient level, and e_{nash} be the Nash equilibrium level of effort, we formalize what stated above as: $e_{eff} \leq e_c < e_d \leq e_{nash}$.⁴ These conditions guarantee that, at the aggregate level, positive rents from productive use of the energetic resource can be maintained. That is, the average product of labor is assured to be above the opportunity cost of labor independently from the share of defectors, providing the incentive for agents to increase their energy consumption (as they can earn positive profits for positive levels of effort). It is further assumed that $f(0, R) = 0 = f(E, 0)$ for the obvious reason that strictly positive levels of energy and resource are required to generate income via the function $f(E, R)$.

An example of a function satisfying the above conditions is the familiar Cobb-Douglas formulation with decreasing returns to scale: $f(E, R) = E_t^\alpha * R_t^\beta$, $\forall E \geq 0, R > 0$ and $\alpha + \beta < 1$. These well-behaved functions guarantee the existence of an optimal solution to the aggregate payoff maximization problem: $\exists! \text{ argmax}(\text{aggregate payoff}) = E_{eff}$ s. t. $f'(E_{eff}) = \frac{w}{p}$, where $\frac{w}{p}$ is the ratio between the opportunity cost of labour w and the market price p of the energy-demanding good produced.

We know that the aggregate level of effort in equilibrium if all agents play according to the Nash equilibrium will be above E_{eff} as each individual will consume more energy than is efficient.

⁴ Note that a direct implication of such constraints is that $E_{eff} \leq E \leq E_{Nash}$. See Dasgupta and Heal (1979) for a pioneering contribution on exhaustible resources.

Let's turn to the resource dynamics and its interaction with the social dynamics occurring as a result of human action. On the one hand we account for a positive rate of growth of energy, as confirmed by current data on production.⁵ In order to consider the entire picture, however, we must include the function $f(E, R)$, which depicts the aggregate energy use by the individuals. Formally,

$$\dot{R} = r(R) - f(E, R) \quad (1)$$

Where $r(R)$ is the energy growth rate (in the absence of usage) and \dot{R} indicates the time derivative of the resource stock, i.e. the overall energy growth rate resulting from the interaction of production and utilization. Note that the energy creation side, which is captured by the first term in the right-hand side of (1), is exogenous with respect to the frequency of prosumers (f_c), which affects instead the second term (representing the energy consumed by the community for productive tasks). For the sake of simplicity, in this paper we assume the resource stock R to be constant and equal to an exogenously fixed level \bar{R} ; that is, we focus on communities whose demand for energy is small relative to the energy production side. This framework, however, can easily be extended to model bigger actors whose collective action may result in an imbalance between energy growth rate and community usage due to overconsumption.

With these notions in mind, we are now ready to shift attention to the strategies and tradeoffs faced by (non.cooperative) traditional consumers and prosumers.

The individual payoff given resource \bar{R} and the behaviour of all community members is given by:

$$\pi_i(e_1, e_2, \dots, e_N) = p \frac{e_i}{E} f(E, \bar{R}) - w e_i \quad (2)$$

Equation (2) applies to all individuals, independent of their types, as it represents the amount they can make "in isolation", i.e. without the help of fellow community members. This amount is proportional to the aggregate payoff, in relation to the individual's energetic uptake e_i , which enters positively in the first term in the right hand side of (2) and negatively in the second term representing the work-related costs. To account for benefits that accrue solely to the prosumers (see section 1 and 2 for the rationale behind these cooperator-exclusive benefits), which may offset the disadvantage inherent in their more sustainable energy utilization depending on the relative abundance of cooperators, we introduce equation (3):

$$\begin{aligned} \mathcal{U}_i = \pi_i + \omega(f_c) * \max\{\pi_d(E) - \pi_i(E), 0\} &= e_i \left(p \frac{f(E, \bar{R})}{E} - w \right) + \omega(f_c) * \\ \max\left\{ (e_d - e_i) \left(p \frac{f(E, \bar{R})}{E} - w \right), 0 \right\} & \quad (3) \end{aligned}$$

Note that the above equation boils down to $\mathcal{U}_c = \pi_c + \omega(f_c)(\pi_d - \pi_c)$ for the prosumer and $\mathcal{U}_d = \pi_d$ for the traditional energy consumer (defector). While the latter

⁵ The 2008 Observ'ER inventory on electricity production in the world reports a 3.5% growth rate over the 1997-2007 decade for total production, with a 3.8% growth rate for conventional production and a 2.6% rate for renewable electricity production. The yearly 2006-2007 growth rate is estimated at 4.5% overall, with conventional production growing at a 4.7% rate and renewable production at 3.6%.

can only count on their productive activities to generate income, the former can also tap in the community. As participation in the Enerweb is voluntary and observable, these benefits are denied to non-members; further, it is assumed that the community benefit function $\omega(f_c)$ is increasing and concave in the number of participants. Notice also that it multiplies the payoff difference between defectors and cooperators, as the cohesion of the prosumer group will be higher the biggest the reduction in energy consumption they undertake to be sustainable (leading to a larger negative productivity gap with respect to traditional energy consumers).

For explanatory purposes, consider the following functional form for the prosumer reward function: $\omega(f_c) = f_c^\lambda$. Graphically, for two distinct values of λ equal to 0.2 and 0.5, it would yield the following:

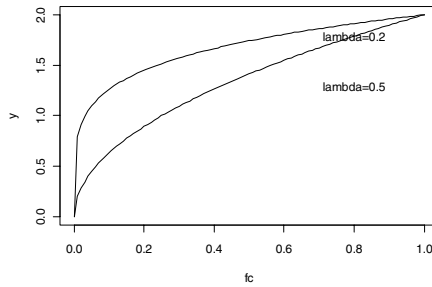


Fig. 2. The community reward function and its dependence to the frequency of cooperators

The analysis of the behavioral evolution of agents facing decisions on their energy practices is conducted by means of replicator dynamics. By so doing, we avoid the complete rationality requirements typical of models of optimization, while retaining some convergence towards the imitation of successful behavior. Agents, rather than rationally best responding to the actions of others as in Nash equilibrium, update their strategies when given the option, and switch to the strategy of the agent with which they are randomly matched if the utility of the latter is above the individual's⁶. It can be shown that such strategy revision takes place with a probability which is proportional to the payoff difference with respect to the average: if, for example, the latter is well above the payoff of a cooperator, he or she is more likely to notice the need to switch than if the average was only slightly above the agent's payoff.

Formally, this leads to:

$$\begin{aligned} \dot{f}_c &= f_c(U_c - \bar{U}) = f_c(1 - f_c)(U_c - U_d) = \\ &= f_c(1 - f_c)(\pi_D - \pi_C)(\omega(f_c) - 1) \end{aligned} \tag{4}$$

⁶ See Taylor and Jonker for their pioneering work on the replicator dynamics.

Again the dotted superscript stands for time derivative: equation (4) models the evolution of cooperating types. We are interested in the nontrivial stable solutions, i.e. those in which positive amount of both types coexist (with $f_c \neq 0$ and $f_c \neq 1$).

4 Conclusions

One of the main conclusions that we have derived by looking at the steady states of the dynamic system given by (1) and (4) is that coexistence of prosumers and traditional consumers is stable for a wide region of parameters. We are in the process of complementing the analytical results obtained by this and other variations of the stylized model proposed here with more detailed agent and resource characteristics by means of agent-based simulations. This will allow us to get more specific insights into the research and policy questions that need to be addressed if one takes seriously the paradigm shift needed to cope with today's energetic and climatic challenges.

Bottom-up approaches may prove to be one effective way to revolutionize current energetic practices; however technological and societal barriers may discourage the spread of adoption. The recently published UKERC Energy 2050 project [15] states that "Microgeneration offers a radically different approach to meeting energy needs, but capital cost and performance are currently barriers for many technologies. However, it could be important in meeting future residential heating needs, and could help catalyse change towards low carbon lifestyles". Our theoretical effort supports this view.

References

1. Barrett, S.: Climate Treaties and "Breakthrough" Technologies. *The American Economic Review* 96(2) (2006)
2. Bergman, N., et al.: A Review of Microgeneration in the United Kingdom. Part 1: Policy and Behavioural Aspects. In: *Proceedings of the Institution of Civil Engineers* (in press, 2009)
3. Carraro, C.: The Interaction between Carbon and Technology Policies. *European Review of Energy Markets* (forthcoming)
4. Dasupta, P., Heal, G.: *Economic theory and exhaustible resources*. Cambridge University Press, Cambridge (1979)
5. Gowdy, J.M.: Behavioral economics and climate change policy. *Journal of Economic Behavior & Organization* 68, 632–644 (2008)
6. Kallis, G., Martinez-Alier, J., Norgaard, R.B.: Paper assets, real debts. An ecological-economic exploration of the global economic crisis, critical perspectives on international business 5(1/2) (2009)
7. Jackson, T.: *Motivating Sustainable Consumption - a review of evidence on consumer behaviour and behaviour change*, A report to the Sustainable Development Research Network, Policy Studies Institute, London (2004)
8. Levin, S.A.: Complex Adaptive Systems: Exploring the Known, the Unknown and the Unknowable. *Bulletin of the American Mathematical Society* 40(1) (2002)
9. Observ'ER: *Worldwide electricity production from renewable energy sources* (2008)

10. Sethi, R., Somanathan, E.: The Evolution of Social Norms in Common Property Resource Use. *The American Economic Review* 86(4)
11. Osès-Eraso, N., Viladrich-Grau, M.: On the sustainability of common property resources. *Journal of Environmental Economics and Management* 53 (2007)
12. Stern, N.: *The Economics of Climate Change: the Stern Review*. Cambridge University Press, Cambridge (2007)
13. Tavoni, A.: When fairness bends rationality: incorporating inequity aversion in models of regretful and noisy behavior. Submitted to *Journal of Public Economic Theory* (2009)
14. Taylor, P., Jonker, L.: Evolutionarily Stable Strategies and Game Dynamics. *Mathematical Biosciences* 40(2) (1978)
15. UKERC: *Low Carbon Society Report* (2009)
16. Watson, J., et al.: *Unlocking the Power House: Policy and System Change for Domestic Microgeneration in the UK*, SPRU, Brighton, UK (2006)

P2P Business and Legal Models for Increasing Accessibility to Popular Culture

Eitan Altman¹, Sulan Wong², and Julio Rojas-Mora³

¹ INRIA Sophia Antipolis, 2004 route des Lucioles, 06902 Sophia Antipolis, France
eitan.altman@sophia.inria.fr

² University of A Coruña, Faculty of Law, Campus Elviña, 15071 A Coruña, Spain
swong@udc.es

³ LIA, University of Avignon, 339 Ch. des Meinajaries, Agroparc BP 1228,
84911 Avignon Cedex 9, France
julio.rojas@univ-avignon.fr

Abstract. Recent publications have reached conflicting conclusions on whether allowing users to have unlimited free access for downloading music and films is beneficial or not for the content production industry (CPI). Not only do models differ in their conclusions, but there has also not been an agreement on the validity of data on the impact of free access to music and films using P2P on the profits of the CPI and on the demand for CDs and DVDs. We provide in this paper a model that allows to study this question using elementary mathematical tools. In particular, it includes the effect of sampling on the willingness to buy. Preliminary numerical experiments show that benefits are maximized by avoiding any control measures against unauthorized sampling over P2P networks.

Keywords: Sampling, P2P Networks, Legal Actions, Business Model.

1 Introduction

Fast evolution of the P2P technology in last years has increased the wide spread access to popular culture (music and films). It has promoted file sharing at the expense of free riding. This evolution created a major conflict between internautes and the content production industry (CPI). While a legal battle is being held in courts and in the legislation area, the conflict also finds its echoes in scientific analysis of experimental data and of the economy of P2P networks. In particular, some publications conclude that the music and film industry gains from P2P where as other conclude the contrary. The gains can be attributed to a sampling effect: by downloading music, a customer can get more information that may increase its willingness to purchase a hard copy of the CD. On the other hand, the availability of free copies for download considerably decreases the demand for costly electronic copies sold by the companies and may

constitute for many users a satisfactory alternative to the original CD thus decreasing again the incomes of the content providers industry¹.

The aim of this paper is first to propose an economic model that predicts the impact of the P2P networks on the access to culture on the one hand, and on the incomes of the CPI on the other. In particular we aim at understanding the impact of actions of the CPI on these two performance measures. Based on the economic model, we identify numerically conditions for win-win business models in which the CPI can see their benefit increase while increasing the accessibility to music and films.

We consider the following two categories of actions that can be taken by the CPI:

1. **Sampling:** Making contents available for download at their own site at a cost that they can determine.
2. **Measures against unauthorized downloads:** Investment in fighting (e.g. in courts) against non-authorized downloads.

We provide in the next section an overview of the economic aspects of sampling and of legal measures against unauthorized downloads. We then introduce an economic model and compute the expected income of the CPI taking into account the impact of their actions on the availability of the files at the P2P networks, on the demand for sampling of the file and for purchasing it.

2 Legal Actions and Economic Research on Sampling

Thanks to new technologies such as P2P networks, consumers have found tools that allow them to freely explore, choose, listen and select what music to purchase. Sampling has escaped from the monopolistic control of the music producer as an advertisement strategy, and has positioned itself in the hands of consumers. We provide below a short overview on the legal and economic aspects of sampling.

2.1 The Discussion of Sampling at American Jurisprudence

After the RIAA tried to sue more than 30,000 individual file sharers since 2003², the CPI has changed judicial harassment for a concerted strategy with Government and ISPs that, by using “cease and desist” letters, seeks to discourage file sharing with the threat of Internet service suspension. Still, we will review the two main legal cases in American jurisprudence that have looked at the subject of sampling. In *A&M Records et al. v Napster*³, the defendant argued that its

¹ According to the 2009 IFPI⁴, by 2008 its 3.7 billion dollars digital music business model had, on an international level, a 25% estimated growth. The RIAA² reported that sales of digital music have grown in 2008 by 30%, which represents 1.6 billion dollars, and constituted 32% (shipments) of the total market value and 2.7 billion dollars in total sales. The market for physical recorded music have fallen 28% to 5.8 billion dollars.

users, in a behavior that resembled fair use, sampled several musical products to make a decision on what to purchase. The Court concluded that sampling is commercial in nature, can not be done by consumers without the producer's authorization, and, hence, can not be identified as *fair use*. Napster, based on [5], even appealed on the positive effects that sampling creates on the market. The Court of Appeals again agreed with the plaintiffs, for whom file-sharing not only hurts CD sales, but also blocks the new markets which the copyright owners could and are entitled to create. Moreover, the Court of Appeals said that although sampling might have a positive effect on the market in general, the increase in sales does not deprive the owners of the right to license their works.

Five years later, the argument about the effect of sampling is reviewed in *BMG Music, et al. v. Cecilia Gonzalez* [6], who was sued for 30 songs downloaded from P2P platforms. Gonzalez argued that she could not have infringed copyright, because her downloading activity was just a previous step before purchasing music. The Court, in line with the Napster ruling, denied any identity of the sampling with the doctrine of fair use and found in the *Grokster* ruling [7] empirical support that led it to conclude that downloaded music "is a close substitute for purchased music"

2.2 Trial Costs and Attorney's Fees

RIAA, in its legal strategy, presents the defendants a pre-litigation extra-judicial settlement that ranges between US\$ 3,000 and 5,000. People who accept the offer are not brought to trial. A conviction in trial would mean between US\$ 750 [8, §504 (c) (1)] and US\$ 150,000 [8, §504 (c) (2)] in statutory damages per musical work infringed. However, any civil action [8, 17 §505] on copyright can lead to the recovery of litigation costs by any party involved in the process and the court may fix attorneys' fees to be recovered as part of the costs. On 2007, the United States District Court for the Western District of Oklahoma [9] ordered the plaintiff to pay Debbie and Amanda Foster US\$ 61,576.50 as attorney's fees, and close to US\$ 7,000 for litigation costs. On 2008, Tanya Andersen was awarded US\$ 103,175 in Attorney's fees and US\$ 4,659 in costs, by the United States District Court for the District of Oregon [10].

2.3 Effects of Sampling on Sales: Theoretical and Empirical Results

Experience goods [11] identify assets that need to be consumed before knowing their satisfaction level. Consumers make an initial selection based on information from indirect sources and will continue testing until the cost of a new trial exceeds the expected growth in satisfaction. Peitz and Waelbroeck [12], assuming that music is an experience good, argue that P2P networks offer the possibility of sampling in music. In their basic model the benefits obtained by the CPI can be increased due to a more informed consumer which would be willing to pay more for albums he really wants to buy. In an extended model, profits will grow even with lower prices of the albums, thanks to savings the CPI would have in costs of marketing and promotion.

Liebowitz [13,14] opposes this position and, based on data collected by the CPI on sales of CDs, argues that the growth of file sharing through P2P networks is the cause of the fall in per capita album sales. To him, sampling simply gives the consumer a more accurate view of his musical taste, but it does not increase his music buying level [15]. Countering RIAA's argument, Liebowitz believes that downloading an album can not be correlated with one or more unsold albums, and a replacement rate of about 5 to 6 CDs downloaded per each unsold CD is more realistic. Analysis made on surveys using different methodologies agree that sampling hurts music sales. Rob and Waldfogel [16], interviewed college students in the United States, finding this impact close to 9%. Michel [17] and Hong [18], based on data from the Consumer Expenditure Survey (CEX), found a drop in sales of 13% and 7.6% respectively. Using data from the European Consumer Technographics survey, Zentner [19] concludes that file-sharing reduces the probability of buying music by 30%, and music sales by 7.8%.

On the other side, using data from German college students, Oberholzer-Gee and Strumpf [20] could not find a direct relationship between file-sharing and declining sales of CD's, obtaining an effect "statistically indistinguishable from zero". A more recent study [21] commissioned by the Dutch government, argues that even when buying and file-sharing sometimes go hand in hand, they are not mutually exclusive, i.e., there is no direct relationship between downloading files protected by copyright and purchasing music in physical format. One of the findings points that file sharers are not more or less willing to buy music than other people, and those file sharers that buy music do not buy more or less music than non file sharers, but they acquire more value added products, like concert tickets and promotional items (*licensing*).

Andersen and Frenz [22], analyzing microdata from the Canadian population as a whole, concluded that there is no clear evidence to say that file sharing affects music sales. However, the group that shares files has a clear positive relationship between the number of downloads per month and music CD's buying, of about 0.44 albums bought for each album this group downloads. Furthermore, they observe a negative indirect influence of album prices in album sales. Price, along assortment of authors, performers and genres (even those not available in traditional stores), and desire to discover new music, is also the main factor that pushes the consumer to download music instead of buying CD's, according to Sandulli [23], who in a study based on the Spanish music market, finds a "strong heterogeneity" in the relationship between the consumption of music via P2P networks and the consumption of music through physical format (CD's).

3 P2P Business Model for the CPI

Define U to be the set of policies available for the CPI for impacting the demand. A policy \underline{u} has the following components:

1. $u^{(1)}$ is the price per sampling an item at the site of the CPI.
2. $u^{(2)}$ is the cost payed by the CPI for legal and other measures to dissuade sampling at P2P networks.

Let $a(u)$, the availability of a file in a P2P network, have the following linearized form²:

$$a(u) = a(0) + \alpha \cdot u^{(1)} - \beta \cdot u^{(2)}; a(u) \in [0, 1] \quad (1)$$

The initial availability of a file increases with the price charged by the CPI, since higher prices make the use of P2P a more attractive alternative. On the other hand, it is negatively affected by the investment made by the CPI on control measures. As reported in [22] and [23], price is the most important factor in the decision of making P2P sampling over buying music. Even if the CPI uses an aggressive pricing scheme, the effect of price is still important, so $\alpha \approx \infty$. Conversely, huge investment on control measures has been largely ineffective in reducing availability of music on P2P networks, so $\beta \approx 0$. For simulation, it is possible to make $\alpha \gg \beta$. This fact was also obtained from analytical consideration in [25], where it was shown that very popular files see their demand decrease by CPI's measures but not their availability.

3.1 Sampling as a Branching Process

We propose a dynamic model for the sampling: a sample may change the information of the internaute not only on the particular sampled item but also on other related ones. For example, it may increase or decrease its interest in discovering more products by the same artist. This may lead as a result to new sampling by the internaute. We thus assume that as a result of a sample, a demand for a random number of samples is generated. This means that we model the sampling as a branching process.

Assume that the internaute plans initially to use X_1 samples. This is a sample of the first generation. Sample i of generation n will cause further creation of $K_n^{(i)}$ samples. The total number of extra samples is given recursively by $X_n = \sum_{i=1}^{X_{n-1}} K_n^{(i)}$. Taking expectation and making $\overline{K}(u) = E_u[K_{n-1}^{(i)}]$, we get $E_u[X_n] = E_u[X_{n-1}] \cdot \overline{K}(u)$. We conclude that the total expected number of samples will be: $N = X_1 / (1 - \overline{K}(u))$.

For computation purposes, we shall again use an approximating version of this expression. Using a linear approximation, the expected number of samples $N_{p2p}(u)$ at a P2P network and $N_{cpi}(u)$ at the CPI's site, respectively, as a result of a single initial sample would be: $N_{cpi}(u) = N_{cpi}(0) - k^{(1)}u^{(1)}$, $N_{p2p}(u) = N_{p2p}(0) - k^{(2)}u^{(2)}$.

We found this approximation unsuitable, specially for the case of the sampling on the CPI site, since it does not model the fact that for high values of $u^{(1)}$ and of $u^{(2)}$, we may expect an asymptotic behavior of N_{cpi} and of N_{p2p} , respectively. To obtain this, we used the following nonlinear approximations instead:

$$\begin{aligned} N_{cpi}(u) &= \min(N_{cpi}(0), N_{cpi}(0) - k^{(1)} \cdot \log(\delta u^{(1)})) \\ N_{p2p}(u) &= \min(N_{p2p}(0), N_{p2p}(0) - k^{(2)} \cdot \log(\gamma u^{(2)})) \end{aligned} \quad (2)$$

² One can use alternative models from [24] for the availability of a file at the P2P network as a function of u .

Next we present a simple first order linear model for the sampling probability. We have:

$$\underline{p}_s(u) = \underline{p}_s(0) + C\underline{u} \tag{3}$$

where $C_{ii} \leq 0$ and $C_{ij} \geq 0$, $p_s^{(1)}(u)$ is the probability of sampling in a P2P network, and $p_s^{(2)}(u)$ is the probability of sampling at the CPI site. Sampling on the CPI site is done by purchasing an electronic version of the song.

We assume that an item is sampled at the CPI's site if it is not available at a P2P network. A successful sample is one in which the requested item was available. The expected amount of successful sampling generated as a result of an initial sampling attempt, at the P2P network and at the the CPI's site, respectively, are given by:

$$\begin{aligned} S_{p2p}(u) &= p_s^{(1)}(u) \cdot a(u) \cdot N_{p2p}(u) \cdot d \\ S_{cpi}(u) &= p_s^{(2)}(u) \cdot (1 - a(u)) \cdot N_{cpi}(u) \cdot d \end{aligned} \tag{4}$$

where d is the potential demand for an item. The expected demand for purchase is:

$$D(u) = (1 - p_s(u))dq_n + \left(p_s^{(1)}(u)a(u)N_{p2p}(u) + p_s^{(2)}(u)(1 - a(u))N_{cpi}(u) \right) dq_s \tag{5}$$

where $p_s(u) = p_s^{(1)}(u) + p_s^{(2)}(u)$ is the probability of sampling, q_s is the probability that an item is purchased if sampled, and q_n is the probability that an item is purchased if not sampled.

Sampling generates traffic that benefits ISPs. We assume that incomes due to subscription are split in a way that a content provider receives an income proportional to the demand it generates. The proportionality coefficient is denoted by c_{isp} . The income of the CPI is thus:

$$R(u) = (u^{(1)} - \tau \cdot u^{(2)} + w)D(u) + c_{isp}S_{cpi}(u) \tag{6}$$

where $\tau \cdot u^{(2)}$ is the prorated fraction of the control costs per sample at the CPI and w is the difference between the price of selling an item (hard copy) minus the marginal cost to produce it.

4 Simulation

We take for (1) the values $\alpha = 0.5$ and $\beta = 1.5 \cdot 10^{-9}$. This corresponds to a very popular file for which the measures taken by the service provider have a negligible impact on its availability (see discussion below eq (1)).

Take $u^{(2)} = 0.01 * (84,000 * 30,000) = 25 \cdot 10^6$, obtained from average attorney fees in P2P cases (US\$ 84,000, see end of Section 2.1) and number of cases (around 30,000 [3]). For $u^{(2)}$, according to [21] the actual number of samples per user in P2P networks is difficult to quantify. Still, they use a conservative estimate of under 300 songs downloaded per user in a year, so we will use a

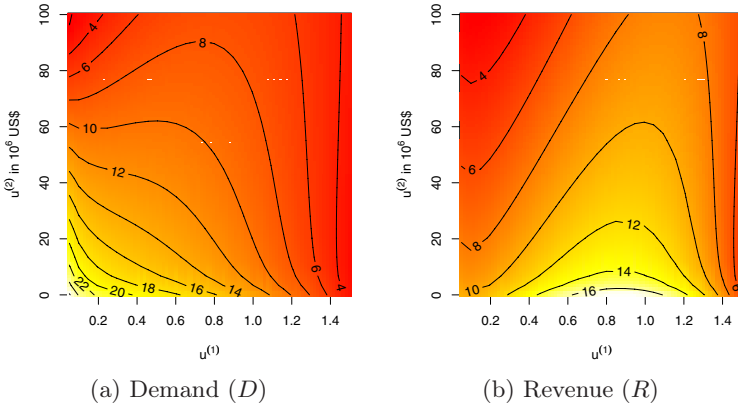


Fig. 1. Contour plots for simulation data

value of $N_{p2p}(u) = 200$. Because the effect from control measures is very small, $N_{p2p}(0) \approx N_{p2p}(u)$, we will use $N_{p2p}(0) = 300$. By requesting that (2) coincides with the values at $N_{p2p} = 200$ and at $N_{p2p} = 300$, we obtain $k^{(2)} = 30$ and $\gamma = 10^{-6}$.

From iTunes most know price, we will fix $u^{(1)} = 0.99$. For $u^{(1)}$ sampling in CPI sites can be set at $N_{cpi}(u^{(1)} = 0.99) = 20$ songs per customer (Apple [26] reports this number per iPod sold). We assume that if the price was zero, users of the CPI site would download at least the same number of songs as in P2P networks, so $N_{cpi}(0) = N_{p2p}(0) = 300$. This gives $k^{(1)} = 7.5$ and $\delta = 30$, so $N_{cpi}(u^{(1)} = 0.99) \approx 20$.

With current levels of u , let's assume $C_{12} \approx 0$. From [21] we can fix $\underline{p}_s(u) = (0.7, 0.4)$. The best case sampling probability should be close to one for the CPI site and not very far from the actual value for P2P networks, so let's use $\underline{p}_s(0) = (1.0, 0.5)$. To get better regions of $\underline{p}_s(u) \in [0, 1]$, we made $C_{11} = 0.61812$, $C_{12} = 10^{-9}$, $C_{21} = 0.29391$ and $C_{22} = 5.5 \cdot 10^{-9}$.

For [4, 5] and [6] we used $d = 1$. From [21, 80] we fixed $q_n = q_s = 0.14125$, as the buying behavior for file sharers and non-file sharers is the same. According to [2] when reporting the number of singles sold at CPI sites, we set $\tau = 1/1,033 \cdot 10^6$. From [27, 360] we make $w = 3/12$ (approx. US\$ 3 divided by the average number of songs of a CD). Finally, we fixed $c_{isp} = 0.01$.

Demand, shown in Figure 1a grows when both $u^{(1)}$ and $u^{(2)}$ decrease, but when $u^{(2)}$ grows beyond certain threshold, demand reaches a maximum level only when the price grows, showing that users increase their use of P2P networks. Revenue finds a maximum level when $u^{(2)}$ is zero and $u^{(1)}$ is between 0.6 and 1.0, as we can see in Figure 1b and follows this pattern for any level of $u^{(2)}$. Thus, by avoiding any judiciary control measures against unauthorized sampling over P2P networks, the profits of the CPI are maximized.

5 Conclusions

We proposed a modeling approach that takes into account various aspects of sampling that could have an impact on the expected income of the CPI. We studied in particular the impact of various strategies of the CPI. We used real data to trim the parameters of the simplified model (based mainly on first order approximations). Numerical experiments show that the CPI is better off, in terms of revenues, by abandoning the vast investment made in the prosecution of file sharers.

Acknowledgement. This work was partly supported by the EuroNF Network of Excellence.

References

1. IFPI: Digital Music Report 2009. Technical report, IFPI (2009)
2. RIAA: 2008 Year-End Shipment Statistics (2008)
3. Electronic Frontier Foundation: RIAA v The People: Five years later (2008)
4. United States Court of Appeals for the Ninth Circuit: A&M Records, Inc., et al. v. Napster, Inc. Nr. 00-16401 (2000)
5. Fader, P.S.: Expert Report. A&M Records, Inc. et al. v. Napster, Inc. 114 F. Supp. 2d 896 (2000)
6. United States Court of Appeals for the Seventh Circuit: BMG Music, et al., vs Cecilia Gonzalez. No. 05-1314 (2005)
7. Supreme Court of the United States of America: Metro-Goldwyn-Mayer Studios et al. v. Grokster, Ltd., et al. No. 04-480 (2005)
8. USC: (United States Code)
9. United States District Court for the Western District of Oklahoma: Capitol Records, Inc., et al., v. Debbie Foster and Amanda Foster No. Civ. 04-1569-W (2007)
10. United States District Court, District of Oregon: Atlantic Recording Corporation et al v. Tanya Andersen No. 05-933-AS (2008)
11. Nelson, P.: Information and Consumer Behavior. *The Journal of Political Economy* 78, 311–329 (1970)
12. Peitz, M., Waelbroeck, P.: File-Sharing, Sampling and Music Distribution. SSRN eLibrary (Dicembre 2004)
13. Liebowitz, S.J.: Will MP3 downloads annihilate the record industry? The evidence so far. *Intellectual property and Entrepreneurship* 15, 229–260 (2004)
14. Liebowitz, S.J.: Testing File-Sharing's Impact by Examining Record Sales in Cities. SSRN eLibrary (2006)
15. Liebowitz, S.J.: Economists Examine File-Sharing and Music Sales. In: Illing, G., Peitz, M. (eds.) *Industrial Organization and Digital Economy*. MIT Press, Cambridge (2006)
16. Rob, R., Waldfogel, J.: Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students. *Journal of Law and Economics* XLIX (April 2006)
17. Michel, N.J.: The Impact of Digital File Sharing on the Music Industry: An Empirical Analysis. *Topics in Economic Analysis & Policy* 6(I) (2006)

18. Hong, S.H.: Measuring the Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster. SSRN eLibrary (2007)
19. Zentner, A.: Measuring the Effect of File Sharing on Music Purchases. *The Journal of Law and Economics* 49 (April 2006)
20. Oberholzer-Gee, F., Strumpf, K.: The effect of file sharing on record sales: An empirical analysis. *Journal of Political Economy* 115(1) (2007)
21. Huygen, A., Rutten, P., Huveneers, S., Limonard, S., Poort, J., Leenheer, J., Janssen, K., van Eijk, N., Helberger, N.: Ups and downs: Economic and cultural effects of files sharing on music, film and games. Technical Report 34782, TNO and SEO and IViR (February 2009)
22. Andersen, B., Frenz, M.: The Impact of Music Downloads and P2P File-Sharing on the Purchase of Music: A Study for Industry Canada (2007), http://www.ic.gc.ca/eic/site/ippd-dppi.nsf/eng/h_ip01456.html
23. Francesco, D.S.: CD music purchase behaviour of P2P users. *Tecnovation* 27, 325–334 (2007)
24. Gayer, A., Shy, O.: Publishers, artists, and copyright enforcement. *Information Economics and Policy* 18, 374–384 (2006)
25. Wong, S., Altman, E., Ibrahim, M.: P2P Networks: The interplay between legislation and information technology. INRIA Technical Report RR-0689 (2009)
26. Jobs, S.: Thoughts on Music (February 2007), <http://www.apple.com/hotnews/thoughtsonmusic/>
27. Peitz, M., Waelbroeck, P.: An Economist's Guide to Digital Music. *CESifo Economic Studies* 51(2-3), 359–428 (2005)

Disruption of the Digital Media Distribution

Margaret Tan and Miguel A. Morales-Arroyo

Institute for Media Innovation
Special Interest Group on Interactive Digital Enterprises – SIGIDE
Nanyang Technological University, Singapore

Abstract. Advancing Internet technologies and the new interactive digital media are affecting the entertainment industry's distributional models. The traditional distributional channels are being restructured and re-shaped especially in its delivery as well as marketing of such digital products and services. This paper provides a conceptual framework to analyze the impacts of the online market arising from the new digital media platform in the entertainment industry. It discusses some of the disruptions and identifies opportunities for the entertainment players to re-think their business models in order to generate revenues to sustain its growth.

Keywords: Interactive digital media, new media distribution, distributional disruption.

1 Introduction

The advancing Internet technologies have provided increasing interactivity in the delivering of web-based digital products and services such as online music, movies and games in the entertainment industry. In other words, as the Internet platform becomes more open and interactive with Web2.0 technologies, especially in the context of peer-to-peer or social networking systems, consumers can now easily obtain their online digital products or services at very competitive prices [1, 2], and for some even free [3].

Since 2000 the music industry has experienced contracted revenues [4, 5] and although some have blamed or attributed the reduced revenues to the illegal online downloads [6], a more rational approach is required to understand the phenomena. Not surprisingly, the movie or film industry, in terms of video rentals [4], too has also been affected as it is so much easier to download or watch the movie titles online. It is therefore important that traditional distributors understand the new circumstances [7, 8] so as to restructure the way they operate in order to generate more sustainable incomes. Indeed, the online entertainment industry faces new challenges and approaches to redefining their business models which in turn will impact the way the goods are produced, distributed, and marketed [9, 10]. Prior to the Internet, most of these entertainment players operate much on a silos-basis [17]. Their distribution channels were independent as they took the financial risk and at the same time created vertical structures for production and distribution. The organizational structures of the entertainment chain have been hierarchical with relatively rigid distribution channels

[8, 16, 17]. The role of each player on the value chain was clearly defined. It was characterized by a protectionist agenda, and the structure was oligopolistic [18]. Although, there were some interactions in some products, but they kept their independence in the distribution. In general, the music, film, and games contained in physical entities, such as tapes, DVD's, film, CD's and cartridges were protected either by copyright [9, 14] or some technical features [37]. Production and distribution of physical entities have a cost, and since it uses physical objects, there are limits to how much of it can be produced or reproduced, and of course, in this context, economies of scale make sense.

Media distribution has been studied from different perspectives. It has been analyzed as supply chain [8] and thus, the effects of online distribution in the reduction of marginal cost have been identified [1] as well as its value added strategies to achieve more efficient distribution [2]. Further, the conditions that benefit or hamper the distribution of content, both in the traditional and online markets, have also been identified [41]. Since copyright has been used as a control mechanism [9, 14] by the media industry, the authors are motivated to analyze its implications on distribution. Thus, the objective of this paper is to provide an analysis of how the online market space disrupts and affects the traditional distributional channels. The paper also discusses the opportunities that players may have to create or strategize in order to generate sustainable revenue streams.

2 A Conceptual Framework

This paper proposes a conceptual framework that looks at the new approaches to distributing interactive digital media (IDM) based on two dimensions: the ownership of intellectual property and market channels as shown in Figure 1. These two dimensions are chosen as they tend to portray the characteristics of such entertainment products or services. For instance, most entertainment products hinge heavily on protecting the copyrights of the content creators or producers, hence, we think the first dimension, intellectual property, is important in the analysis. Further, with Web2.0 technologies, and as markets move online with networking systems, the way the digital products and services are being delivered will take a new form, hence, we think the second dimension, distributional channel is equally important to understand how the markets will be shaped.

The first dimension defines the ownership of the content creator in terms of its intellectual property (IP) and it covers all types of artistic and literary creation of information and knowledge such as trade secrets, patents, trademarks, design rights, and copyrights. IP law provides the copyright holders of the creative production complete entitlement to reproduce their work. In other words, it secures a monopoly over their work and provides them with the entitlement to reproduce and market copies of their work for a period of time or to entitle others on their behalf [11].

In the context of online markets, copyrights are essential to protect illegal downloads of music, movies, and games. In figure 1, we portray the arrow to show the IP dimension that begins with full control of one's copyright content to a situation where the content can be contributed by many people in the public domain (that is from bottom to top – depicting 'one-to-many' or from 'private-to-public' domain).

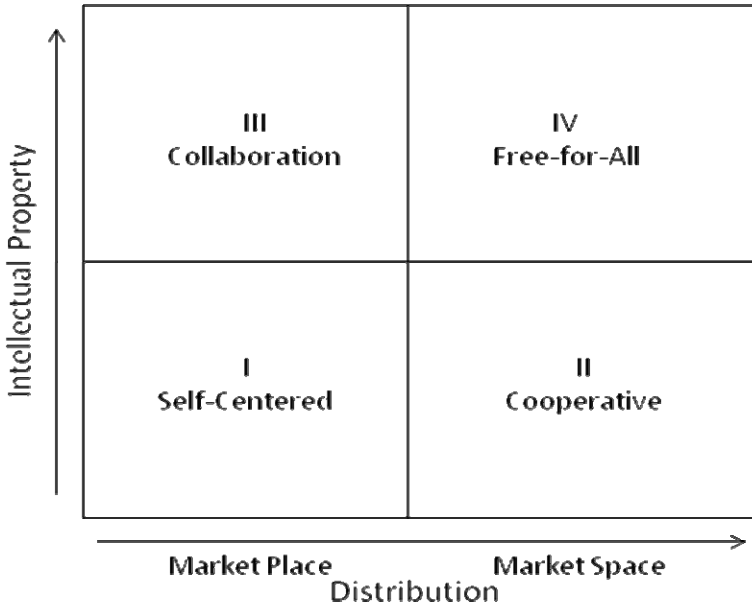


Fig. 1. A Conceptual Framework of IDM Market Channels

Public domain is described as a situation where the content can be created or contributed by various people and thus it would be difficult for anyone party to control or own the creations or productions [12].

The second dimension depicts the intermediaries, the distributional channels and they can take various forms of descriptions such as aggregators, syndicators, or just simple distributors. Aggregators are specific distributors in which content from different providers or producers is distributed by the same intermediary [13]. Sometimes aggregators are also syndicators. Syndication is a type of distribution in which limited rights are given for the media content which is sold to various distributors – this happens especially in the movie industry where its rights prescribe the number of runs for the movie. It provides exclusivity in a specific market.

In the traditional market, the distribution of content tends to be in the recorded form encapsulated in physical entities, such as CDs, DVDs, and shrink-wrapped video games. It can also be distributed using different platforms like IPTV which includes Video-on-demand (VoD) and broadcast [13]. As the Internet can eliminate cumbersome physical process of returning a rented CD or DVD [14], it can provide, theoretically, access to abundant content [3]. The x-axis in figure 1 shows the distributional channels beginning with the physical ‘market place’ and moving to the online ‘market space’ to the right. With online market space, traditional distributors who have been distributing the physical entities will also distribute digital content. In this context, the traditional distribution structure will be disrupted [8, 15].

We will analyse each of the quadrant in Figure 1 in terms of its characteristics and how the two dimensions, IP and the distributional channels have been restructured or disrupted by the new interactive digital media.

2.1 Quadrant I: The Market Place under Intellectual Copyright

Quadrant I typifies the traditional market place where the distributions of physical objects are legally protected by copyright regulations or laws. The distributors are in control of the value chain as they assume financial and marketing risks and the industry is characterized by a protectionist agenda with an oligopolistic structure [16, 17, 18]. Sometimes, the distributors synergize the channels to the markets by integrating backwards with vertical production, resulting in a hierarchical chain that is relatively rigid [8, 16, 17]. There are three types of distributors in the movie and video industry – broadcast (network and cable), syndicators, video distributors (DVD) - where the delivery includes cable and local stations broadcast or video rental and retail outlets. Besides the distributors in the value chain, there are also the advertisers who can deliver the content free [19]. In terms of generating revenue, providers use different distribution windows to strategise its revenue; for example, theatrical pay-per-view, re-runs, and DVD rentals among others.

In the music sector, the music distribution channel comprises a set of sequential activities, serially interdependent, with linear value creation and high vertical integration [8]. The distributors would obtain the rights from the content creators, and then subcontract a manufacturer to produce the physical entity (CD) which they would then sell the music titles in the CDs to the whole sellers who would then distribute them to retailers in the shops [20].

We call this quadrant ‘self-centered’ since each entertainment sector is concerned solely with its own needs and interests. Given its oligopolistic structure, they are relatively independent of outside influence.

2.2 Quadrant II: The Market Space under Intellectual Copyright

Quadrant II is also characterized by the traditional control of the intellectual copyrights of the creators or producers. However, the distributional channels are disrupted by the more pervasive online market which we called it the ‘market space’. In the distributions of digital entities, the copyright content tend to come under threat by peer-to-peer online sharing infrastructure and platform. With the new interactive media, and especially with Web2.0, the distributional channels seem to take a different form as technology makes it possible for consumers to produce or re-produce copies. Web-based or online distribution is a norm on the Internet [21]. Global sharing is therefore simple especially for those having the common language and also having with similar interests. For instance, it is possible for some music creators to get access to their audiences without the intermediation of music labels. Radiohead, a British music band, sold their album, “In Rainbows”, on its website and fans could pay as much as they like [22]. However, it is not possible for every band to market directly to their audience since few bands have the brand recognition. In fact, the Internet while facilitating new types of distributions has eliminated some intermediaries especially in viral distribution which occurs when a content provider permits any one to display his content without any restriction from the author - the case of YouTube videos. For instance, anyone can embed YouTube videos in his/her website, however, the author does not have any control over the destination [23].

New ways to discover content are also possible with web tools such as blogs, social networks (e.g. Facebook) as well as engines that track user behavior, and use that information to suggest products or services [24]. For example, in the “Wisdom of Crowds”, it allows users to express their opinion by voting and is based on the concept that a diverse group with independence of thought and decentralization of knowledge makes better decisions [25, 26] including disseminating summaries or headlines using web feeds software (RSS).

We call this quadrant ‘co-operative’ since the activities needed to deliver content require the participation of multiple actors from content to service providers such as search engines and firms that provide payment services. With the flexibility, players in the ‘market space’ face multiplicity of roles in the dynamic relationships, for example, content providers have to generate co-values, find ways to share revenues, as well as getting use to the new conditions where restrictions (like resources, space, and time) are not as pronounced as in the physical world [21, 27].

2.3 Quadrant III: The Market Place in the Public Domain

Quadrant III is typified by the distribution of media content in the public domain. The term ‘public domain’ applies to creative works not assigned to IP regulations and is often contributed by various participants much similar to ‘shareware’ in the information technology. In other words, the content is not specifically controlled or owned by anyone so it is considered as public property. In the creative commons concept, creative works enter into public domain by four possible paths: the copyright that protects the work expires; the copyright holder fails to renew the copyright; the copyright holder decides to place the work in public domain; and finally, the copyright law does not protect the work [28, 30].

However, the term ‘public domain’ needs some clarifications as various countries define the span and scope of public domain differently, hence having different jurisdictions [29]. There has been little content produced commercially in the public domain. One exception is the movie “It’s a Wonderful Life” from Frank Capra. This movie was in public domain from 1974 to 1990 since the copyright owner did not file a renewal application. During the period, television stations were able to broadcast the movie without paying any royalty rights, hence reducing their costs [31].

We call this quadrant ‘collaborative’ as creators or producers of digital content need to collaborate to take advantage of the possibilities that offers the public domain.

2.4 Quadrant IV: The Market Space under Creative Commons

Quadrant IV is typified by the distribution of content under Creative Commons (CC) licenses and public domain. One of the burdensome characteristics of today’s copyright law is the need for creators to clear rights (obtain permission to use a work), and this process can be expensive [32]. Another factor that complicates the creative process has been the copyright term extension. According to some experts, the industry has over extended the law. This is because each time a modification is made, a new copyright has to be filed [30]. Indeed, the entertainment industry uses copyright law to validate their rights over IP, a process that has become expensive such that independent creators have difficulty trying to bring about innovations. The expansion of intellectual property rights can inhibit the innovation process [33].

Indeed, with the concept of creative commons, creators can remix their creations or productions with previous works without having much concern about copyright. Creative Commons' licensees provide the capacity to share content that could be copyrighted, but reducing barriers such as the high legal costs and procedures. In a way, creative commons can work alongside IP regulation [35]. Currently, it is difficult to know what position the media industry will take in the future in relation to the alternatives that Creative Commons' licenses offer. However, the entertainment industry has maintained a protectionist agenda that look after the old ways of distribution [18]. It is interesting to see whether the entertainment industry will be able to embrace the present technology in order to generate sustainable revenues [36].

We call this quadrant 'free-for-all' since the creators and producers are all willing parties who may be interested in sharing their content and distribute them to the consumers at perhaps, little or no costs.

3 Viable IDM Market Channels

In order for the entertainment industry to ensure the viability and sustainable of their distributional channels in the advancing digital platforms, it is therefore important that the creators, producers, and distributors understand the forces in the markets. Looking at the conceptual framework, we believe that quadrant I – 'self-centered' - provides the traditional approach that may be uncompetitive in the future yet it is unlikely that quadrant IV – 'free-for-all' - is attractive as a business model because of its unpredictability and lack of structure with regards to its contents as well as markets. However, can quadrant II – 'co-operative' or quadrant III – 'collaborative' - be attractive in building new business models?

3.1 Moving from 'Self-centered' to 'Co-operative' Distribution

It seems that various copyright holders in the industry are moving from Quadrant I to Quadrant II, that is, from the 'self-centered' to the 'co-operative' area. In this context, as content undergoes tremendous transformations from being encased in a physical entity to an online digital form, the traditional distributional channel will be disrupted. Therefore, the monopolistic situation with good revenue will be affected [3] as the enforcement of digital rights will be difficult, with easy to share online content [36].

Traditionally the music and video industries keep control of the distribution via three mechanisms: copyright enforcement, licensing physical products through distribution windows, and anti-piracy tools [37, 38, 29]. For instance, the lucrative video sector that used distribution windows to address different non-competing market will be affected. The distributors who used to have control over the distribution process by defining time, location, and number of runs, and targeting mass audiences [39, 40, 41] will find some discord with its revenue as they no longer are able to package or bundle unpopular items with the popular ones to reduce unintended losses [17]. Take for example, many of us would have experienced that in a music CD, the content providers tend to include one or two good songs packaging the rest with unpopular songs. This is also the case with some unsuccessful films that were also bundled for syndication purposes such that successful movies were bundled with unsuccessful ones and broadcasted on pay-tv or public TV to reduce costs.

In terms of the ease of use to share online content, the pace of technology makes it difficult to find acceptable definition of “fair use” in the Internet environment. When content is contained in a physical object, fair use is associated with the number of times, number of views, number of pages and so on. Indeed, the media industry has tried to lobby for legislation that forces hardware manufacturers and software developers to implement measures that limit the possibility to re-create or produce copies [14, 34]. This is, however, not easy especially with advancing technologies. Indeed, the system may become obsolete before the legislations are approved [34, 11]. The Recording Industry Association of America has stopped taking legal steps against consumers who were sharing music on the Internet. The new strategy is to work with Internet Service Providers (ISP), who will then monitor the users such that if a user is sharing music, the user will be notified. When the behavior persists, the bandwidth provided to the user will be reduced. However, if the user refuses to comply, the ISP may then discontinue the bandwidth access [35].

The music and film industries have also used anti-piracy tools to reduce the dissemination of illegal content, but these measures have not been successful. Unfortunately, anti-piracy tools can be broken almost as rapidly as they are generated, and this information spreads quickly in the Internet. Independently of how fast the anti-piracy tools are implemented, there is always the possibility to record with a video camera or recorder directly from the player [14, 37, 9, 42]. Some of the anti-piracy tools create annoyances for users with usage constraints [10]. As video players become commoditized, distributors have fewer alternatives other than content differentiation. However, absolute exclusivity will be difficult to implement given the instantaneous sharing on the Internet [40]. Indeed, the industry needs to find ways to optimize its distribution, and to look for strategies for re-purposing the content in the interactive market place as using copyright, anti-piracy tools, proprietary hardware and software to build a walled garden around content may not be sustainable.

3.2 Impact of Players Moving from ‘Self-centered’ to ‘Co-operative’

The recording music industry has a market size worth US\$36.9 billion in 2000 [43]. However, in 2007, there was a reduction of 26 per cent to US\$27 billion [4]. The bulk of the income seemed to come from the sale of CD’s and DVD’s, that is, the songs in the tangible objects. Essentially, the physical units sold had reduced, and digital products sold had been increasing [5]. Besides, the International Federation of the phonographic Industry (IFPI) reported that 15 per cent of the revenue for 2007 came from digital products [6]. Overall, it was reported that the global music and video market had a contraction of more than 2 per cent from 2003 to 2007 [4].

As a consequence of the contraction of the physical market, the retail music and video rental stores were closing down. An estimated 25 per cent of music and video specialists closed its doors in UK from 1999 to 2004 [44]; in Canada, the Music World having 72 stores across the country filed for bankruptcy in 2007 [45]; in US, about 800 music stores closed in 2006, including 89 from Tower Record [46]. The recent casualty is Virgin Megastores which would be closing down in the coming months [47]. Blockbuster, a worldwide movie retailer and video rental, had the following financial results in the reds: -29.2 per cent in 2002, -16.6 per cent in 2003, -21.1 per cent in 2004, -10.3 per cent in 2005, and 1.0 per cent in 2006 [4].

3.3 Opportunities Moving from 'Self-centered' to 'Co-operative' Distribution

Indeed, the music industry has been forced to find new ways to address the demands of their users. Last.fm, a social music network, introduced a free service that allows its fans to listen to songs on-demand. This service is a partnership between CBS and the four major music labels with more than 150,000 independent labels and artists. Users can stream music or pay to download to an MP3 copy of the song through Amazon.com. The streaming service is funded by advertising revenue [48]. It is interesting to note that musicians are becoming less dependent on the distributors to reach their audiences. Artists are under pressure to get publicity, and record labels have to find new ways to generate revenue. Indeed, the Internet has provided opportunities for fans to organize and invest on the musicians' work. Scars on 45, a British band, won 15,000 pounds to record an album. One of the members uploaded the group's music on a site called Slicethepie that provides opportunity to artists to record their albums by requesting fans to invest on their favorite musicians [49].

The broadcasting sector is also experiencing substantial changes. For cable operators and satellite TV suppliers, there is intense pressure to keep their digital services attractive. This is because broadcast via Internet Protocol TV (IPTV) has the potential to transform the broadcasting sector. For instance, the telecoms are tapping on the opportunity to offer bundled services, which is cheaper and simpler for consumers such as offering voice, data, and video content - a triple play that provides a survival strategy [50]. In addition, VoD services play a key role in the competition between traditional pay-TV and IPTV providers. Such factors - the interactive capabilities, the enhanced watching experience, and better content - are user requirements especially in the personalized markets [51]. The latest delivery platforms require novel methods for tracking and evaluating audience interest.

Although moving to quadrant IV is really not an option to generate good revenue nor is it a sustainable business model for distributors, however, it is interesting to note that there may be several advantages for creators. For instance, the reduction in the legal cost of clearing copyrights and the transaction barriers may enable creators to be more innovative [32].

4 Conclusion

This paper presented a preliminary conceptual framework based on two dimensions, the intellectual properties of the creators or producers and the distribution channels in the physical market place or virtual market space, to analyze the evolving business models of the entertainment industry. It appears that as content can be progressively contributed by various parties of interest in the industry, the issue of copyrights and intellectual properties can enter into a grey area whereby it is almost difficult to control even by regulations and laws. The space of the creative commons or public domain is one such aspect. Of course, such products or services would also impose a structural change with regards to its delivery or distributional mechanism.

Unfortunately, both the music and movie industries have been affected with the new media platform. This is because as the compression algorithms get better and users are able to get higher broadband speed, sharing music and movies on the Internet is a

common practice. It is unfortunate that both sectors have been slow in recognizing the opportunities that the market space could offer. The measures used to protect the industry such as lawsuits against music fans have instead created a public relationships fiasco rather than working to its benefit.

Indeed, with the new media platform, distributors of digital content and services should tap on the opportunity to collaborate to generate value and strategies with the creators or producers. Of course, finding new ways to excite the consumers with relevant content and tools, for example, the possibility of defining metrics that can measure user's level of interest on specific content or products would be an interesting proposition. The distributors must focus on the needs of the users in the market space so as to reach out to them in a novel and interactive way.

Acknowledgments. This work is supported by the Singapore National Research Foundation Interactive Digital Media R&D Program, under research Grant NRF.2007IDM-IDM00276.

References

1. Bakos, Y., Brynjolfsson, E.: Bundling & competition on Internet. *Mkg Sc.* 19, 63–82 (2000)
2. Lam, C.K.M., Tan, B.C.Y.: The Internet is changing the music ind. *CACM* 44, 62–68 (2001)
3. Anderson, C.: The Long Tail. *Wired Magazine* 10, 12, 171–177 (2004)
4. Datamonitor: Global Music & Video. Report, Datamonitor (2008)
5. RIIA. Year-End Shipment Statistics (2007), <http://76.74.24.142/81128FFD-028F-282E-1CE5-FDBF16A46388.pdf>
6. IFPI: Digital Music Report - Revolution Innovation Responsibility. Report, IFPI (2008)
7. Rabinovich, E., Bailey, J.P., Carter, C.R.: A Transaction-Efficiency Analysis of an Internet Retailing Supply Chain in the Music CD Industry. *Decision Sciences* 34, 131–172 (2003)
8. Graham, G., Burnes, B., Lewis, G.J., Langer, J.: The transformation of the music industry supply chain. *Intl Jrl of Operations & Production Mgt* 24, 1087–1103 (2004)
9. Kaplan, D.: Broadcast flags & war agst digital TV piracy. *Fed. Com. Law* 57, 325–344 (2005)
10. Newkirk, C.D., Forker, T.: Does YouTube's Bright Past Mean a Cloudy Future? *Intellectual Property & Technology Law Journal* 19, 1–6 (2007)
11. Sparrow, A.P.: Music Distribution and the Internet. Gower Publishing, England (2006)
12. May, C., Sell, S.K.: Intellectual Property Rights - A Critical History. In: Marlin-Bennett, R. (ed.) *iPolitics: Global Challenges in the Information Age*. Lynne Rienner, Boulder (2006)
13. Hjelm, J.: Why IPTV? - Interactivity, Technologies and Services. Wiley, U.K (2008)
14. Waterman, D., Ji, S.W., Rochet, L.R.: Enforcement and Control of Piracy, Copying, and Sharing in the Movie Industry. *Review of Industrial Organization* 30, 255–289 (2007)
15. Picard, R.G.: Changing bus models of online content serv. *Intl. Jrl. Media Mgt.*, 60–68 (2000)
16. Kerr, A.: *The Business and Culture of Digital Games*. Sage California (2006)
17. Vogel, H.L.: *Entertainment Industry Economics*, 7th ed. Cambridge Uni Press, NY (2007)
18. Currah, A.: Hollywood versus the Internet. *Jrl of Economic Geography* 6, 439–468 (2006)
19. Owen, B., Wildman, S.: *Video Economics*. Harvard Uni Press, Cambridge (1992)
20. Edwards, P.: *Music Industry*. Personal Communication (2008)
21. Werbach, K.: Syndication. *Harvard Business Review* 78, 85–93 (2000)

22. Siew, A.: Thank you, Radiohead. *The Straits Times* 36 (November 27, 2007)
23. TICSC.: The Internet Content Syndication. White Paper (2008),
http://www.internetcontentsyndication.org/downloads/whitepapers/content_creation.pdf (accessed on November 7, 2008)
24. Segaran, T.: Programming Collective Intelligence, 1st edn. O'Reilly Media, CA (2007)
25. Maney, K.: Techies hot on concept of 'wisdom of crowds, but it has some pitfalls, USA Today, 4 (September 13, 2006)
26. Surowiecki, J.: *The wisdom of crowds*, Doubleday, NY (2004)
27. Normann, R., Ramírez, R.: *Designing Interactive Strategy*. John Wiley & Sons, NY (2001)
28. Standford, L.: *Copyright and Fair Use* (2007),
http://fairuse.stanford.edu/Copyright_and_Fair_Use_Overview/chapter8/8-a.html
29. Saasten, J.v.: Domestic and International Syndication. In: Eastman, S.T., Ferguson, D. (eds.) *Media programming: strategies and practices*. Thomson, Belmont (2009)
30. Samuels, E.: The public domain revisited. *Law Review* 36, 389–200 (2002)
31. Bates, J.: Company Town Yule with Less 'Wonderful Life'? *Tune In. LA Times* (1993)
32. Lessig, L.: The Failures of Fair Use and the Future of Free Culture. In: Basilico, S. (ed.) *Cut: Film as Found Object in Contemporary Video*, Milwaukee Art Museum (2004)
33. Boyle, J.: *Shamans, Software, and Spleens: Law and the Construction of the Information Society*. Harvard University Press, Cambridge (1996)
34. Benkler, Y.: *The Wealth of Networks*. Yale University Press, New Haven (2006)
35. McBride, S.E., Smith, E.: Music Industry to Abandon Mass Suits, *WSJ* (December 19, 2008)
36. Seidenberg, S.: Copyright in the age of You Tube. *American Bar. Ass. Jrl.* 95, 46–51 (2009)
37. Lyon, M.H.: Tech protection measures for digital audio & video. *SCCHT Law Jrl.* 23, 643–665 (2007)
38. Moore, S.M.: *The Biz*. Silman-James Press, Los Angeles (2002)
39. Blume, S.E.: The Revenue Streams: An Overview. In: Squire, J.E. (ed.) *The Movie Business Book*, A Fireside Book, Simon & Schuster, New York (2004)
40. Tulsiani, B., Mitskaviets, I., Card, D.: *The Internet TV Value Chain*. Jupiter Research (2007)
41. Wildman, S.S.: Interactive Channels & the Challenge of Content Budgeting. *International Journal on Media Management* 10, 91–101 (2008)
42. Carter, E.L., Lunt, S.: Podcasting & Copyright. *SCCHT Law Jrl.* 22, 187–217 (2006)
43. IFPI Recording Industry Report, Intl Federation of the Phonographic Industry (2001)
44. Research, V.: *Music & Video Retailers*. Report, Veredic Research (2005)
45. Kirsch, V.: Music World Files for Bankruptcy. *Guelph Mercury* 10 (November 13, 2007)
46. Smith, E.: News in Depth: For the music world, the tune gets sadder. *WSJ* (March 22, 2007)
47. Nakashima, R.: All virgin megastores in US to close by June. *Business Times* (March 5, 2009)
48. Adegoke, Y.: Last.fm, music labels launch free music on-demand. *Reuters*, Janaury 24 (2008)
49. Ilett, D.: The stock market for songwriters websites are giving bands the chance to record albums. *The Daily Telegraph* 19 (January 26, 2008)
50. Eastwood, G.: *The Future of TV*. Business Insights Ltd. (2007)
51. Eastwood, G.: *The Future of Convergence*. Report, Business Insights Ltd. (2006)

Evolutionary and Pervasive Services

Elia Conchione

Architect, Soluta.Net, Italy, EU
econchione@soluta.net

Abstract. The strategy is to have a system that does not move data across the architecture when it is not required. The goal is also to support different architectures, namely: Information, Model and Application. The first is to be stored and maintained where the information is actually created (by the service providers), the Model in a separate repository; models are horizontal respect to the information and are common across services. Application architecture is managed by the Registry and it 's used as a directory of services. This separation of concerns as implemented in this project helps to avoid data replication and it aims at providing a better alignment between IT and the business.

Keywords: Modelling, negotiations, e-procurements, MDA [1], p2p, peer to peer, metamodelling, agreement, SMEs.

1 Small Organisations and e-Business

The SMEs are the backbone of European economy, in the enlarged European Union of 25 countries, some 23 million SMEs provide around 75 million jobs and represent 99% of all enterprises.

The subprime mortgage problem, which emerged in the summer of 2007 in the United States, has led to global financial market turmoil and growing concern about a U.S. economic slowdown. This situation of course don't help our SMEs.

Inside this scenario, it has become increasingly important the ability of enterprises to manage innovation and the creation of ecosystems to cater for short-term, short notice collaboration. Enterprises of the future will be nodes in innovation ecosystems requiring flexible, spontaneous collaboration activities.

Being more follower than leader of the change process, small organizations seem to need favorable conditions to accelerate the diffusion of the Internet and adoption of ICT technologies [2] [3] and thus to avoid a digital divide between larger and smaller enterprises and among geographical areas [4].

The digital divide that really matters, then is between those with access to a network and those without. The good news is that the gap is closing fast. The EU has set a goal of 50% access by 2015, but a new report from the World Bank notes that 77% of the Europe's population already lives within range of a network.

2 The Two Digital Divides

At the Lisbon summit in March 2000, the European Union representatives set the goal of becoming the world's most dynamic and competitive knowledge-based economy

by 2010 with the need to promote an ‘Information Society for All’, and to address the issues of the digital divide in the adoption of Internet and e-business use [4].

Comparing big and small company, the gap as defined using these per-head measures looks enormous, the growth rates tell a different story. Over the past 25 years, computer penetration has been increasing faster in SMEs than in holdings, which has not been surprising given the software saturation in complex organization. But the same is also true of internet usage, which grew by around 13% per quarter in high-income firm in the late 1990s, compared with 10% per quarter in low and middle-income one. Holdings are ahead, but the SMEs are catching up fast.

3 There Must be a Better Way

When it comes to enterprise software, in particular, this way of delivering technology creates a somewhat perverse set of economics. Software is a service at heart, albeit an automated one, but it is sold much like a manufactured good. Company have to pay large sums of money up front, bear much of the risk that a program may not work as promised, and cannot readily switch vendors.

IT firms, for their part, have to spend a lot of resources on marketing and distribution, rather than concentrating on developing software that works well and is easy to use. In our markets it is a great advantage to be first, so vendors are tempted to release programs even if they are still riddled with bugs. And because equity analysts rightly consider software firms a risky investment, such firms must grow quickly to justify their relatively high share prices, pushing them to sell more programs than customers need.

All this explains several peculiarities of the software business[5]. One is the fact that many of the licences sold are never used, a phenomenon known as “shelfware”. More importantly, many software firms have grown so fast, often mortgaging the future, that they collapse when they reach Euro 1 billion in annual revenues, sometimes never to recover. Then there is the end-of-quarter rush, spurring many firms to do anything to get deals signed and meet analysts' expectations [6].

4 Major Obstacles

Adoption of a new technology is often expensive for many reasons. The benefits gained by the user and the costs of adoption are notable. These benefits are simply the difference in profits when an organization migrates from an older technology to a new one. Benefits may also include such “non-economic” factors as the satisfaction of being the first on the block to be converted from the older technology to the new one [7].

Large organizations that have large market shares are more inclined to undertake adopting a new technology, both because they have a greater opportunity to gain profits from the adoption but also because the availability of funds to these organizations is greater [8].

More profitable organizations are more likely to have the financial resources required for purchasing and adopting a new technology. In addition, they may be better able to attract the necessary human capital and other resources that are necessary [9].

For all previous reasons the small organization should be supported in removing a set of obstacles. This requires a strategic long-term vision and planning. The key challenge for policy makers is to identify the European and national policies which directly address the removal of these obstacles, through specific actions.

The *eEurope* 2002 Action Plan was endorsed by EU Member States at the Feira European Council in June 2000. The Action Plan's objective includes an action to encourage SMEs to 'go digital'. Within this action, main obstacles SMEs face as they engage in e-business has been identified [10].

In IT vision, the lack of suitable technical and managerial staff with sufficient knowledge and expertise is a major barrier.

This shortage affects technical knowledge related to ICT, but also entrepreneurial and managerial expertise needed for operating in a networked economy.

5 The Solution

The proximity to sources of information and services is thus key to success, especially for small and medium enterprises. The proximity to facilities and information flows has been considered very important in the past years, and it is still useful to understand the growth and competitiveness of companies in the present.

The evolutionary and pervasive service platform implements the concept of Digital Business Ecosystem (DBE) [11] [12]: companies that join the community can register their services and make them discoverable by the other members of the community. The DBE thus realizes the concept of extended dynamic cluster.

It is very important to stress that this approach focuses on services and not on the technological infrastructure that is used for implementing and running business services: small and medium enterprises lacking the physical and technological infrastructure to promote and distribute their services are only required to formalize the interface of their services in order to be able to exploit the potential of the system. Once they do so, their services become visible and usable inside the ecosystem.

By using a peer-to-peer infrastructure, the project provides a service registry that collects information about business services provided by all the subscribed organizations and makes them discoverable and usable.

The main concept about this system is that anybody in the community can get to know the existence of a service and be able use this service; the importance of a service does not depend on the dimension of the enterprise which exposes it, but rather on its perceived usefulness. The ranking mechanism for services promotes the services that are used most frequently, or that have received a positive feedback by their users. Any organization in the community can decide to build its own service by using and modifying models of services that have a top ranking, introducing a next step in the evolution of the ecosystem.

A business service in the evolutionary and pervasive service platform can be whatever: it can be a flight ticket reservation system as well as flowers delivery online service. By joining the community, companies can enlarge their market visibility and improve the quality of services provided by cooperating with other actors.

6 Evolutionary and Pervasive Service

This is the solution and the implementation designed for the Evolutionary and Pervasive Service Platform; this solution takes into account the project solutions defined in the previous specifications and summarized as follows:

- decoupling between: model (IT and Business), service data, service endpoint
- alignment between business and IT: “What you see on the platform is what you get when you use it”
- support to the EOA (Ecosystem Oriented Architecture) paradigm: dependency, versioning, metamodel bases

In the architecture description we also present components that are not part of the Evolutionary and Pervasive Service Platform. These elements provide functionalities that are used by system actors so as to improve the components integration.

7 Architectural Strategies

The strategy is to have a system that does not move data across the architecture when it is not required. The goal is also to support different architectures, namely: Information, Model and Application. The first is to be stored and maintained where the information is actually created (by the service providers), the Model in a separate repository; models are horizontal respect to the information and are common across services. Application architecture is managed by the Registry and it 's used as a directory of services. This separation of concerns as implemented in this project helps to avoid data replication and it aims at providing a better alignment between IT and the business, details follow below.

Inside the Evolutionary and Pervasive Service Platform, registered services are associated to models that are stored in the Repository.

A registry entry contains only attributes that are used for: consuming the service, retrieving the business model instance, retrieving information about the service vendor, retrieving the WSDL associated with the service, retrieving models that are associated with the service

The service is thus described by models that are associated with it. An user that wants to find a certain type of services must first of all search the models that contain the desired contents, and then search for all the services associated with these models.

The Registry works like a DNS server, because it doesn't contain any meta-data that describes services but rather contains only information for associating a service to a model or an ID. Like a DNS server, it must have a fully decentralized architecture for ensuring availability and reliability, and also an automatic synchronization mechanism.

The Registry also manages the lease for registered services and uses this information to keep them discoverable. We believe this approach avoids the risk of an overload of “keep alive” requests from service providers willing to renew their services. Every registry node can use an ad-hoc algorithm for checking service availability and then unregister those that do not respond.

The Repository is used for storing models [13] and their descriptions, so it must use a database to save information and allow queries on meta-data.

We also introduce the concept of template (borrowed from Jini) to improve models search.

To describe models we use tags which are similar to folksonomies [14]. This is a way for creating a model description that is driven by the community growing around models development.

We deem the use of standards very important in a project, and that using WSDL for web [15] [16] services, the JAXR [17] API for storing models and the SQL language for performing queries is a guarantee of quality.

We use a data warehouse for performing data mining on the data which is collected by inspecting business model instances. The data warehouse can also contain model meta-data for performing powerful queries on clients' requests.

The information contained in business model instances is collected by a service crawler that retrieves them by querying the service provider, inspecting the business model instances and reporting information to the data warehouse.

8 Interactions between Components

Interactions between components are explained using *following picture*, Figure 1, which describe the steps that are performed by operations provided by the registry and by the repository.

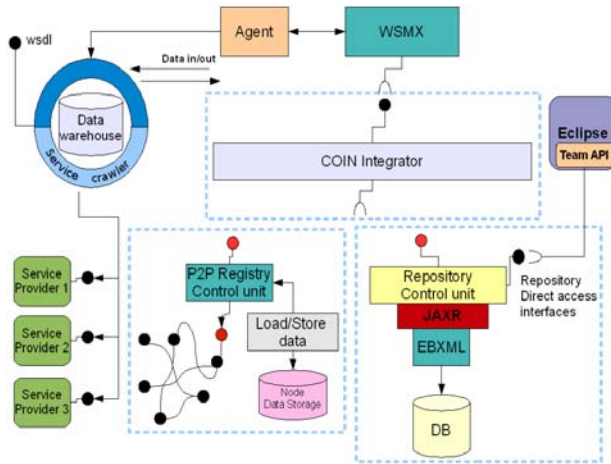


Fig. 1. Evolutionary and Pervasive Service Platform architecture

9 Registry

The Registry is used to store services that are available to users. Information is stored inside a RegistryEntry object.

The Registry is distributed over a P2P network, which provides the load balancing between nodes: every node contains $\approx (1/n * total-data) + backups$ (where n is the

number of nodes). This architecture is fully decentralized and so it doesn't have a single point of failure. When a node joins or leaves the network, the data is automatically redistributed to avoid data loss.

Registry functionalities are similar to DNS servers, which receive a domain name as input, and return the IP associated with it. The Registry, similarly, receives a ServiceID and returns the corresponding EndPoint which can be used for consuming the service.

The Registry also manages the concept of “lease”. After a service registration, a dedicated checker tests if the service is still alive once a predefined time slot is elapsed. The figure below illustrates this scenario [18].

The service provider must supply a special method that responds to a “check” request. If the service doesn't respond, the entry is deleted from the Registry.

The lease is managed at Registry side, because we want to avoid the critical situation of overload that may happen when all the services perform the lease renew at the same time. With the proposed solution, nodes manage the lease using a built-in ad-hoc strategy.

10 Repository

The Repository is used for storing models and meta-models.

The Repository [19] realizes the evolutionary concept, that is summarized below:

- supports the tracking of model versions, extensions and copies: models are never removed/modified, a change is a copy with a link to the original, models gain “ageing” concept
- supports model dependencies: solving “Versioning Hell” is one of the goals that the repository is targeting, models gain “context aware” information, the more connected, the more valuable

Users can also get access to the Repository via a special interface that exposes the services using the WSDL standard. This solution allows custom plugins development, like a new one that performs the component integration between the “Repository control unit” and the “Eclipse Team API” [20].

11 Data Ware House

This module, that is external to the Evolutionary and Pervasive Service Platform, collects information from the repository and also from the business model instances that are located on the service providers.

Operations that involve data mining are performed inside this component for improving performances and effectiveness. Custom data mining algorithms can be implemented inside it. This choice improves performances because queries regarding service properties, like discount percentage, products amount etc., are directly submitted to a dedicated optimized repository.

The current status is maintained by a service crawler, that queries services providers and collects information inspecting the business model instances that are returned.

12 Summary and Conclusions

Evolutionary and Pervasive Service Platform is the founding principle that drives the entire architecture[21] of the Open Negotiation Environment, this allows the solution to functionally scale, to have software components that are not bind to any specific negotiation and are adaptable to custom needs. P2P designs have a lot of advantages this has provided also benefit in the daily organization of the work easing the interoperability issues among partners and software components.

The architecture adopts a decentralized approach to avoid having a central node (a single point of failure) this would jeopardize the adoption of the platform. The architecture is configured in order to behave like a self supporting organism that grow over time: nodes of participant organizations are attached to each other sharing resources.

The technologies for the decentralization are built around p2p protocols, RESTful and JAXR iso standard.

By using a peer-to-peer infrastructure, the project provides a service registry that collects information about business services provided by all the subscribed organizations and makes them discoverable and usable.

References

1. Many, Model Driven Architecture
2. Association for Computing Machinery, Transactions on Internet Technology
3. Many, Rest Representational State Transfer
4. Nachira, F.: Towards a Network of Digital Business Ecosystems Fostering The Local Development (2002)
5. Stanoevska-Slabeva, K., Faust, R., Hoyer, V.: D1.2 – Software Requirements and Use Cases (2007)
6. The Economist (2003)
7. Hall and Khan (2003)
8. Bronwyn and Beethika (2003)
9. Dorfman (1987)
10. European Commission (2000)
11. Digital Business Ecosystem Project
12. Pierfranco, F.: Architecture for Digital Ecosystems: Beyond Service-Oriented Architecture (2007)
13. Many, Omg Document Number: Formal/2007-12-01, Mof 2.0/Xmi Mapping (2007)
14. Many, Folksomy
15. Fielding, R.T., Taylor, R.N.: Principled Design of the Modern Web Architecture (2002-2005)
16. Pautasso, C., Zimmermann, O., Leymann, F.: Restful Web Services vs. Big Web Services: Making the Right Architectural Decision (2008-2004)
17. Many, The Java Api for Xml Registries (Jaxr),
18. Many, Hazelcast, Clustering And Highly Scalable Data Distribution Platform for Java
19. Many, Electronic Business Using Extensible Markup Language
20. Many, Eclipse
21. University of California, Irvine, Architectural Styles And The Design Of Network-Based Software Architectures, Doctoral Dissertation

Future Internet Markets and Business Strategies: An Analysis and a Proposal

Man-Sze Li

IC Focus, 42 Clifton Road, London N8 8JA, UK
msli@icfocus.co.uk

Abstract. The Internet as a market does not comprise a single, linear value chain of actors. Companies are in the business to win, to outperform their competitors and to make profits. Developments of the Internet technology supply market to date suggest that there are six areas of focus for supplier companies' business strategies, leading to different approaches to out performing competitors and making money. The new business strategies arising from the Internet market re-define the relationships between the provider and the user. These relationships are not based solely on the calculus of economic inputs and outputs. There is a distinction between the value proposition of what is offer to the user and the value proposition of what is on offer to the provider. The business strategies aligned with the open source models, the long tail and the power of free exploit the economic fact of the margin cost of software production and delivery approaching zero. A utility approach to Future Internet markets could contribute to greater and even better choice for users.

Keywords: Future Internet, Internet markets, market competition, business strategies, business models, universal service, utility.

1 Introduction

The Internet was born of a research initiative commissioned by the US Department of Defense in 1969. Historically, it was a research spin-off. Also historically, it was a technical construct. But with the commercialisation of the Internet in the 1990s, the Internet also became a global business opportunity, arguably on a scale unparalleled in history. Today, the Internet as a maturing, worldwide, universal network of information and communications, is also becoming a social machine. Technologically, the network is beginning to resemble a global computer, a global database and a global virtualised resource. It is also increasingly wireless based, and is starting to fuse the virtual with the physical world. In business terms, the network is becoming a business system. In social terms, the network is the people and the collective intelligence of the people. The Internet has become a critical part of an economy's infrastructure, indeed a critical infrastructure for the overall society in the developed world and an embodiment of aspirations for the least developed countries. Some argue that the Internet is becoming a public utility (see, for example, [1]).

Whatever one might think the Internet is (or *should* be), one thing is certain: whatever is happening in the world, we can see at least fragments of that on the Internet. But does the Internet simply mirror what is going on in the world? This paper seeks to provide an answer in relation to development of the Internet market(s).

The paper discusses the structure and dynamics of this market, with a view towards the Internet of the future. New developments are on the horizon, in the marketplace as well as in research and development. The “Future Internet” will be much faster and smarter, more secure, embracing not just information and content but also services and real world objects (“things”) [2]. If history is a guide, the potential of the Future Internet will be primarily driven by innovative services and applications. The tussle between the different actors is reflected in the business strategies of those who seek to exploit the Internet as a commercial opportunity in competitive markets.

Six areas of focus for business strategies for Internet technology market supply are presented below. The paper calls for fresh thinking on the service infrastructure of the Internet, and proposes a utility-based approach to Future Internet markets.

The approach of the paper is business-oriented. It is based on experiences in the field, supplemented by academic and policy literature. Key literature is referenced to support the arguments presented. It is noted that while there is huge market and policy interest in the business strategies and models that tap into the Internet’s commercial potentials, the volume of papers from the research community is relatively small.

2 Internet Market(s)

When the Internet was first conceived, communication over the wire was provided by “the telephone company” of individual countries, overseen by or being part of the government. Commercialisation of the Internet formally began in 1995, when NSFNET was switched off as “the Internet” and the backbone traffic was routed instead through interconnected networks of (by then) commercial network providers, who had added data to their core telephony business. At the same time, independent Internet service providers (ISPs) sprang up to offer, to the general public, Internet access as well as some online content behind their “walled garden”. Therefore, the basic shape of the Internet as we know it today was already in place. It was a landscape comprising, on the supply side, commercial network providers, ISPs, content providers and equipment providers; and on the demand side, users, or individual consumers taking out contracts with the ISPs.

2.1 The “New” Landscape

Today, to this landscape has been added a new category: a wide variety of service providers exploiting the Internet as a communications platform. These include notably those dealing with computing or information technology, and those offering intermediary facilities for content creation and dissemination. In the meantime, the businesses, business models, identity and balance of power of the original groups of players have fundamentally changed. Network providers are no longer just shifting traffic around, and the telecommunications industry now also includes cable, satellite and mobile communications. The ISPs as an independent, distinct group of providers has largely

been eroded, while Internet access has become a line of business rather than the business of ISPs. Content providers can be said to be everyone, thanks to the Web which enables all with access to the Internet to publish. Equipment providers have expanded in range and scope beyond recognition. The “traditional” network equipment and customer premises equipment providers are joined by vendors of all kinds of devices hooked onto the Internet, including several billion mobile devices.

In short, while the supply side of the Internet market can be broken down into five broad segments, that segmentation masks a vastly complex constellation of actors with different focus, speciality and target customers.

On the demand side, “users” are not only vastly more numerous, they are also far more diverse and complicated than before. Depending on the context of use, people can access the Internet as a private individual, a consumer of goods and services, an employee, a citizen, or a member of any community to which the user belongs.

Moreover, in the long chain of communication conducted on the Internet, there are now three broad classes of intermediary users, again depending on context. These include the entities to which users belong (organisations and communities), the different kinds of communicating devices and things, and the different kinds of software-based or software-enabled services making use of other software-based or software-enabled services. The inter-relations between all of these, and with the end-user, are linked to the inter-relations between the providers on the supply side.

2.2 Market Dynamics

The above shows that the Internet as a market does not comprise a single, linear value chain of actors. The Internet of course does not replace the economy. However, it fuels the dynamics of the market. It offers new opportunities for market actors, existing as well as new, actual as well as potential. It also offers a new location – a new *marketplace* – for the exchange of some goods and services.

As a communications medium, the Internet facilitates market (in the sense of a generic, globalised market) or markets (in the sense that the global market comprises many different markets). It is useful and indeed important to distinguish between the Internet technology market and the many markets that are facilitated by Internet technologies; e.g. the failure to make such a distinction fuelled the dotcom bubble.

It is true that communications and information technologies are now indispensable to the operation of all big businesses and entire industry sectors; for many businesses, they are also accepted as a vital tool for gaining competitive advantage. However, an enabler, no matter how essential, is not core business. Not does it necessarily add value. A flourishing technology market does not automatically create prosperity in other markets it facilitates, let alone the economy as a whole.

A more fruitful line of enquiry on market dynamics and the Internet is to establish how the exploitation of Internet technologies impact on market mechanisms and structures relative to what was the norm before the pervasiveness of the Internet. It is here that technologies have a disruptive impact in the sense of radical innovation. It is also here that the business strategies of the Internet technology providers are crucial for the development of a new mode of exchange available solely on the Internet – the “Internet-based markets”. Such markets could, and may, in due course change the practice of business, break down the boundary between work and other forms of activity, and deeply influence the way in which we approach and interpret what is going on in the world.

3 Business Strategies

Economic textbooks typically tell us that technology is an important (even the) determinant of production cost. Improvements in technology make it possible to increase productivity. Technology contributes to economic growth. However, in a free market economy, commercial companies do not generally exist for this purpose. On the whole, companies are in the business to win, to outperform their competitors and to make profits. That is the essence of business. Simply defined, a business model spells out how the company makes money¹ [3]. In practice, the *what* of a business (business model) is inextricably linked to the *how* (business strategy). The tussle among providers and consumers in a competitive landscape is the most basic attribute of a marketplace [4]. In a competitive marketplace, business strategies, rather than business models per se, ultimately determine winners and losers.

During the dotcom era, there was a fashionable belief that the Internet renders business strategy obsolete. After the market crash, Michael Porter published an article arguing that the opposite is true [5]. Because the Internet tends to weaken industry profitability without providing proprietary operational advantages, it is more important than ever for companies to distinguish themselves through strategy. According to Porter, the winners will be those that view the Internet as a complement to, not a cannibal of, traditional ways of competing.

Developments of the Internet technology supply market to date suggest that there are six areas of focus for supplier companies' business strategies, leading to different approaches to out performing competitors and making money.

1. *Business strategy involving protection of intellectual property.* Owners of intellectual property acquire market power and reap competitive advantage by being legally granted exclusive rights to a variety of intangible assets. Among the common types of intellectual property, most relevant for Internet technologies are trade secrets and patents. The source code of a lot of commercial software available is a trade secret, e.g. Microsoft's Windows operating system, Apple's Mac OS X and iPhone operating system. Other ideas and inventions implemented into software could also be a trade secret, e.g. Google's search algorithm. The essence of a trade secret is that the economic benefit to its holder derives specifically from the secrecy (and not necessarily from the value of the asset being protected). In contrast, patents provide the right, within the time limit of the patent, to exclude others from making use of the invention, subject to compensation to the patent holder (usually in the form of fee-based licenses). Many technology companies assiduously build up a patents "vault" for purely defensive purposes, with no intention of putting the patents to use.

2. *Business strategy involving the bundling of technologies.* By aggregating several technologies into a single offering, a company may achieve one or several of the following which would not have been possible (or would have been less likely) if the technologies were made available separately: a greater - even dominant - market share, a bigger market, lower cost, higher price. Proponents of technology bundling

¹ Of course making money is not necessarily the same as generating or creating value, on which some definitions of business model focus instead. This distinction is illustrated by new business strategies in the Internet landscape – the subject of this section.

usually point to a superior market offering through bundling, in the sense that the overall offering is greater than the sum of its parts, and provides a more consistent experience to the user.

Probably the most famous example of technology bundling in the Internet technology field is the case brought against Microsoft first by the US Government (settled in 2001) and then by the European Commission (ongoing). The latest charge brought by the European Commission rests upon the claim that the company's bundling of the Internet Explorer (Web browser) with Windows (operating system) is an "abuse" of its dominant position in the computer market. According to the Commission, the bundling distorts competition on the merits between competing Web browsers, undermines innovation in the provision of browser-based services to consumers, and ultimately harms consumer choice. The merits or otherwise of the Microsoft case aside, the specific details show the complex inter-relations between different technologies in a particular market, and the thin line between technology inter-relations and technological inter-dependencies².

3. *Business strategy involving standardisation.* Put it simply, a standard is an agreed, repeatable way of doing something. The Internet as a technical construct was founded upon standards, originally technical protocols to interconnect existing networks. However, as the Internet develops, what to standardise, how to standardise, when to standardise and who (are best placed to) standardise, are questions that have no easy answers, or answers that can be readily agreed upon by the stakeholders.

Claims to standards and various perceived attributes of standards – notably openness and interoperability – are a hallmark of the Internet technology supply market. However, such claims are derogated by the practical difficulty of verifying standards conformance and compliance in an objective and authoritative manner.

Compared with the early days of the Internet, there is now a general reluctance among market actors to standardise technology "early", for fear of upsetting the existing market dynamics and a given market structure. In the Internet technology field, standardisation has become increasingly defensive, ad hoc, fragmented and short term. Often, it is no more than rubber stamping the status quo, i.e. the dominant technology. In addition, the process has proved to be useful for rallying an ecosystem of business partners and expediting the market momentum of the technology in question. The more general point is that standards are "leveraged" in support of specific commercial interests. But standards may not be "open". "The market" may not be the best arbiter of competing standards. In this respect, standardisation is at risk of losing its purpose and value of creating a stable foundation, or a level playing field, for facilitating innovation and ultimately delivering choice. Standardisation may even become an instrument of the market incumbents to create road blocks to innovation.

There is a further business strategy in respect of standardisation: ignore the rest and pursue own goal via publishing APIs as open technical specifications. This has been

² By way of example, complementarity between technologies, which could be argued to improve users' experience and therefore constitute an added value, might also create lock-in and erode choice. This applies especially to bottleneck asset or key technical interfaces between software programs (application programming interfaces or APIs). Function placement as defined by interfaces is highly non-trivial. It defines market structures [4].

successfully practised by the prominent Web 2.0 companies in building up their individual ecosystems³.

4. *Business strategy involving open source.* The term “open source” originates from “open source software” [6]. As a focus for business strategy in general, open source may be broadly understood as the peer development and production of the “source” of an artefact through open – typically public – collaboration. Also, the output of the collaboration process resides in the public domain, so that it is in principle available to all to exploit. The contrast is the closed, centralised and centrally controlled model of development, typically within the confines of a single company. The Internet (in the sense of a technical construct overall) could be considered as a manifestation of the open source model.

The business strategy involving open source relies on several key attributes. First, tapping into the collective intelligence of an army of volunteers for the “source” offering. Second, providing added value to the typically free source offering as a value proposition for the company’s own offering. Third, developing the distribution channel for the company offering. Fourth, managing the process of change in the source offering, including compatibility between its offering and the source offering and associated licensing. Fifth, managing the relationship with the volunteers. Sixth, ensuring the viability of the source offering development process. The business strategy can be divided into three main subclasses, in terms of added value to the functionality of the offering, added value in the form of customer service, and added value in the delivery of the offering. The first includes providing a commercial version of the source offering (with extensions/modules/plugins/add-ons), and complementary products to the source offering which may be bundled together into a single commercial offering. The second includes technical support and consultancy services. The third includes making the source offering available in an alternative way, notably software as a service. These possibilities are not mutually exclusive.

5. *Business strategy involving the “long tail”.* The term the “long tail” was coined by Chris Andersen, originally with reference to the emerging digital entertainment economy [7]. The long tail business strategy focuses on market niches of specialised and individual preferences, as opposed to the mainstream market of blockbusters; in other words, mass customisation as an alternative to mass-market fare. Examples include the online retailers such as Amazon, Netflix and Rhapsody, as well as Google, who makes most of its money from small advertisers, and eBay, who built its business based on auctioning niche and one-off products.

By catering to niches and creating micro markets, long tail companies expand the overall range and scope of the market. They also threaten the established market players whose business is built and concentrated on hits. Intrinsic to the long tail argument is the notion that the digital world is characterised by abundance, as opposed to the physical world of scarcity. The marginal cost of producing digital goods is close to and even equivalent to zero. Over a long time frame, the total average cost of producing digital goods similarly declines towards zero.

³ By tapping into their established position in a particular market, these companies reap the benefit of positive feedback to their published specifications, and further cement their market dominance by acting as a de facto standardisation hub.

As a business strategy for Internet-based markets, the long tail is particularly powerful in its focus on the smaller customer, customisation and choice. It challenges technology providers' traditional pre-occupation with large customers (in the business market) and the mass product (in the consumer market). It is also particularly significant for those technologies or software that may be delivered as services.

6. *Business strategy and the power of "free"*. Internet technology and applications started as free. Commercialisation of the Internet changed a lot of that. But commercialisation spawns its own counter culture. It appears that somebody somewhere inevitably starts offering alternatives or entirely new ways of doing things for free on the Internet. Because of the Internet, such offerings quickly spread (i.e. viral in marketing terms). Also because of the network effect of the Internet (i.e. things become more valuable to the user as there are more users), the offerings can pose a significant challenge to the established commercial products or services. Moreover, Google, one of the most valuable companies in the world in terms of market capitalisation, has demonstrated that gratis can be good business, by creating an entirely new market through the function of search.

On its own, free is not a business model - what is free by definition does not make money. Free however can be leveraged in a business strategy in a number of ways. What is free tends to attract users, often huge numbers of them, constituting a potential customer base. Things, both physical and virtual, can then be sold to them. However, in order to be viable as a business proposition, these things must be somehow "relevant" to the user. Better still is if these things are "valuable" to the user. The distinction between what is relevant and what is valuable is important. Advertising is about monetising relevance to the free. Up-selling is about monetising added value on top of what is free. Because advertising offers no intrinsic value to the user, it is unlikely to be a sustainable business proposition in the long run. Research suggests that consumers do not trust advertising, do not want to view advertising, and do not need advertising [8]. This is reflected in the price of online ads and click-through rates, which has been falling regardless of the general state of the economy.

When it comes to creating value on top of what's free, the business strategy associated with the open source model, discussed above, applies. There is however a key difference - what is offered for free in the present type of business strategy is provided directly by the company, who is moreover *the* intermediary for accessing the free offering. Even if the provider company does not own the actual content as much as it does not own the user community, it controls the channel to access. There is, therefore, another business possibility to leverage what is free - monetising access⁴.

The power of the free can be the glue for a new kind of voluntary lock-in. The site can become so useful to the user and its services can become so integrated with the user's daily activity that the user has no incentive to leave. Internet technology has a key role in making a site useful and the experience of the site uniquely compelling; free applications can provide the launch pad for entire ecosystems.

⁴ This could relate to participation in the site (which is problematic because it contradicts with building a vast base of users), or to referral, where the user is directed elsewhere on the sole basis of what he does on the site, appropriately contextualised. The greater is the user engagement (e.g. evaluation, recommendation, feedback), the more fine-tuned is the contextualisation, and the more useful the referral becomes for the user.

4 A Utility-Based Approach to Future Internet Markets

Since the 1980s, basic telephony as a universal service [9] has been widely enshrined in telecommunications legislation in US, Europe, and other parts of the world. The attributes of a universal service in telecoms typically involve affordability, accessibility, availability and quality. Fundamental to the universal service, which justifies its regulation in a market economy requiring legally binding obligations of the provider and in some cases arguments for government subsidy, is the notion that such services are a utility. Simply stated, a utility is essential for the basic conditions of living in society; therefore, the individual has the *right* to the service in question.

In classical economics, the doctrine of utilitarianism prescribes maximisation of utility as a moral criterion for the organisation of society⁵. Morality aside, the concern about the provision of universal or utility services is whether market efficiency alone can produce the optimal utility from the standpoint of society, and ensuring that rights are fulfilled as in “nobody is left behind”. There may not be sufficient economic incentives for providers to provide for adequate availability, provide access to all, hold the price down, and/or meet certain criteria for quality.

An important consideration is that a utility is subject to the scarcity principle of the resources in question, either by nature (as in natural resources) and/or by infrastructure limitations (as in the power plant for energy or the network infrastructure for telecommunications, which takes time and costs a lot to build). For all these reasons, a utility provider enjoys considerable market power: a large degree of control over the inputs (including the infrastructure as a capital investment) and economies of scale for the output (increasing returns which are disproportionate to the inputs once the infrastructure is built, and the capital outlay is compensated for over time). Because the resource in question is “essential”, it cannot be readily substituted for. In other words, its demand is not elastic – consumers cannot switch to substitutes even when confronted with large increases in prices, or artificial high prices not justified by the cost of provision. The lock-in effect of utility explains why concerns about market power and competition have been a driving force for regulation.

At the time of writing, the European Union is putting forward the argument to extend universal service to broadband communication services, which would require network providers to extend broadband coverage to all geographic areas of the EU, regardless of whether it makes economic sense or not from the viewpoint of the provider. However, if the Internet is seen as part of the fabric of economy and society (a view held by governments of the developed world [10]), then there is a case for carefully considering whether the utility argument might not apply also to additional Internet services other than broadband. For example, if basic business functions are migrated to the Internet, and those who can afford such services enjoy – for the market as a whole – uncompetitive advantages over their typically smaller rivals, basic functions could be regarded as utility services. The new Internet-based markets show

⁵ This could entail, for example, the greatest benefit for the greatest number (Jeremy Bentham and John Stuart Mill), the greatest benefit of the least advantaged (John Rawls), or just the greatest benefit (Derek Parfit). Other theorists focus on the starting position, rather than the outcome, as in free exchange from a just starting position (Robert Nozick), which also echoes Rawls’ supplementary view of fair equality of opportunity.

that the overall pie of aggregated demand could be larger where utility is a stimulus for innovation. If broadband communications are regarded as a utility, then a similar argument may be applied to essential IT services that both businesses and consumers must rely upon in order to make use of and make use out of broadband. These essential services may include basic security, identity management, search and discovery, electronic payment, ease of access, ease of use and ease of management. Users should have the genuine freedom to switch between providers of all these services, and combine services available from different providers. This freedom goes beyond “plug and play” and “open access”.

The basic cost of putting an “IT infrastructure” (or software platform) in place is intrinsically far lower than that of say a power plant or the railway network or the next generation fibre optics telecommunications network. Business strategies aligned with the open source models, the long tail and the power of free clearly exploit the economic fact of the margin cost of software production and delivery approaching zero. The processes involved in moving information around is becoming as much a “commodity” as the mechanisms for moving communication network traffic around; in the latter case causing telecommunications providers to re-invent their business models and re-brand themselves. Moreover, the value of information is generated not by the provider of the processing systems or by the provider of the communication networks, but by the end user of the Internet. Users should therefore have a greater control over what, how and even why they are being charged.

Research work in developing the Interoperability Service Utility (ISU) [11] under the auspices of the European Commission’s research framework programme, is addressing these concerns. An Internet that is fit for future business requires a transformation of the Internet infrastructure beyond today’s basic network connectivity to encompass utility services on top of which new value could be created [12]. Competition in open markets is best advanced by a level playing field. The Internet has demonstrated that technology can have a democratising effect not only to open up the boundary of the field, but also to continuously challenge and enrich the scope of the “levelers” in the field.

5 Conclusions

The history of the Internet has already shown the pivotal, gate-keeping role of the ISPs – they control access to the networks. The ISPs have declined as power players not because that gate-keeping role has disappeared; it simply means that the battleground for services has moved elsewhere - away from basic network services, which are increasingly seen as a “commodity”, and “up the stack” to the application-oriented software-based services. The economics of the Internet has changed.

In his recent work [13], Jonathan Zittrain argues that the “generative” power of the Internet is largely responsible for its spectacular growth. Zittrain defines the generative power as the ability for lots of people to build on a technology platform, and share what they do with others, without the permission of the platform-maker. The views associated with the Internet market as a collection of ecosystems have concentrated minds on the role of generic technologies in creating markets; the need to collaborate with a wider range of partners including third party developers, consultants

and channels; and a new type of market and mode of exchange brought forth by Internet-based companies.

The above views are supported by the new business strategies which re-define the relationships between the provider and the user, and relationships which are not based solely on the calculus of economic inputs and outputs. There is a distinction between the value proposition of what is offered to the user and the value proposition of what is offered to the provider. The two value propositions should not automatically be assumed to be in alignment. Importantly, Internet services, as digital goods, are not subject to the fundamental economic premise of scarcity. With an in principle abundance of market supply, the pricing of such services should arguably reflect the basic production and distribution cost of services being almost nothing.

With changing market structures, market dynamics and economic premises, the utility approach to Future Internet Markets is one proposal which calls for fresh thinking on the service infrastructure of the Internet, and how that infrastructure may be better leveraged in order to deliver greater and even better choice for users.

References

1. Reding, V.: Seizing the Opportunities of the Global Internet Economy, OECD Ministerial Meeting. In: Future of the internet economy, Seoul, Korea, June 17-18 (2008)
2. European Commission: Future Internet 2020: Visions of an Industry Expert Group (2009), http://ec.europa.eu/information_society/activities/foi/library/EPR.pdf
3. Rappa, M.: Business Models on the Web (2002), <http://digitalenterprise.org/models/models.html>
4. Clark, D., et al.: Tussle in Cyberspace: Defining Tomorrow's Internet. In: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications, pp. 347–356. ACM, New York (2002)
5. Porter, M.: Strategy and the Internet. In: Harvard Business Review (March 2001)
6. The Open Source Definition (Annotated), Version 1.9, published by the Open Source Initiative, <http://www.opensource.org/docs/definition.php>
7. Andersen, C.: The Long Tail, Wired Magazine Issue 12.10 (October 2004); Also: The Long Tail: Why the Future of Business Is Selling Less of More, Hyperion, New York (2006)
8. Clemons, E.: Why Advertising is Failing on the Internet. TechCrunch (March 22, 2009), <http://www.techcrunch.com/2009/03/22/why-advertising-is-failing-on-the-internet/>
9. US Communications Act of 1934 and its update in US Telecommunications Act of 1996
10. The Seoul Declaration for the Future of the Internet Economy, OECD Ministerial Meeting, June 17-18 (2008), <http://www.oecd.org/dataoecd/49/28/40839436.pdf>
11. European Commission: Enterprise Interoperability Research Roadmap V4 (2006) and V5 (2008), http://cordis.europa.eu/fp7/ict/enet/ei-research-roadmap_en.html
12. European Commission: Future Internet Enterprise (FInES) Cluster Position Paper, V1 (2009), http://cordis.europa.eu/fp7/ict/enet/fines-positionpaper_en.html
13. Zittrain, J.: The Future of the Internet and How to Stop It. Yale University Press, New Haven (2008), <http://futureoftheinternet.org/>

Impact of Microfinance Operations on the Livelihood of the Clients: A Review of the Existing Literature

Aliar Hossain¹, Mohammad Kamal Hossain¹, and Karim Mohammed Rezaul²

¹ Applied Research Centre for Business and Information Technology (ARCBIT),
Guildhall College

a.hossain@guildhall.ac, karunu2003@yahoo.com

² Centre for Applied Internet Research (CAIR), Glyndŵr University,

Plas Coch Campus, Wrexham, UK

morekba786@yahoo.co.uk

Abstract. The contribution of microfinance in the development of global economy is undoubtedly admissible. The research has been carried out to explore the impact of microfinance on the livelihood of the clients. The study emphasized to detect clear-cut impact of microfinance on poverty eradication, livelihood security, and on the role of Information and Communications Technology (ICT) system on microfinance operations. There are considerable debates about the effectiveness of microfinance as tools for direct reducing poverty. A plenty of researches explore that microfinance has had a significant positive impact in changing livelihood by reducing poverty and it brings a positive shift in attitude towards life and improvement in the standard of living. Clients have been successful to increase their household income, decrease economic vulnerability, and they enjoy better nutrition, health facilities and greater empowerment, gain the opportunity to pursue education, and in some cases, they have been able to wipe out poverty completely. The purpose of this paper is to illustrate the impact of microfinance on the livelihood of poor people especially who participated in microfinance operations. Toward this objective, around 50 research articles have been critically reviewed and tried to extract the findings of the impact of microfinance operations. However, due to page limitations we are unable to provide all our review information. The research also highlights the application of ICT in order to speed up the transaction processes in microfinance operations.

Keywords: Microfinance, Poverty, ICT system, Sustainable income.

1 Introduction

Poverty, now-a-days, is a global issue. It is a regular phenomenon for the under and lesser developed countries, though developed countries can not claim that they are absolutely free from poverty and they are seen to introduce various anti-poverty programs as well. Statistics shows that “among 1.4 billion poor about 1.2 billion live in the developing countries and only 200 million live in the developed countries” [1]. Poverty is unevenly spread among urban and rural area and almost every country

exhibits a rural-urban gap on various poverty indicators. The rural poor constitute 36% of the total rural population. Although urban poverty is a growing phenomenon, rural poverty still accounts for over 80% of the total number of poor people in 114 developing countries [2].

On the other hand, evidence shows that women are the most affected by poverty and “they are most disadvantaged group” [3] compared to men. Over the past decade, the number of women living in poverty has increased disproportionately compared with the number of men [4] and they lead their lives without any economic and productive works. Because the prevailing social customs act against poverty alleviation by denying women’s participation in different anti-poverty programs. But now-a-days’ poverty often forces women to work especially outside their homes, which is unusual in the context of social set up of rural areas like Bangladesh. Despite poverty forces the poor women to work outside their homes, they cannot throw out the poverty from their lives because of enjoying uneven social facilities in the male dominated society. Moreover, the existing economy system denies fair chances to the poor people. Across the world, poor people are excluded from the financial services of formal financial institutions [5] like banks, insurance companies, mortgage or house building societies. Microfinance operation has brought them under these services. “Microfinance is the provision of financial services” provides to “low-income, poor and very poor self-employed people” [6] who do not have access to the formal financial institutions. It is revealed in different researches that the participants of microfinance operations have been able to upgrade their socio-economic status by involving themselves into various poverty alleviation programmes.

1.1 Evaluation of Microfinance

Historically, the existence of microfinance was among the poor in various shapes and form [7]. Early beginners of micro credit programs in the world were the cooperative societies. Evidence shows that group formation for undertaking economic activities was an ancient concept of modern microfinance institutions (MFIs) and in the 12th century there was a cooperative type association named “kou” in Japan. Though these types of traditional and informal institutions could not achieve sustainability [8] and as a result, formal institution for the poor emerged.

Microfinance in formal shape started out from Bangladesh in 1970’s in a limited scale and Professor Dr. M Yunus, the great visionary, is the unique and only claimant the development of this concept. He has conceptualized micro credit intervention for its extensive use to eradicate poverty of millions for the world’s poorest people. The mission of micro credit has been guided by four core themes: (i) reaching the poorest; (ii) reaching and empowering women; (iii) building financially self-sufficient institutions; (iv) ensuring a positive measurable impact on the lives of the clients and their families [9].

Before the microfinance era, poor did not have access to credit and loans due to the widespread belief that the poor could not repay loans [10], though this belief has proven absolutely incorrect and “the poor are indeed responsible to manage credit and repay loans” (ibid, 2005). At present, this concept has proliferated worldwide including developed countries. Currently 1000 to 2500 MFIs are serving some 67.6 million clients around the world [11].

1.2 Approach of Microfinance

There are many different approaches of microfinance. Grameen Bank has identified fourteen different microfinance approaches. These are- associations, bank guarantees, community banking, co-operatives, credit unions, Grameen group, individual, intermediaries, NGOs, peer pressure, rotating savings and credit association (ROSCA), small business and village banking [12]. A common approach found throughout the world is the ROSCA. A ROSCA consists of a group of community members, usually 30-40 members, the group which is called village organizations (VOs), who meet regularly to pool their savings. The pool is then lent out to one member of the group, who repays it, at which time it is lent out to another group member, and so on until each group member takes a turn borrowing and repaying the pool of savings [5].

2 Impact of Microfinance Operations

Microfinance is contemplated to be an effective tool of poverty alleviation and an approach to empower the poor especially poor women. That's why; assessment of impact of microfinance operations is inevitable. "Independent studies show that microcredit has a host of positive impacts on families that receive it" [13]. Microfinance clients have been successful to increase their household income, decrease economic vulnerability, and they enjoy better nutrition, health facilities and greater empowerment, gain the opportunity to pursue education, and in some cases, they have been able to wipe out poverty completely [10].

However, it is frequently heard to say that the fortune of the microfinance clients is as like as before despite claiming the success of microfinance by different researchers, even it is learnt from media that some suicidal incidents are bound up with microfinance operations as clients have been pressurized to repay their daily installment despite their inability or they are hiding to escape from pay loan back taken from different microfinancer institutions (MFIs). These incidents demonstrate quite opposite scenarios of microfinance success and its positive impact on the socio-economic level of the target groups and that's why Wrenn E [12] commented that "not all commentators are as enthusiastic about the role of microfinance in development and it is important to realize that microfinance is not a silver bullet when it comes to fighting poverty".

On the other hand, the positive impact of microfinance rests on its policies as well as procedures and the management involved executing the policies according to procedures. In some cases, microfinance operation has not been successful because of inefficiency of management despite having good policy and procedures or ill-designed programs having efficient management. According to Bhatt N and Tang S [14] the future success of microfinance will hinged upon how MFIs develops design tailoring to specific clients. They highlight the importance of future research to develop MFI vehicles, technologies, performance assessment, sound design of MFI and its management. So, "well-designed programmes can improve the incomes of the poor and can move them out of poverty" [15] and if MFIs and their management "understand the needs of the poor and try to meet theses needs, projects can have a positive impact on reducing the vulnerability, not just of the poor, but also of the poorest in the society" [12].

Therefore, considerable debate remains about the effectiveness of microfinance as tool for directly reducing poverty and about the characteristics of the people it benefits [16] and it is extremely difficult to measure absolute impact of microfinance operations on the livelihood of the destitute people because of not having unanimity about the definition of poverty line.

“The impact of microcredit is mainly assessed in terms of the income gains for the borrowing households, the less perceptible beneficial impact on various aspects of human development is no less important”[17]. There are some common variables to measure the impact of a typical microfinance programs “includes income, diversity of income sources, volatility of income, employment generation, assets acquisition, volatility of household consumption, level of household consumption, education for children, poverty reduction, women empowerment and reduction in social exclusion of women” [8].

2.1 Impact of Microfinance on Poverty

The term poverty is a vague and ambiguous concept. It has no unambiguous connotation that if an individual does not have some particular stuff or assets means he or she is living below the poverty line. Somebody defines poverty is the lack of income. But “poverty is more than just a lack of income” [12]. Increasing income and reducing poverty is not same thing because the income increased of the poor may be spent on gambling or buying alcohol, which does not necessarily reducing poverty. To reduce poverty needs to focus on helping the poor to “sustain a specific level well-being” [18].

Despite a plenty of criticism of microfinance operations, it is proved by many researchers that microfinance has had a significant impact in reducing poverty [19],[20],[21],[22], and [23] by increasing income, other resources, number of income earners. The more effective impact of microfinance operations is noticeable on extreme poverty compare to moderate poverty [24].

Many researchers have compared poverty scenario between clients and non-clients of microfinance and scenarios before and after joining microfinance interventions. In a study conducted by Panjaitan-Drioadisuryo et al. [19] in Lombok, Indonesia, finds 90% of sample clients of Bank Rakyat Indonesia (BRI) increased their income by 112% and these families exceed the poverty line. Only 10% respondents’ income did not increase because of misusing the money by their husbands.

MkNelly and Dunford [20] and Elizabeth Dum [21] found in their study in Ghana and Peru respectively that microfinance clients increased their income by 50% higher compared to non-clients and “only 28% clients live below the poverty line compared to 41% of non-clients” [21]. Khandaker S R, [22] states that “microfinance participants do better than non-participants in per capita income, per capita expenditure and household net worth and the incidence of poverty among participating households is lower compared to non-participating households. Setboonsarng S and Parpiev Z [23] carry out a research on microfinance and Millennium Development Goals in Pakistan. They draw a comparative impact assessment of microfinance between borrowers’ households of Khushhali Bank (KB) and non-borrowers. They found strong positive impacts of KB borrowing on agricultural production, especially in animal raising activities and agricultural farming. KB clients possess on average PRs 17,705, PRs. 12,814, and PRs. 882.5 higher values of livestock, farm equipment and rental income from farm equipment than

that of non-borrowers (*ibid*). Though they did not find any significant impact on non-agricultural enterprise, durable assets, consumption, education, health care and empowerment (*ibid*).

Another indication of poverty reduction is the increase of number of income earners. In Lesser Developed Countries (LDCs) like Bangladesh, most of the households have only one income earner [3]. Microfinance contributes directly or indirectly to increase number of earners. Choudhury M O [8] states that “micro credit delivery to women assists to generate employment for the able bodied male members who may not have access to the credit facility” and MFIs that provides “credit to one woman generates employment for 2.5 persons, male or female, in the family and on hire for wage”. Though the use of loans by male relatives is viewed as the loss of women’s direct control over loans and a reinforcement of unequal gender relationships and interpreted as disempowering [25].

The destitute people of the village suffer from hunger acutely, which is another dimension of poverty and they become happy if they can enjoy foods two or three times in a day. Several researches showed that microfinance operations let them freed from hunger. Mina M S and Alam S S [26] in a study showed that 82% of respondents had self-sufficiency of food for 2-6 months and only 18% had for 6-7 months in a year before joining microfinance operation but after joining the credit programs the figures was just reverse to before joining program. Rahman M M et al. [27], however, revealed different statistics in their study that clients of microfinance operations spent a lesser portion of income to their family basic food items after joining credit program. They found that though the amount of fooding expenses increased remarkably accounting for 274.41% higher than before joining operations, but after joining credit program they were spending only 66.75% portion of income for their family basic food items which was 18.84% lower than before joining credit while major portion (85.59%) of income used to spend for that.

2.2 Impact of Microfinance on Livelihood Security

Despite having debate about the extent of poverty alleviation by microfinance, it is admissible to all that microfinance plays a vital role in alleviating poverty. Different researches reveal that microfinance boosts client’s income and thereby reduces poverty. However, rising income does not necessarily mean that client’s livelihood is secured or durable. When poor people are able to protect themselves from different shocks such as natural, social, political and economical disasters, then their livelihood is said to be secured and durable. The way of being secured and durable life is having tangible assets like land, and other resources, non-tangible assets such as saving etc. along side sustainable income. “Impact must be assessed on each of these issues if a true picture of the impact of microfinance is to be obtained” [12].

Rahman M M et al. [27] pointed out in their study that no clients of microfinance had saving before joining Grameen Bank, but a remarkable advancement in saving was found after joining micro credit program while “their savings was on an average Tk. 625” each of clients at the survey year. But they did not mention if clients’ saving were voluntary after meeting their all daily expenses or they did this to comply with mandatory rule of saving of MFIs. They found in their research that value of all tangible resources of the clients went up by 955.9% at the survey year after joining Grameen Bank microfinance program. Though an insignificant change (only 13%)

was taken place in land ownership. But drastic change was observed in livestock and poultry after joining credit program that was 1606.05% higher than before joining followed by house and modern amenities (265.43%).

Latif [28] in a study on Micro-credit and Savings of Rural Households in Bangladesh revealed that the micro-credit programs of the MFIs targeted to alleviate poverty by supplying small credits to the rural poor in self-employment activities, has a distinctly positive role in influencing household savings. For households of similar income categories, the saving-income ratios are found to be significantly higher among the participants than the non-participants.

Mina M S and Alam S S [26] showed in their study that 55% of sample respondents had been able to develop a saving after meeting their expenses, but they did not reveal how much money they were able to save each week or month and what was the savings scenario before joining microfinance? Because, the range of savings could be very little amounting to large in context of their socio-economic condition and the same group might have savings before joining microfinance program and if so, how much improvement has in this regard been developed? On the other hand, they stated that beneficiaries had been able to increase their assets on an average by 90%, out of them business assets (development of a business) were significant (1475%) followed by housing (107%) and development of cultivation (72%) as well as livestock (72%) was not insignificant as well (ibid).

3 Role of ICT System on Microfinance Operations

A well-developed ICT system is very important for knowledge share and communications between the service providers of micro-finance and the entrepreneurs (receiver). The success of microfinance operation hugely depends on the effective management, communication and training with the different department of the service provider. The development of broadband or Internet connectivity in the rural areas of developing countries (e.g., Bangladesh) can improve the services of information exchange and the accessibility in the areas for health, education, local government management and awareness of some international issues such as climate change and global warming.

In many cases, it is obvious that the policy makers or the key people related to the microfinance operation are not aware about the sustainable development. With this limited knowledge, they can focus and train only the traditional way of doing business. Microfinance has lots of advantages, which the authors have explained in this report, though its major disadvantage is lack of long-term sustainability. The primary aim of the microfinance is to eliminate poverty, which means the focal point is the financial sustainability. Then the question arises about the social and environmental sustainability. Establishment of sustainable broadband Internet system in rural areas of Bangladesh could make a favourable environment for all stakeholders to share the information, develop ideas and knowledge including social and environmental development, project management skills and entrepreneurship. On top of that, it will reduce the bureaucracy and increase the transparency of business environment and the education level.

Traditional method of employing staff in remote villages is very expensive and even it is not so easy to reach poor borrowers due to proper transportation systems. With the latest generation of sophisticated ICT systems it is possible to record all

sorts of transactions in the rural area and transfer them to the central server. Both clients and MFIs can benefit by using ICT. For example, clients do not require queuing in the line for any sort of transactions. If a controlled and secured system, such as personal digital assistant (PDA) or smart card systems is implemented, clients can conduct transaction anytime which could be more convenient banking services in a rural area and also it is possible to provide much faster loan processing than the traditional system. On the other hand, MFIs can receive benefits from better performance management due to faster and better quality data processing [29]. MFIs can increase their outreach by making services available 24 hours and closer to clients and thereby increase customer satisfaction and loyalty by introducing new types of services, such as money transfers and direct deposits. MFIs can also reduce transaction cost by reducing staff time when applying ICT based systems.

4 Conclusions

The study has been tried to demonstrate the effectiveness of microfinance on the households, people life cycle who participated in the microfinance programs. The impact of microfinance on poverty alleviation is, now-a day's, debated issue what has been noticed in the different research reviewed. Most of the researchers found that microfinance has had significant effect on the livelihood of the participants. Some critic researchers, however, argue that it is not a silver bullet or panacea of poverty alleviation. Actually microfinance can not make the clients to run in the competitive race but to walk from the crippled position only. To alleviate poverty requires a toolbox with many different categorized parts. Microfinance is not a sole solution of poverty alleviation but it can be considered one of the most effective parts of the toolbox. It must also be coupled with other social programs that are flexible to meet the diverse needs of destitute families [10].

The plan aiming for poverty alleviation designed by MFIs suffers from many criticisms. Indeed, the critics have raised question if microfinance alleviates poverty at all [10]. The critics placed five pitfalls of microfinance operations. (1) Microfinance does not reach the poorest of the poor and it is happening because of discriminated behavior of the loan officers [30], (2) MFIs are rarely achieved financial sustainability [31], (3) it is potentially harmful to women, because their husbands imagine the matter as a degradation the man's sense of masculinity if wives generate more income [32] and hence violence on them starts, (4) it creates a large debt for some poor people who are unable to repay the loans [33] and (5) it is not universally applicable especially for the young, the old, the sick, or physically and mentally handicapped [34].

Therefore, it is recommended to take appropriate measures to bring behavioral changes of loan officers to treat evenly to all clients. Moreover, MFIs should provide due training to the clients before giving loan and while utilizing loans and there should have monitoring management whether clients are utilizing their loans properly. While microfinance has been embraced as an effective tool for poverty reduction, many people in rural areas are still struggling with the appropriate utilization of this service. In order to create the awareness of positive use and impact of microfinance MFIs must have to give priorities to the modern ICT applications. Since an ICT system can lower the cost of communication, transportation and the overall business cost,

the authors recommend the development and application of ICT systems for microfinance operations.

References

1. UNDP, Human Development Report 1991. Oxford University Press, Oxford (1991)
2. IFAD (1992), cited in Khondkar, M.: Conceptualization of Poverty: The Bangladeshi Rural Women's Experience. *Dhaka University Journal of Business Studies* 20(2), 109–136 (1999)
3. Khondkar, M.: Conceptualization of Poverty: The Bangladeshi Rural Women's Experience. *Dhaka University Journal of Business Studies* 20(2), 109–136 (1999)
4. Platform for Action, Beijing Declaration and Platform for Action adopted by the Fourth World Conference on Women: Action for Equality, Development and Peace, Beijing, United Nations Department of Public Information, New York, USA (1995)
5. Baru, J.C., Woller, G.M.: Microfinance: A Comprehensive Review of the Existing Literature. *Journal of Entrepreneurial Finance and Business Ventures* 9(1), 1–26 (2004), <http://marriottschool.byu.edu/selfreliance/workingpapers/library/997.pdf>
6. Otero, M.: Bringing Development Back into Microfinance. In: Conference paper on New Development Finance held at the Goethe University in Frankfurt, sponsored by IPC, Ohio State University and the Goethe University (1999), http://www.helpendpoverty.com/file_download.asp?f=8 (accessed on 22/02/2009)
7. Banuri, S.: Impact of Microfinance on Poverty, Income Equality and Entrepreneurship. This is a term paper submitted by the author in partial fulfillment for Research design 1, Richardson, Texas (2006), http://www.utdallas.edu/~s6030200/sb_cp6_rd_microfin_pov_122006.pdf (accessed on 13/01/2009)
8. Choudhury, M.O.: An Impact Assessment of Micro-credit Programs for Rural Women-Experience of Bangladesh. *Journal of the Institute of Bankers Bangladesh* 48, 1–47 (2001/2002)
9. Nath, D.K.: Assessment of Microcredit Programme in Bangladesh: Government Sector. In: Ahmed, S., Hakim, M. (eds.) *Attacking Poverty with Micro credit*, 2nd impression 2004, pp. 49–70. The University Press Limited, Dhaka-1000 (2004)
10. Swope, T.: *Microfinance and Poverty Alleviation* (2005), <http://www.rollins.edu/olin/rurj/mpa.pdf> (accessed on 17/02/2009)
11. Sengupta, R., Aubuchon, C.P.: The Microfinance Revolution: An Overview. *Federal Reserve Bank of St. Louis Review* 90(1), 9–30 (2008)
12. Wrenn, E.: *Microfinance - Literature Review* (2005), http://www.dochas.ie/documents/MicroFinance_literature_review.pdf (accessed on 07/02/2009)
13. Yunus, M.: Expanding Microcredit Outreach to Reach the Millennium Development Goals: Some Issues for Attention. In: Ahmed, S., Hakim, M.A. (eds.) *Attacking Poverty with Micro credit*, 2nd impression 2004, pp. 7–22. The University Press Limited, Dhaka-1000 (2004)
14. Bhatt, N., Tang, S.: Delivering Microfinance in developing countries: Controversies and policy perspectives. *Policy Studies Journal* 29, 319–333 (2001c)
15. Hulme, Mosley: *Finance Against Poverty* London: Routledge (1996), Volume-1 and 2 Interregional Training Seminar Report (1990) on How to Improve Banking for Rural Women, Nicosia, Cyprus, 19-30, MIM, IFAD, INSTRAW (November 1990)

16. Chowdhury, Mosley, Simaowitz: The Social Impact of Microfinance. *Journal of International Development* 16(3), 291–300 (2004), <http://www.interscience.wiley.com> (accessed on 31/01/2009)
17. Mahmud, W.: Introduction. In: Ahmed, S., Hakim, M.A. (eds.) *Attacking Poverty with Micro credit*, 2nd impression 2004, pp. 1–5. The University Press Limited, Dhaka-1000 (2003)
18. Wright, G.A.N.: *Microfinance Systems: Designing Quality Financial Services for the Poor*. Zed Books Ltd., London and New York, and The University Press Limited, Dhaka (2000)
19. Panjaitan-Drioadisuryo, et al.: Gender, Self-Employment and Micro credit Programs: An Indonesian Case Study. *The Quarterly Review of Economics and Finance* 39, 769–779 (1999)
20. MkNelly, Dunford: Impact of Credit with Education on Mothers and Their Young Children's Nutrition (Ghana). *Freedom from Hunger Research*, Paper No. 4, Davis, CA (1998)
21. Dum, E.: *Microfinance Clients in Lima, Peru: Baseline Report for AIMS Core Impact Assessment*. University of Missouri-Columbia (1999), <http://www.mip.org> (accessed on 10/02/2009)
22. Khandker, S.R.: *Fighting Poverty with Microcredit: Experience in Bangladesh*. Oxford University Press, New York (1998)
23. Setboonsarng, S., Parpiev, Z.: *Microfinance and the Millennium Development Goals in Pakistan: Impact Assessment Using Propensity Score Matching*. ADB Institute Discussion Paper No. 104 (2008), <http://www.adbi.org/discussion-paper/2008/04/18/2526.microfinance.millennium.dev.goals> (accessed on 16/02/2009)
24. Khandker, S.R.: *Micro Finance and Poverty – Evidence Using Panel Data from Bangladesh*. Policy Research Working Paper, The World Bank (2003)
25. Goetz, Sengupta: Who Takes the Credit? Gender, Power and Control Over Loan Use in Rural Credit Programmes in Bangladesh. *World Development* 24(1), 45–63 (1996)
26. Mina, M.S., Alam, S.S.: Impact of NGO Activities on the Development of Rural Economy of Bangladesh: A case Study. *The Journal of Finance and Banking*, University of Dhaka 4(1&2), 20–38 (1995)
27. Rahman, M.M., et al.: *Micro-Credit: A Sharp Weapon for Economic Employment of Rural Women*. *Development Review* 13(1&2), 99–105 (2001)
28. Latif: *Micro-credit and Savings of Rural Households in Bangladesh*. *The Bangladesh Development Studies*, vol. XXVII, No-1 (March 2001)
29. Hishigsuren, G.: *Information and Communication Technology and Microfinance: Options for Mongolia*. ADB Institute Discussion, Paper No. 42 (February 2006), <http://www.adbi.org/files/2006.02.dp42.ict.microfinance.mongolia.pdf> (accessed on 7 May, 2009)
30. Simanowitz, A., Walter, A.: *Ensuring Impact: Reaching the Poorest while Building Financially Self-Sufficient Institutions and Showing Improvement in the Lives of the Poorest Women and Their Families*. In: Daley-Harries, S. (ed.) *Microcredit Summit Campaign. Pathways Out of Poverty*, Kumarian, Bloomfield (2002)
31. Morduch, J.: The Microfinance Promise. *Journal of Economic Literature* 37(4) (1999), <http://links.jstor.org/sici?sici=00220515%28199912%2937%3A4%3C1569%3ATMP%3E2.0.CO%3B2-F> (accessed on 09/02/2009)
32. Cheston, S., Kuhn, L.: *Empowering Women Through Microfinance*. In: Daley-Harries, S. (ed.) *Microcredit Summit Campaign. Pathways Out of Poverty*, Kumarian, Bloomfield (2002)
33. Buss, T.: *Microfinance in International Perspective: An Overview of the Issues*. *Journal of Economic Development* (1999), <http://www.spacf.com/sample.html> (accessed on 13/01/2009)
34. Versluysen, E.: *Defying the Odds: Banking for the Poor*. Kumarian Press, Greenwich (1999)

Combining Organisational and Coordination Theory with Model Driven Approaches to Develop Dynamic, Flexible, Distributed Business Systems

Javier Vázquez-Salceda¹, Luigi Ceccaroni², Frank Dignum³,
Wamberto Vasconcelos⁴, Julian Padget⁵, Siobhan Clarke⁶,
Paul Sergeant⁷, and Kees Nieuwenhuis⁸

¹ Universitat Politècnica de Catalunya, Barcelona, Spain

² Tech Media Telecom Factory SL, Barcelona, Spain

³ Universiteit Utrecht, The Netherlands

⁴ University of Aberdeen, United Kingdom

⁵ University of Bath, United Kingdom

⁶ Trinity College Dublin, Ireland

⁷ Calico Jack Ltd, United Kingdom

⁸ Thales Nederland B.V., Netherlands

Abstract. Enterprise systems are increasingly behaving as nodes in a digital, dynamic ecosystem. On the one hand, this new situation requires flexible, spontaneous and opportunistic collaboration activities to be identified and established among business parties. On the other hand, it also requires engineering approaches able to integrate new functionalities and behaviours into running systems and active, distributed, interdependent processes. In this paper we present a multi-level architecture, combining organisational and coordination theories with model driven development, for the implementation, deployment and management of dynamic, flexible and robust service-oriented business applications.

1 Introduction

A new generation of networked business applications based on the notion of software services that can be dynamically deployed, composed and adapted will make it possible to create radically new types of applications. These applications will be able to communicate, flexibly reconfigure at runtime, adapt to their environment and dynamically combine sets of simple, specialised, independent services into more comprehensive business services. This will require profound changes in the way software systems are designed, deployed and managed, replacing existing “design in isolation” engineering with new approaches able to integrate new functionalities and behaviours into running systems and active, distributed, interdependent processes. Technical progress in the area of service-oriented architectures (SOAs) in recent years has been impressive, with new models, tools and standards emerging to cover a wide range of core and related functions. The main areas of progress include:

- interoperability (SOAP [15], WSDL [18] and OGSi [12]);
- discovery and management (UDDI [11] and WS-Management [3]);

- orchestration and choreography (WS-BPEL [5], XPDL [19], ebXML [10] and WS-CDL [17]); association of semantics with Web-services (OWL-S [16] and WSMO [20]).

Furthermore, advances have come from a variety of sources, including enterprise interoperability, Grid computing, software engineering, database and knowledge-base theory, artificial intelligence, object-oriented systems and logic. This rapid progress has, for the first time, raised the realistic possibility of deploying large numbers of services in intranets and extranets of companies and public organisations, as well as in the public Internet, in order to create communities of services that are always connected, frequently changing, open or semi-open, and form the baseline environment for software applications. However, this shift brings about not only potential benefits, but also serious challenges for how such systems and applications should be designed, managed and deployed. Existing approaches in some important areas (such as security, transactions and federation) tend to cover only technology issues such as, for example, how to secure a protocol or connect federated directories, without considering the paradigm change that occurs when large numbers of services are deployed and managed over time. In particular, existing approaches do not offer satisfactory answers to these questions:

- How to manage workflows in non-trivial environments, where not all services are owned by the same organisation? Since we cannot assume that all parties are either benevolent or that they will deliver results unless explicit obligations are defined and enforced, should workflows be agreed upon by all parties before they can be executed?
- How to align the configurations and settings needed by a service to operate with those of the operational environment?
- How service execution is affected by issues of trust, rights, obligations and prohibitions?
- What if critical applications simply cease to function if services provisioned from third parties disappear or malfunction?
- How to deal with knowledge representation, when connecting or binding together two or more actual entities or services using different ontologies?

These issues point to the need for a “social layer” as part of the service interaction context. From an engineering perspective, new approaches are needed that take an holistic view of service environments, and take into account not only the properties of individual applications, but also the objectives, structure and dynamics of the system as a whole. In recent years, research in fields as diverse as social science, management science, economics, biology, distributed systems and multi-agent systems, analysed, modelled and explained a wide range of social phenomena often seen in human and animal societies and tried to apply those results to computational systems. In particular, techniques have been developed, that:

- Make it possible to characterise and model the organisational structures commonly used by humans to organise themselves in societies with particular needs;
- Capture coordination patterns that are often used between humans to solve common problems (e.g., to sell goods or achieve specific goals);

- Characterise autonomous actors in an environment and model their potential, rational behaviour (in order to predict, for example, how individuals will act in the context of a given set of “rules”).

The rest of this paper, which aims to describe how the ALIVE¹ architecture deals with the issues mentioned above, is organised as follows. Section 2 describes a detailed view of the architecture, while in section 3 some use cases are presented. The final section gives the conclusion and outlines directions for future research.

2 Proposed Architecture

The proposed ALIVE architecture combines *model driven development* (MDD) [13] with coordination and organisational mechanisms, providing support for *live* (that is, highly dynamic) and *open* systems of services. ALIVE’s approach extends current trends in engineering by defining three levels in the design and management of distributed systems: the Service Level, the Coordination Level and the Organisation Level, illustrated in Fig. 1 and explained below.

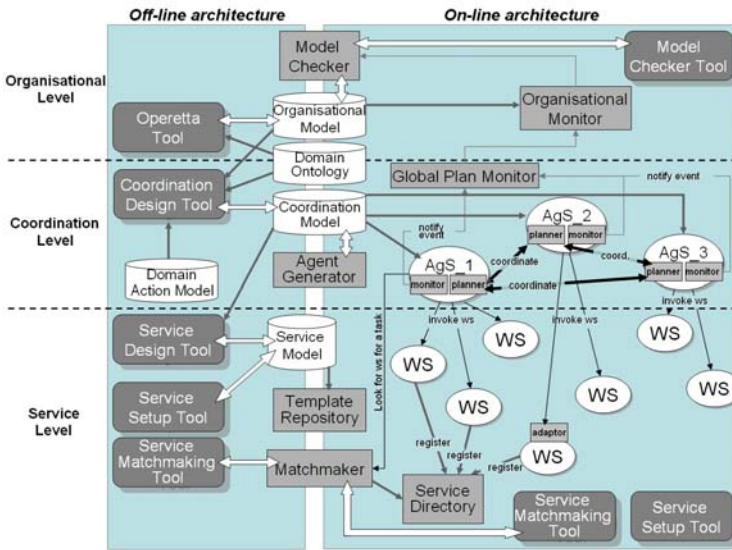


Fig. 1. ALIVE Multi-Level Architecture

The *Service Level* extends existing service models, in order to make components aware of their social context and of the rules of engagement with other components and services, by making use of semantic Web technologies. This “semantification” is particularly useful when highly dynamic and frequently changing services (the WSs in

¹ The ALIVE research project (<http://www.ist-alive.eu/>) is funded by the European Commission within the 7th Framework Programme for RTD (contract FP7 215890).

Fig. 1) are present in the system, as the meta-information in each service description (stored in a *service directory*) eases such processes as finding substitute services if a given service does not respond to invocations.

The *Coordination Level* provides the means to specify, at a high level, the patterns of interaction among services, using a variety of powerful coordination techniques from recent research in the area [49], based on agent technology. For these a special kind of services are introduced at this level: *agentified services*. These services are organisational-aware, that is, they are aware of the system objectives and manage task allocation and workflow generation and agreement. Also, at the coordination level agreed workflows can be adapted *while* the system is running – this is very useful when the system has to react to failures or exceptions (e.g., failing payment or booking systems).

The *Organisational Level* provides a social context for the Coordination and the Service levels, specifying the organisational rules that govern interaction and using recent developments in organisational dynamics [14] to allow the structural adaptation of systems over time. This is important when frequent changes of rules and restrictions are expected.

The ALIVE architecture can be seen as a service-oriented middleware supporting the combination, reorganisation and adaptation of services at both design- and run-time. These activities follow organisational patterns and adopt coordination techniques. Furthermore, the MDD paradigm integrated in the architecture offers significant assistance to system developers, as it provides semi-automated transformations between models of the three levels described above, as well as the capacity for multiple target platforms and representation languages. In the remaining part of this section, a more detailed description of the architecture is provided. Complementary theoretical aspects are described in [1] and related methodological aspects are described in [2].

2.1 Off-Line Support

The main objective of the off-line part of ALIVE's architecture is to give support not only to the design phase of distributed systems based on the ALIVE methodology (see [2]), but also to support the continuous maintenance of deployed systems during their lifetime.

The *Operetta Tool* is an organisational modelling tool. Its function is to create and manage the *organisational model* of a given distributed system. It is possible to connect Operetta to a domain ontology, use the entities of the ontology in the organisational model, and ensure the consistency of the entities across different models of the organisational level and a consistent connection between organisational model and coordination model. The functionality of this tool is extended by the Dynamic Organisational Modelling plug-in, for those applications which need extra methods and components to handle dynamic organisational change. One important aspect of the organisational models built by this tool is that they abstract away from the low-level details of the services that may be invoked. The designer is able to design the whole organisational level of a given distributed system through abstract concepts such as *objectives*, *roles*, *obligations*, *violations*, *sanctions* and high-level interaction diagrams that only identify critical states (called *landmarks*). The interaction with some end users (see section 3)

has shown that this abstract specification is easy to understand to non-software specialists, easing the validation of designs by end users.

The *Model Checker* is used to verify some properties of organisational models. E.g., if a role of manager depends on a role of salesman to get information then an interaction scene is defined where this information exchange is modelled. This component not only is used at design time but can also be used on-line during the execution of the system. The Model Checker Tool thus complements the OperettA Tool, and is linked directly to it through a button in the OperettA interface.

The *Coordination Design Tool* is used by designers and system administrators to create and manage the coordination model of a given application. The tool assists the user in the definition of actors and tasks, the generation of the agents that will perform the actual coordination tasks and the inspection of predefined and generated plans. It also supports importing and inspecting the task descriptions defined in an application *domain action model*. A distinctive trait of the coordination models created by this tool (in comparison with other orchestration and choreography technologies [519,1017]) is that the coordination models also abstract away from the low-level details of the services that may be invoked. The designer is able to design the whole coordination level of a distributed system by means of *actors, tasks, plans* and *plan coordination mechanisms*. The Agentified Services (see section 2.2) are the ones that connect, at execution time, the abstract tasks with the actual services that are invoked. Apart of the dynamism this solution brings at execution time, this also allows end users to inspect and better comprehend coordination models.

The *Service Design Tool* is used by designers and system administrators to generate or inspect service descriptions, edit service templates and register them in the *service directory*. It also connects with the *Service Matchmaking Tool* (a human interface to the matchmaker component), allowing designers and system administrators to search for services matching a given task description or implementing a given service template and registering it in the *service directory*.

Finally the *Service Setup Tool* allows designers and system administrators to check and modify the setup of the running environment, including the URIs of different resources, facilitating components, pre-defined services and service containers.

2.2 On-Line Support

The main objective of the on-line part of ALIVE's architecture is to give support to the deployment and execution of distributed systems. This part also covers on-line monitoring mechanisms to detect failures or deviations from expected functionality, and provides the mechanisms to correct, recover or adapt to them.

The *Service Directory* is one of the main components of the architecture. It is a repository for service interface descriptions, service process models and service profiles. It thus supports several query mechanisms for the discovery of specific services based on their syntactic and/or semantic descriptions. The service directory can be used by any component or service, and its main clients are the matchmaker components. Depending on the requirements and size of the distributed system, one or many service directories may be deployed. *Matchmakers* are an important component for service recognition. They are responsible for the discovery of appropriate services that fulfil the

requirements of a given task. Depending on the level of abstraction of the task to be fulfilled, matchmakers may query a service directory directly or by the use of service templates. The *Service Templates* are intermediary descriptions linking higher-level goals or tasks (as specified in the Coordination Model) with specific service interactions. A template includes a parameterised process model for a class of services in terms of pre- and post-conditions. If the described process is of some complexity, such description may include an abstract workflow fragment indicating required sub-steps in the process. Parameters in a template are specified as abstract types linked to the variables or parameters in the process model. The parameters are dynamically bound at execution time into concrete ontology instances and/or concrete service process models when the template is selected by a matchmaker. Service template definitions are stored in a *Template Repository*.

Agentified services are organisation-aware services able to coordinate with others according to a given organisation and coordination model. They are able to:

- Incorporate the description of an organisation role, including its objectives, rights, obligations and prohibitions;
- Build, at run-time, local plans to fulfill the role’s objectives;
- Coordinate its activities with other agentified services, thus building a partial global plan [8].

Agentified services can interact with normal Web services by means of standard SOAP and REST interfaces. Furthermore Agentified services communicate coordination-related issues to other agentified services using protocol-based conversations expressed in a coordination language (based on GPGP) implemented over SOAP. The exchanged plans are abstract workflows possibly with tasks referring to abstract services rather than to concrete ones (e.g., “map service” instead of “Google Maps”). When a plan is agreed upon, an agentified service will look (via the matchmaker component) for services that can fulfil the abstract tasks, binding them together.

External Services (i.e., existing, third-party services that have not been designed following the ALIVE framework) can be invoked at execution time according to their service description. Usually, external services are not consumed directly; instead, this is done via service adaptors. *Service adaptors* allow external services to be utilised for suitable organisational tasks. Typical examples of adaptation are type translation services to adapt a service interface to the entities and data types used by a given organisation.

One or several *Monitoring Components* are able to aggregate and analyse events related to the execution of services, the fulfilment of coordination plans and the achievements of role and/or organisational objectives. During the on-line execution, events are generated by the components (viz., the agentified services), whenever deviations, exceptions or failures are detected that cannot be handled by the agentified service itself or the existing coordination plan in place. In such situations the current organisational model is evaluated and then either (a) the objectives affected by the detected issue may be re-evaluated (their priority may be lowered or they may even be dropped completely), or (b) more significant changes in the organisation model may be required (for instance, changing the rights of a role). In case (a) the agent’s coordination modules will create a new plan based on the updated organisational objectives. In case (b) the updated

model is sent to the Agent Generator component to (re)generate the agentified services that populate the system. Depending on the set-up preferences of the administrator, the monitoring component may be a separate component used by several agentified services or may be a federation of several components inside the agentified services themselves.

Finally, there are graphical tools to support system administrators in the management of a distributed system. The *Monitor Tool* allows the administrator to inspect the status of a system's execution, getting information from the different components in the on-line architecture (especially from the monitoring components). In this way the administrator can keep track of the events generated at execution time and inspect how the system handles them. The *Service Set-up Tool* can be used by system administrators to check and modify the setup of the running environment. Finally the *Service Match-making Tool* (also used off-line) allows administrators to search manually for services that match a given task description (by using the matchmaker component) when the system is not able to cope with a particular issue or the administrator wants to change manually an automatically selected service for another that is considered more suitable for reasons not modelled within the ALIVE framework.

3 Application Scenarios

The ALIVE multi-level architecture is especially useful for scenarios where changes can occur at either abstract or concrete levels, and where services are expected to be continuously changing, with new services entering the system and existing services leaving it at run-time. For example, when there is a significant change at a high level (e.g., a change in the organisational structure), the service orchestration at lower levels can be automatically reorganised. Another example is the automatic adaptation of higher levels when lower ones suffer significant changes (e.g., the continuous failure in some low-level services). The proposed architecture is currently being applied to three real-life systems:

- *Dynamic orchestration of distributed services on interactive community displays* – The concept of *interactive community displays* (ICDs) is to build a virtual space within which people in urban communities can interact and share knowledge, experiences, and mutual interests. ICDs integrate urban information (in real time) and create public spaces in the Internet for people living in or visiting a city. ICDs are being developed all over the world in the context of digital cities [6]. An important question concerns why urban information spaces attract people given this era of globalisation. The Internet has triggered global businesses, and at the same time enables us to create rich information spaces for everyday life. While the Internet makes research and businesses global, life is inherently local. Business requires homogeneity to allow global competition, while life is heterogeneous, reflecting the different cultural backgrounds [7]. TMT Factory, one of the partners of the ALIVE project, is creating a personalised recommendation tool which provides city services to residents and tourists through ICDs (see Fig. 2). These ICDs, installed in urban environments, display tourist information and other services, dynamically personalised according to user preferences, geographical location and local laws. The ALIVE architecture provides support for the organisational modelling of the

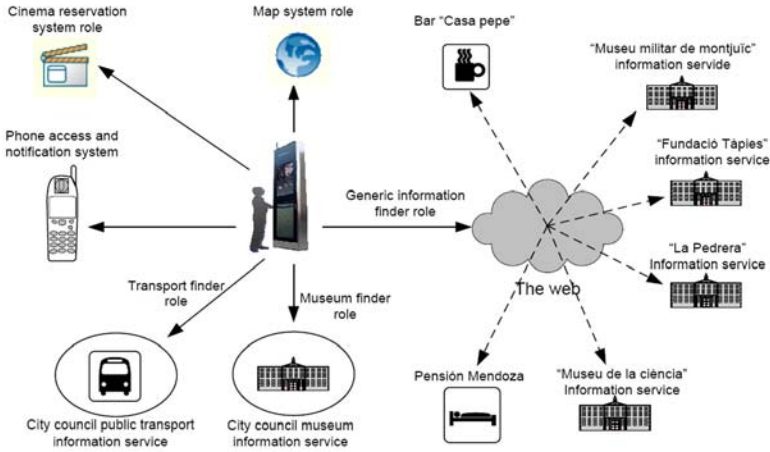


Fig. 2. Example of the ICD interaction with different organisational services, modelled by decomposing the system functionalities in terms of clearly-defined roles

different entities that offer services to users and dynamically adapts service composition in case of service failure.

- *Dynamic crisis management.* Simulations of crisis scenarios have the potential to improve the organizational structures and the quality of response of the parties involved. As a crisis evolves, organizational structures must be systematically updated to reflect the changes in the numbers of personnel and the seriousness of the crisis. Simulation allows the effectiveness of these organizational changes to be evaluated. In the real world, simulations are enacted using active personnel. However, such simulations are expensive, both in terms of the cost of execution and the cost of the time required for the emergency service personnel involved. THALES B. V., another partner of the ALIVE project, is creating an emergency simulation tool for the Dutch emergency forces. This simulator effectively explores diverse crisis management scenarios when a natural disaster has escalating city- and nation-wide consequences, allowing policy makers to evaluate current and potential emergency policies and procedures. In particular The ALIVE architecture provides rich ways to describe all stakeholders in different emergency scenarios, their roles, responsibilities and dependencies and suggest the coordination strategies that can emerge according to a given set of policies.
- *Communication in entertainment domains.* Calico Jack Ltd., another partner, is developing an Entertainment Communication Router (ECR), a system that exploits the richness of online entertainment systems by coupling them with social networking resources. The ECR is designed to address the fragmentation of communication resulting from an increasing diversity of channels. This fragmentation makes it difficult to optimise the delivery of messages within constraints imposed by communication norms and a dynamic context of preferences, presence and availability. The system manages the roles, identities and social relations that users have in different fora (e.g. Facebook) to support rich forms of communication both between players

and with their wider social environment. In this case the ALIVE architecture provides support for high-level context definition and management. This includes the description of the information routing rules that reflect the complex, dynamic and sometimes conflicting constraints that emerge from the users' context.

4 Conclusions

The ALIVE framework aims to support the design and development of distributed systems suitable for highly dynamic environments, based on model-driven engineering, and three interconnected levels: service, coordination and organisation.

The crucial distinction of the ALIVE approach from existing ones is that it provides an organisational context (such as, for instance, objectives, structures and regulations) that can be used to select, compose and invoke services dynamically. ALIVE also provides a notion of organisational awareness to some components (such as the agentified services at the Coordination Level or the matchmaker component at the Service Level) that can direct system execution in order to achieve higher-level organisational objectives. One of the effects is that exceptions can be managed not only at the lower level (as in other service-oriented architectures) but at higher levels, looking for alternative ways to fulfil a task or even a full abstract workflow by agreeing upon a new plan of action. Furthermore, organisational and coordination models are defined at a level of abstraction that allows non-expert end-users to support better the design and the maintenance of the system.

The first version of the ALIVE tool suite is now under development and will become available through the project's Sourceforge site².

Acknowledgements

This research has been carried out within the FP7-215890 ALIVE project, funded by the European Commission. Javier Vázquez-Salceda is partially funded by the Ramón y Cajal program of the Spanish Ministry of Education and Science.

References

1. ALIVE project: ALIVE Theoretical Framework, http://www.ist-alive.eu/index.php?option=com_docman&task=doc_download&gid=3&Itemid=49 (retrieved 20090527)
2. ALIVE project: ALIVE Methodology Document, http://www.ist-alive.eu/index.php?option=com_docman&task=doc_download&gid=2&Itemid=49 (retrieved 20090527)
3. DMTF: Web Services Management, <http://www.dmtf.org/standards/wsman/> (retrieved 20090527)
4. Ghijssen, M., Jansweijer, W., Wielinga, B.B.: Towards a Framework for Agent Coordination and Reorganization, AgentCoRe. In: Sichman, J.S., Padget, J., Ossowski, S., Noriega, P. (eds.) COIN 2007. LNCS (LNAI), vol. 4870, pp. 1–14. Springer, Heidelberg (2008)

² <http://sourceforge.net/projects/ict-alive/>

5. IBM: Business process execution language for web services version 1.1 (July 2003), <http://www.ibm.com/developerworks/library/specification/ws-bpel/> (retrieved 20090527)
6. Ishida, T. (ed.): Understanding Digital Cities: Cross-Cultural Perspectives. MIT Press, Cambridge (2000)
7. Ishida, T.: Digital City Kyoto. Communications of the ACM 45(7), 76–81 (2002)
8. Lesser, V.: Evolution of the GPGP/TAEMS domain-independent coordination framework. In: AAMAS 2002: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1–2. ACM, New York (2002)
9. Matskin, M., et al.: Enabling Web Services Composition with Software Agents. In: Proc. of the Conference on Internet and Multimedia Systems, and Applications. Honolulu (2005)
10. OASIS: The Framework for eBusiness (April 2006), <http://www.ebxml.org/> (retrieved 20090527)
11. OASIS: Universal Description, Discovery and Integration (UDDI) version 3.0.2. OASIS (2002), http://www.uddi.org/pubs/uddi_v3.htm (retrieved 20090527)
12. OGSi Alliance: OGSi Specifications, <http://www.osgi.org/Specifications/HomePage> (retrieved 20090527)
13. OMG: Model Driven Architecture, <http://www.omg.org/mda/> (retrieved 20090527)
14. van der Vecht, B., Dignum, F., Jules, J., Meyer, C., Dignum, V.: Organizations and Autonomous Agents: Bottom-up Dynamics of Coordination Mechanisms. In: 5th Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems. Estoril (2008)
15. W3C: Simple Object Access Protocol (SOAP) 1.1. W3C Note, <http://www.w3.org/TR/2000/NOTE-SOAP-20000508> (retrieved 20090527)
16. W3C: OWL-S - Semantic Markup for Web Services (2004), <http://www.w3.org/Submission/OWL-S/> (retrieved 20090527)
17. W3C: Web Service Choreography Description Language (WS-CDL) (2005), <http://www.w3.org/TR/ws-cdl-10/> (retrieved 20090527)
18. W3C: Web Service Description Language, WSDL (2001), <http://www.w3.org/TR/wsdl> (retrieved 20090527)
19. WfMC. XML Process Definition Language (XPDL). Document Number WfMC-TC-1025: Version 1.14 Document Status - Final (2005)
20. WSMO working group: Web Service Modeling Ontology. ESSI cluster, <http://www.wsmo.org/> (retrieved 20090527)

Author Index

- Altman, Eitan 130
Antoniadis, Panayotis 74
Avesani, Paolo 19
Azgomi, Mohammad Abdollahi 9
- Broser, Christian 35
- Ceccaroni, Luigi 175
Clarke, Siobhan 175
Conchione, Elia 149
Courcoubetis, Costas 74
Crespi, Noel 98
- Dignum, Frank 175
Divsalar, Mehrnoosh 9
- Ferronato, Pierfranco 27
Finnegan, Jason 27
Fritsch, Christoph 35
- Galli, Luca 64
Gmelch, Oliver 35
Guarneri, Renata 64
- Hormazábal, Nicolás 82
Hossain, Aliar 166
Hossain, Mohammad Kamal 166
Huhtamaki, Jukka 64
- Ion, Mihaela 45
- Kalogiros, Costas 74
Klingert, Sonja 90
Koralalage, Krishan Sabaragamu 53
- Laga, Nassim 98
Li, Man-Sze 156
Lopardo, Gabriel 82
- Malossini, Andrea 19
Mangiaracina, Riccardo 109
Minh Quan, Dang 90
Morales-Arroyo, Miguel A. 139
- Nieuwenhuis, Kees 175
- Padget, Julian 175
Pernul, Günther 35
- Rezaul, Karim Mohammed 166
Risch, Marcel 1
Rojas-Mora, Julio 130
Rosa, Josep Lluís de la 82
- Saint-Paul, Regis 45
Schillinger, Rolf 35
Sergeant, Paul 175
- Tan, Margaret 139
Tavoni, Alessandro 119
Telesca, Luigi 119
Terzidis, Orestis 90
Toletti, Giovanni 109
Turba, Luca 109
- Vasconcelos, Wamberto 175
Vázquez-Salceda, Javier 175
- Wiesbeck, Sandra 35
Wong, Sulan 130
- Yoshiura, Noriaki 53
- Zhao, Zhenzhen 98