

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Friedemann Mattern

*ETH Zurich, Switzerland*

John C. Mitchell

*Stanford University, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*University of Dortmund, Germany*

Madhu Sudan

*Massachusetts Institute of Technology, MA, USA*

Demetri Terzopoulos

*University of California, Los Angeles, CA, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Moshe Y. Vardi

*Rice University, Houston, TX, USA*

Gerhard Weikum

*Max-Planck Institute of Computer Science, Saarbruecken, Germany*

Simone Fischer-Hübner Steven Furnell  
Costas Lambrinoudakis (Eds.)

# Trust, Privacy, and Security in Digital Business

Third International Conference, TrustBus 2006  
Kraków, Poland, September 2006  
Proceedings

## Volume Editors

Simone Fischer-Hübner  
Karlstad University  
Department of Computer Science  
Universitetsgatan 2, 651 88 Karlstad, Sweden  
E-mail: simone.fischer-huebner@kau.se

Steven Furnell  
University of Plymouth  
School of Computing, Communications and Electronics  
Network Research Group, Plymouth, PL4 8AA, UK  
E-mail: sfurnell@plymouth.ac.uk

Costas Lambrinoudakis  
University of the Aegean  
Department of Information and Communication Systems Engineering  
Karlovasi, 83200 Samos, Greece  
E-mail: clam@aegean.gr

Library of Congress Control Number: 2006931261

CR Subject Classification (1998): K.4.4, K.4, K.6, E.3, C.2, D.4.6, J.1

LNCS Sublibrary: SL 4 – Security and Cryptology

ISSN	0302-9743
ISBN-10	3-540-37750-6 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-37750-4 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 is a part of Springer Science+Business Media  
springer.com

© Springer-Verlag Berlin Heidelberg 2006  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India  
Printed on acid-free paper      SPIN: 11824633      06/3142      5 4 3 2 1 0

## Preface

This book presents the proceedings of the Third International Conference on Trust, Privacy and Security in Digital Business (TrustBus 2006), held in Kraków, Poland, September 5-7, 2006. The conference continues from previous events held in Zaragoza (2004) and Copenhagen (2005), and maintains the aim of bringing together academic researchers and industry developers to discuss the state of the art in technology for establishing trust, privacy and security in digital business. We thank the attendees for coming to Kraków to participate and debate the new emerging advances in this area.

The conference programme included two keynote presentations, one panel session and eight technical papers sessions. The keynote speeches were delivered by Jeremy Ward from Symantec EMEA on the topic of “Building the Information Assurance Community of Purpose”, and by Günter Karjoth from IBM Research - Zurich, with a talk entitled “Privacy Practices and Economics — From Privacy Policies to Privacy SLAs.”

The subject of the panel discussion was “Is Security Without Trust Feasible?” chaired by Leszek T. Lilien from Western Michigan University, USA. The reviewed paper sessions covered a broad range of topics, from access control models to security and risk management, and from privacy and identity management to security protocols. The conference attracted 70 submissions, each of which was assigned to four referees for review. The Programme Committee ultimately accepted 24 papers for inclusion, which were revised based upon comments from their reviews.

We would like to express our thanks to the various people who assisted us in organizing the event and formulating the programme. We are very grateful to the Programme Committee members, and external reviewers, for their timely and rigorous reviews of the papers. Thanks are also due to the DEXA Organizing Committee for supporting our event, and in particular to Mrs. Gabriela Wagner for her help with the administrative aspects. We would also like to thank Sokratis Katsikas, Javier López and Günther Pernul for their past efforts in establishing the conference series, and their valuable advice and assistance in enabling us to take it forward.

Finally we would like to thank all of the authors who submitted papers for the event, and contributed to an interesting set of conference proceedings.

September 2006  
Kraków, Poland

Simone Fischer-Hübner, Karlstad University, Sweden  
Steven Furnell, University of Plymouth, UK  
Costas Lambrinoudakis, University of the Aegean, Greece

# Programme Committee

## General Chair

Simone Fischer-Hübner      Karlstad University, Sweden

## Programme Committee Co-chairs

Steven Furnell      University of Plymouth, UK  
Costas Lambrinoudakis      University of the Aegean, Greece

## International Programme Committee Members

Alessandro Acquisti	Carnegie Mellon University, USA
Marco Casassa Mont	HP Labs, Bristol, UK
David Chadwick	University of Kent, UK
Nathan Clarke	University of Plymouth, UK
Frederic Cuppens	ENST Bretagne, France
Ernesto Damiani	University of Milan, Italy
Ed Dawson	Queensland University of Technology, Australia
Claudia Eckert	Darmstadt Technical University, Germany
Hannes Federrath	University of Regensburg, Germany
Eduardo B. Fernandez	Florida Atlantic University, USA
Elena Ferrari	University of Insubria at Como, Italy
Juan M. González-Nieto	Queensland University of Technology, Australia
Rüdiger Grimm	University of Koblenz , Germany
Dimitrios Gritzalis	Athens University of Economics and Business, Greece
Stefanos Gritzalis	University of the Aegean, Greece
Ehud Gudes	Ben-Gurion University, Israel
Sigrid Gürgens	Fraunhofer Institute for Secure Information Technology, Germany
Marit Hansen	Independent Center for Privacy Protection, Germany
Audun Josang	School of Software Engineering & Data Communications, QUT, Australia
Tom Karygiannis	NIST, USA
Sokratis Katsikas	University of the Aegean, Greece
Dogan Kesdogan	RWTH Aachen University, Germany
Hiroaki Kikuchi	Tokai University, Japan

Spyros Kokolakis	University of the Aegean, Greece
Klaus Kursawe	Philips Research, Eindhoven, The Netherlands
Leszek Lilien	Western Michigan University, USA
Antonio Lioy	Politecnico di Torino, Italy
Javier López	University of Malaga, Spain
Peter Lory	University of Regensburg, Germany
Olivier Markowitch	Université Libre de Bruxelles, Belgium
Fabio Martinelli	National Research Council – CNR Pisa, Italy
Fabio Massacci	University of Trento, Italy
Jose A. Montenegro	University of Malaga, Spain
Eiji Okamoto	University of Tsukuba, Japan
Martin S. Olivier	University of Pretoria, South Africa
Rolf Oppliger	eSecurity Technologies, Switzerland
Maria Papadaki	University of Plymouth, UK
Ahmed Patel	Centre for Applied Research in Information Systems, Kingston University, UK
Günther Pernul	University of Regensburg, Germany
Andreas Pfitzmann	Dresden University of Technology, Germany
Hartmut Pohl	University of Applied Sciences, FH Bonn-Rhein-Sieg, Germany
Karl Posch	University of Technology, Graz, Austria
Torsten Priebe	Capgemini, Austria
Gerald Quirchmayr	University of Vienna, Austria
Kai Rannenber	Goethe University of Frankfurt, Germany
Christoph Ruland	University of Siegen, Germany
Pierangela Samarati	University of Milan, Italy
Matthias Schunter	IBM Zurich Research Lab., Switzerland
Mikko T. Siponen	University of Oulu, Finland
Adrian Spalka	University of Bonn, Germany
Leon Strous	De Nederlandsche Bank, The Netherlands
Stephanie Teufel	University of Fribourg, Switzerland
Jianying Zhou	I2R, Singapore

## External Reviewers

Isaac Agudo	University of Malaga, Spain
Manos Antonakakis	NIST, USA
Aimilios Apostolopoulos	NIST, USA
Giampaolo Bella	University of Catania, Italy
Rainer Böhme	Dresden University of Technology, Germany

Katrin Borcea-Pfitzmann	Dresden University of Technology, Germany
Colin Boyd	Queensland University of Technology, Australia
Andrew Clark	Queensland University of Technology, Australia
Sebastian Clauß	Dresden University of Technology, Germany
Nora Cuppens-Boulahia	ENST Bretagne, France
Wiebke Drespe	University of Regensburg, Germany
Ludwig Fuchs	University of Regensburg, Germany
Dimitris Geneiatakis	University of the Aegean, Greece
Juhani Heikka	University of Oulu, Finland
Christos Kalloniatis	University of the Aegean, Greece
Costas Karafasoulis	University of the Aegean, Greece
George Karopoulos	University of the Aegean, Greece
Maria Karyda	University of the Aegean, Greece
Tobias Koelsch	RWTH Aachen University, Germany
Stefan Köpsell	Dresden University of Technology, Germany
Hristo Koshutanski	Create-Net, Italy
Ponnurangam Kumaraguru	Carnegie Mellon University, USA
Dimitris Lekkas	University of the Aegean, Greece
Mink Martin	RWTH Aachen University, Germany
Patrick Sinclair Merten	University of Fribourg, Switzerland
Nicola Mezzetti	Università di Bologna, Italy
Björn Muschall	University of Regensburg, Germany
Andriy Panchenko	RWTH Aachen University, Germany
Lexi Pimenidis	RWTH Aachen University, Germany
Carsten Rudolph	Fraunhofer Institute for Secure Information Technology, Germany
Rolf Schillinger	University of Regensburg, Germany
Christian Schläger	University of Regensburg, Germany
Sandra Steinbrecher	Dresden University of Technology, Germany
Martin Steinert	University of Fribourg, Switzerland
Daniela Wanner	University of Fribourg, Switzerland
Andreas Westfeld	Dresden University of Technology, Germany
Nicola Zannone	University of Trento, Italy

# Table of Contents

## Session 1: Privacy and Identity Management

Towards Scalable Management of Privacy Obligations in Enterprises . . . . .	1
<i>Marco Casassa Mont</i>	
A New User-Centric Identity Management Infrastructure for Federated Systems . . . . .	11
<i>Vassilis Poursalidis, Christos Nikolaou</i>	

## Session 2: Security and Risk Management

Information Security Risk Assessment Model for Risk Management . . . . .	21
<i>Dariusz Waurzyniak</i>	
On the Limits of Cyber-Insurance . . . . .	31
<i>Rainer Böhme, Gaurav Kataria</i>	
Towards a Risk Management Perspective on AAI s . . . . .	41
<i>Christian Schläger, Thomas Nowey</i>	

## Session 3: Security Requirements and Development

Towards a UML 2.0 Extension for the Modeling of Security Requirements in Business Processes . . . . .	51
<i>Alfonso Rodríguez, Eduardo Fernández-Medina, Mario Piattini</i>	
A Framework for Exploiting Security Expertise in Application Development . . . . .	62
<i>Theodoros Balopoulos, Lazaros Gymnopoulos, Maria Karyda, Spyros Kokolakis, Stefanos Gritzalis, Sokratis Katsikas</i>	
On Diffusion and Confusion – Why Electronic Signatures Have Failed . . . . .	71
<i>Heiko Roßnagel</i>	

## Session 4: Privacy Enhancing Technologies and Privacy Management

Extending P3P to Facilitate Proxies Which Pose as a Potential Threat to Privacy . . . . .	81
<i>Wesley Brandi, Martin S. Olivier</i>	



A Systematic Approach to Privacy Enforcement and Policy Compliance Checking in Enterprises .....	91
<i>Marco Casassa Mont, Siani Pearson, Robert Thyne</i>	
A Generic Privacy Enhancing Technology for Pervasive Computing Environments .....	103
<i>Stelios Dritsas, John Tsaparas, Dimitris Gritzalis</i>	
Bringing the User Back into Control: A New Paradigm for Usability in Highly Dynamic Systems .....	114
<i>Sebastian Höhn</i>	

## Session 5: Access Control Models

Extending SQL to Allow the Active Usage of Purposes .....	123
<i>Wyndand van Staden, Martin S. Olivier</i>	
FGAC-QD: Fine-Grained Access Control Model Based on Query Decomposition Strategy .....	132
<i>Guoqiang Zhan, Zude Li, Xiaojun Ye, Jianmin Wang</i>	
A Framework for Modeling Restricted Delegation in Service Oriented Architecture .....	142
<i>Muhammad Alam, Michael Hafner, Ruth Breu, Stefan Unterthiner</i>	

## Session 6: Trust and Reputation

Reputation-Based Trust Systems for P2P Applications: Design Issues and Comparison Framework .....	152
<i>Eleni Koutrouli, Aphrodite Tsalgatidou</i>	
Towards Trust in Digital Rights Management Systems .....	162
<i>Jürgen Nützel, Anja Beyer</i>	
Cluster-Based Analysis and Recommendation of Sellers in Online Auctions .....	172
<i>Mikołaj Morzy, Juliusz Jezierski</i>	
Trust Model Architecture: Defining Prejudice by Learning .....	182
<i>M. Wojcik, J.H.P. Eloff, H.S. Venter</i>	

## Session 7: Security Protocols

How to Protect a Signature from Being Shown to a Third Party .....	192
<i>Marek Klonowski, Przemysław Kubiak, Mirosław Kutylowski, Anna Lauks</i>	

Security Analysis and Improvement for Key Issuing Schemes in ID-Based Cryptography .....	203
<i>Saeran Kwon, Sang-Ho Lee</i>	

A Secure E-Tender Submission Protocol .....	213
<i>Rong Du, Colin Boyd, Ernest Foo</i>	

## **Session 8: Security and Privacy in Mobile Environments**

A Sophisticated Solution for Revealing Attacks on Wireless LAN.....	223
<i>René Neumerkel, Stephan Groß</i>	

Information Leakage in Ubiquitous Voice-over-IP Communications .....	233
<i>Thorsten Neumann, Heiko Tillwick, Martin S. Olivier</i>	

<b>Author Index</b> .....	243
---------------------------	-----

# Towards Scalable Management of Privacy Obligations in Enterprises

Marco Casassa Mont

Hewlett-Packard Labs, Trusted Systems Lab  
Bristol, UK  
marco.casassa-mont@hp.com

**Abstract.** Privacy management is important for enterprises that collect, store, access and disclose personal data. Among other things, the management of privacy includes dealing with privacy obligations that dictate duties and expectations an enterprise has to comply with, in terms of data retention, deletion, notice requirements, etc. This is still a green area open to research and innovation: it is about enabling privacy-aware information lifecycle management. This paper provides an overview of the work we have done in this space: definition of an obligation management model and a related framework; implementation of a prototype of an obligation management system integrated both in the context of the PRIME project and with an HP identity management solution. This paper then focuses on an important open issue: how to make our approach scalable, in case large amounts of personal data have to be managed. Thanks to our integration work and the feedback we received, we learnt more about how users and enterprises are likely to deal with privacy obligations. We describe these findings and how to leverage them. Specifically, in the final part of this paper we introduce and discuss the concepts of parametric obligation and “hybrid” obligation management and how this can improve the scalability and flexibility of our system. Our work is in progress. Further research and development is going to be done in the context of the PRIME project and an HP Labs project.

## 1 Introduction

Enterprises that store, manage and process personal data must comply with privacy laws and satisfy people’s expectations on how their personal data should be used. Privacy laws [1,2,3] dictate policies on how personal data should be collected, accessed and disclosed according to stated purposes, by keeping into account the consent given by data subjects (e.g. customers, employees, business partners) and by satisfying related *privacy obligations* including data retention, data deletion, notice requirements, etc.

The management and enforcement of privacy policies in enterprises is still a green field: key requirements include automation, cost reduction, simplification, compliance checking and integration with existing enterprise identity management solutions. In particular the management of *privacy obligations* is open to research and innovation. Privacy obligations [4] dictate duties and expectations on how personal data should be

managed. They require enterprises to put in place *privacy-aware information lifecycle management* processes.

During the last two years we have been active in the *privacy obligation management* [5] space by: (1) researching and defining an explicit model for privacy obligations; (2) formalising the representation of obligations; (3) introducing an obligation management framework and a related *obligation management system* to deal with the explicit scheduling, enforcement and monitoring of privacy obligations.

This paper provides an overview of the current status of this work. Our current obligation management system allows end-user to customise - in a fine-grained way - their personal preferences: related privacy obligations (based on the set of obligations supported by an enterprise) are automatically generated and associated to users' data. However, this causes scalability issues when large sets of personal data have to be managed, because our system generates a large set of associated privacy obligations: their current management is expensive and inefficient. Addressing this aspect is very important for enterprises that potentially have to deal with millions of data records related to customers, employees or business partners.

The integration phase of our work in PRIME [6] and with an HP identity management solution [8, 12] and the feedback we received from third parties (customers, HP businesses, etc.) has helped us to better understand how users are actually likely to define their privacy preferences and which realistic support enterprises can provide in terms of handling privacy obligations. We describe these findings and highlight how they can actually be leveraged to address the scalability issues. The final part of this paper describes our related ideas, based on the concept of *parametric obligations* and a *hybrid obligation management model*. This work is in progress and will be carried on in the context of PRIME and an HP Labs project.

## 2 Management of Privacy Obligations in Enterprises

This section provides a quick overview of our R&D work to manage privacy obligations in enterprises. Details can be found in [4,5,9].

Privacy obligations [4,5,9] are policies that dictate expectations and duties to enterprises on how to handle personal data and how to deal with its lifecycle management in a privacy-aware way. They include: dealing with data deletion and retention, dealing with data transformation (e.g. encryption), sending notifications, executing workflows involving human and system interactions, logging information, etc.

Related work includes EPAL [10] that defines a privacy language, inclusive of a placeholder for obligations, in the context of an Enterprise Privacy Authorisation architecture [11]. This is important work but it does not define obligation policies in detail and subordinate their enforcement to access control. Similar observations apply for XACML [8] and other work in the obligation management space.

In our vision the management and enforcement of privacy obligations must not be subordinated to the management and enforcement of access control policies [4]. For example, deletion of personal data at a precise point in time has to happen independently from the fact that this data has ever been accessed. This fundamental concept is at the very base of our work and differentiates it from related work. A more detailed comparison of our work against related work is provided in [4,5,9].

Based on this concept, we introduced an *obligation management model* [4,5,9], where privacy obligations are “first class” entities, i.e. they are explicit entities that are modeled, managed and enforced. In this model, a privacy obligation is an “object” [9] that includes: *Obligation Identifier*; *Targeted Personal Data* (e.g. data affected by the obligation); *Triggering Events* (e.g. time-based events); *Actions* (e.g. data deletion, sending notifications) – see Figure 1.

```

<obligation ObligationId="OBLID1">
  <target // Reference to the PII Data the obligation is associated to
    <data repository>databaseA </data repository>
    <data structure type=TABLE> CustomerTable </data structure>
    <data attr="ALL" @key:UserId:PSEUDO1 </data>
  </target>
  <events operator="&">
    <event id="e1">
      <type>TIMEOUT</type>
      <date now="no"> 2007/10/13 14:01:00 </date>
    </event>
  </events>
  <actions>
    <action id="a1">
      <type>DELETE</type>
      <data attr="part">
        <item> // Reference to the PII Data attribute
          @key:UserId:PSEUDO1|att:CreditCard
        </item>
      </data>
    </action>
    <action id="a2">
      <type>NOTIFY</type>
      <method>EMAIL</method>
      <to> // Reference to the PII Data attribute
        @key:UserId:PSEUDO1|att:E-Mail
      </to>
    </action>
  </actions>
</obligation>

```

**Fig. 1.** Simple Example of Privacy Obligation

Figure 1 shows a very simple example of a privacy obligation (expressed in XML), associated to the personal data of a user (in the example having the *PSEUDO1* unique identifier) and stored in an enterprise RDBMS database. This obligation dictates the deletion of a personal attribute (credit card detail) at a predefined point in time, along with the need to notify the user via e-mail when this happens.

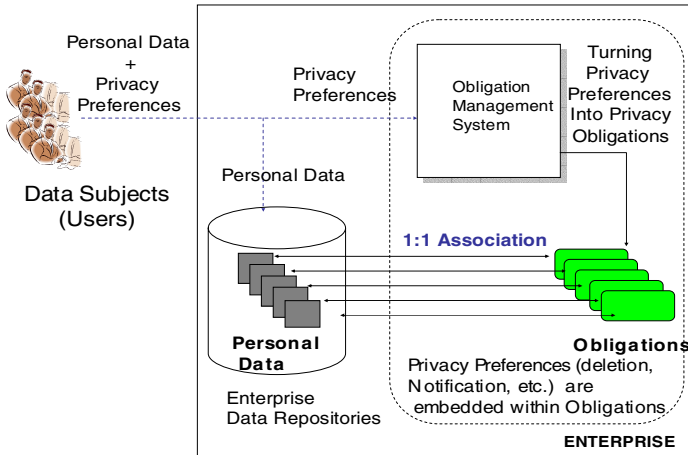
In general, our privacy obligations can target personal data stored in various types of data repositories, including databases, LDAP directories, meta/virtual directories, file systems, etc. This further differentiates our work and approach from related work, that is mainly focused on the management of data in RDBMS databases, e.g. [13].

We designed an *obligation management framework* [4,5,9] and an associated *obligation management system* [4,5,9] to represent these privacy obligations, schedule and enforce them and monitor for their fulfillment. In our system, *data subjects* (i.e. users) can explicitly define privacy preferences (e.g. on data deletion, notifications, etc.) on their personal data at their disclosure time (e.g. during a self-registration process) or at any subsequent time. These preferences are automatically turned into privacy obligations, based on the types of obligations supported by an enterprise. *Enterprise privacy administrators* can further associate other privacy obligations to personal data, for example dictated by laws or internal guidelines.

As a proof-of-concept, a working prototype has been fully implemented and integrated in the context of the EU PRIME project [6]. To demonstrate the feasibility and applicability of this work within enterprises, we also integrated it with HP OpenView Select Identity (an HP state-of-the-art identity management solution [7]) to manage privacy preferences and related privacy obligations during user provisioning and account management processes.

### 3 Scalability Issues

Our obligation management system provides flexible, fine-grained mechanisms to end-users (and enterprise privacy administrators) to express their privacy preferences (e.g. deletion preferences, notification preferences, etc.) on their personal data: based on the types of obligations supported by an enterprise, our system automatically turns these preferences into privacy obligations (by means of translation rules) and manages them. Users have the capability to customize aspects of these obligations (e.g. actual combinations of events and actions) as long as they are supported by the enterprise. The side-effect of this flexibility (at least in the current implementation) is that for each piece of personal data disclosed by a user, one or more privacy obligations can be generated, each of them with its own specific properties and requirements. For example, each user of an e-commerce site could potentially specify different privacy preferences (e.g. deletion date, notification preferences, encryption of data, data minimisation, etc.) and privacy constraints (among the ones supported by the enterprise) on their personal data. Figure 2 shows this approach (architectural details are omitted for simplicity).



**Fig. 2.** Current Model: Direct Association of Privacy Obligations to Personal Data

In case large amounts of users are managed by the enterprise, large amounts of privacy obligations are created and subsequently they must be explicitly scheduled, enforced and monitored by our obligation management system. In general, the

number of managed privacy obligations *linearly grows* with the number of managed users. Despite the fact that the components of our system can be replicated and distributed [9], the overhead of managing large amounts of obligations could be overwhelming, both in terms of computation and in terms of human-based administration.

Related to the latter aspect, the current GUI administrative tools [9] to manage privacy obligations within enterprises can potentially display all the managed privacy obligations along with their current status (to be enforced, enforced & compliant, enforced & violated, etc.). These GUI tools already allow administrators to focus on sub-set of managed obligations, based on some of their properties. However, in case of large amounts of managed privacy obligations, the task of selecting the relevant privacy obligations or having an overall view of the status of monitored obligations could be difficult to achieve.

To summarise, addressing the scalability problem requires to: (1) deal with large amount of personal data (potentially millions of records) and related privacy obligations; (2) do it in efficient and practically usable way; (3) provide adequate administration and obligation lifecycle management capabilities.

These issues were known at the design time of our current prototype: however more urgent and preliminary work was required to research the very concept and properties of privacy obligations. Our first prototype was meant to demonstrate the feasibility of our approach and use it as a starting point to make further experiments.

## 4 Towards Scalable Management of Privacy Obligations

As described in the previous section, the main cause of the current scalability problem is that our obligation management system generates one or more privacy obligations for each piece of personal data that is disclosed: these obligations can potentially be different in their structure and declared constraints (as long as based on the types of obligations supported by an enterprise). We learnt a few lessons by integrating our system in PRIME and with the HP identity management solution and by getting related feedback. This has provided us with more insights and ideas on how to address the scalability problem – in a way we can leverage and extend our current work. Next sections provide more details.

### 4.1 Learnt Lessons

Our obligation management system has been integrated with the PRIME system [6] to provide a comprehensive privacy-enhanced identity management solution both at the user-side and the enterprise-side. At the integration time, it has been clear that it would have not been feasible for the enterprise to support users in defining any arbitrary combination of privacy preferences and constraint specifications, even within the types of obligations that an enterprise potentially supports (e.g. by allowing any possible combinations of related *events* and *actions*). This because of the involved costs, the complexity of developing a general purpose solution and usability aspects for users (e.g. authoring their privacy obligations).

We have learnt that it would be preferable to explicitly provide users with a list of predefined “types” of privacy obligations supported by an enterprise (for given types

of personal data to be disclosed) where these obligation types have *predefined structures* (i.e. predefined combination of events and actions). Each type of privacy obligation clearly states which relevant *privacy preferences* a user can specify (e.g. data deletion time, notification preference, etc.).

In the integrated PRIME system [6], an enterprise describes these “types” of privacy obligations by means of “*Obligation Templates*”. An “*Obligation Template*” is graphically rendered to users at the time they have to disclose their personal data. In doing this, users can intuitively instantiate their related privacy preferences without being exposed to the internal representation of privacy obligations. Figure 3 shows a simple example of *Obligation Template*: preferences and information that need to be instantiated are expressed in the template with the “[?]” notation.

```
<obligation ObligationId="OBLID1">
  <target // Reference to the PII Data the obligation is associated to
    <data repository>databaseA </data repository>
    <data structure type=TABLE> CustomerTable </data structure>
    <data attr="ALL" @key:UserId:[?] </data>
  </target>
  <events operator="">
    <event id="e1">
      <type>TIMEOUT</type>
      <date now="no"> [?] </date>
    </event>
  </events>
  <actions>
    <action id="a1">
      <type>DELETE</type>
      <data attr="part">
        <item> // Reference to the PII Data attribute
          @key:UserId:[?]@att:CreditCard
        </item>
      </data>
    </action>
    <action id="a2">
      <type>NOTIFY</type>
      <method>EMAIL</method>
      <to> // Reference to the PII Data attribute
        @key:UserId:[?]@att:E-Mail
      </to>
    </action>
  </actions>
</obligation>
```

**Fig. 3.** Simple Example of Obligation Template

Once privacy obligations have been instantiated (with the relevant privacy preferences) they are processed by our obligation management system as described in section 2. For example, the instantiation of the *Obligation Template* in Figure 3 is going to be a privacy obligation similar to the one shown in Figure 1.

This approach to “*predefine and standardise*” the types of managed obligations is also consistent with: (1) the feedback we received by customers, HP business divisions and third parties; (2) our experience in integrating our system with the HP identity management solution. In these cases the main drivers were simplification of the overall specification and management processes, both for the enterprise and users.

By using this approach, obligations derived from a predefined “type” (obligation template) have the same structure (i.e. the same template, describing the same combinations of events and actions): the only aspects that differentiate them are the privacy



preferences provided by end-users. These preferences are *embedded* within these obligations. Of course, in case of large amounts of personal data (of related users), our obligation management system still has to generate and deal with a large number of distinct privacy obligations – hence again the scalability issue.

At this point, however, we realised that each set of *structurally identical* obligations requires the same type of management, enforcement and monitoring: as such, each set can be represented by just *an abstract obligation* that is *parametric* to the related data targets and privacy preferences expressed by users. This introduced the concept of *parametric privacy obligation*: its properties and the implication for our obligation management model are described in the next section.

## 4.2 Model of Parametric Privacy Obligations and Hybrid Obligation Management

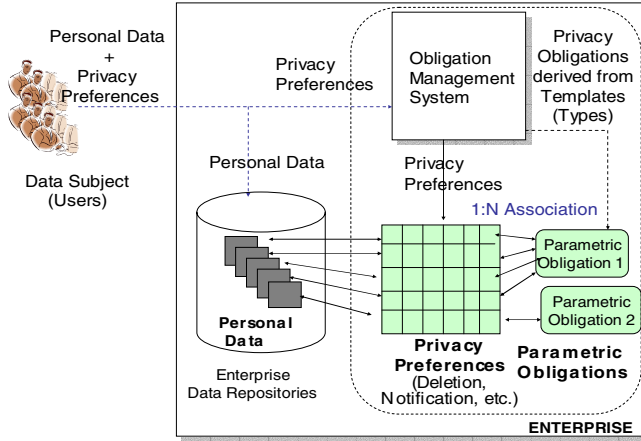
This section describes our current thoughts and ideas on how to address the scalability issues by leveraging the concept of *parametric obligation*. In particular we aim at minimising the set of privacy obligations to be managed by our obligation management system.

A *parametric obligation* is an obligation containing a *parametric* definition of its sub-components, i.e. *Target*, *Events* and *Actions*. Its structure is still based on *obligation templates* defined by enterprise privacy administrators. However, instead of containing explicit values within its Target, Events and Actions (as it happens in our current obligations - e.g. a unique user identifier and a few preferences – such as deletion time), a parametric obligation contains *references* to these values. In particular, the Target refers to the (potentially large) set of data the obligation is associated to.

In this context, privacy preferences *are not anymore embedded* within obligations (as it happens in the current system). These privacy preferences are still managed by the obligation management system but they are stored in a separated, explicit *data structure* (e.g. database tables in a relational database) – referred in this paper as “*Privacy Preferences*” *data structure* - along with a reference to the personal data they are associated to. Because the target of a parametric obligation *refers to a set of personal data* and its events/actions *contain references to related privacy preferences*, this obligation can now dictate how to handle and manage this entire set of personal data. Hence, in this model, a *parametric obligation* can be associated to a *set of personal data* and related privacy preferences - as shown in Figure 4.

As a consequence, a parametric obligation can be used to replace a potentially large set of “traditional” privacy obligations (i.e. the ones used in our current system) as long as they are based on the same “obligation template”. In this new model, each parametric obligation dictates an identical set of duties and expectations to be fulfilled on a set of personal data, individually customised by associated privacy preferences.

As a result, the set of parametric obligations is now reasonably small, depending on the different types of obligations that are explicitly managed by the enterprise. In other words, given a predefined set of obligation types (i.e. obligation templates), the obligation management system will have to manage a correspondent set of parametric obligations. As these sets are meant to be small, this is a step towards addressing the scalability problem.



**Fig. 4.** Association of Parametric Obligations to Personal Data

It is beyond the scope of this paper to describe the actual (XML) representation of parametric obligations. Further research and work has to be done in this area.

Of course, the current obligation management system needs to be extended to deal with these parametric obligations. For each managed parametric obligation it has to: (1) understand what the targeted set of personal data and related preferences are; (2) capture and manage the events that are relevant to all this data; (3) check if any of these events can trigger the execution of specific actions. If so, execute these actions and monitor them.

On one hand, this *extended obligation management system* will have only to manage a small number of (parametric) obligations. On the other hand, however, each parametric obligation could be associated to a potentially large set of personal data along with their related preferences. For each piece of personal data, this system must remember relevant “operational” information (related to associated parametric obligations), such as the local status of the events that might trigger the execution of actions. In case of composite events [9] (including stateful events, such as access counters) additional intermediate information must be stored. This can be done in additional data structures managed by the obligation management system.

Despite the fact that the management of events and actions might relate to a potentially large amount of data, we believe that these operations can now be optimised by using appropriate data structures and ways to manipulate this data via standard data access mechanisms. For example indexed tables could be used within relational databases to store the relevant information (privacy preferences and auxiliary data) and (optimised) SQL queries used to make inferences, extract and update the relevant information. Research is in progress on these aspects.

“Traditional” privacy obligations and parametric obligations can coexist in the same obligation management system: this introduces a *hybrid model and framework* to manage privacy obligations. This model provides users and enterprises with a comprehensive and flexible approach that can adapt to varying needs and requirements. In case large amounts of personal data need to be handled, the support for parametric

obligations allows enterprises to deal with scalability issues by minimising the number of managed obligations - by predefining and fixing the structures of supported parametric obligations. Nevertheless in those cases where more flexibility and customisation is required, this will still be supported and managed by the system.

Hence, depending on the context and requirements, a mixture of the two capabilities can be provided to address at the best needs for scalability, flexibility and customisation.

## 5 Discussion and Next Steps

We believe that the proposed model does not limit the control that users have in specifying their privacy preferences: it actually makes the overall process more effective by (1) allowing enterprises to explicitly declare upfront which types of privacy obligations they can support and (2) letting users make their informed decisions.

Work needs to be done to better understand how to provide suitable administrative and GUI tools to manage parametric obligations and their overall lifecycle (creation, update and disposal). Current GUI tools allow administrators to administer one-by-one every instantiated privacy obligation, by displaying their properties and current status (to be enforced, enforced & satisfied, enforced & violated). In case of parametric obligations this capability has to be extended, as a parametric obligation can potentially refer to a large set of personal data (and related preferences): for each piece of personal data the properties and status of the parametric obligation could be different. We are currently investigating how to provide incremental details on managed parametric obligations via graphical tools that can drill-down the relevant information.

Our next steps involve further research to refine our model of parametric obligations and extend our obligation management framework. This includes: formalizing the format of parametric obligations; designing the engine that processes these obligations; ensuring that our system evolves towards a hybrid system that can support both “traditional” obligations and parametric ones; explore in more details the lifecycle management of parametric obligations. We plan to do this work in the context of the PRIME project and an HP Labs project. We are also planning to get further feedback and input by engaging in technological trials with customers.

## 6 Conclusions

Privacy management is important for enterprises that handle personal data, in particular the management and enforcement of related privacy obligations. This paper provides an overview of our R&D work done in this space to explicitly represent, schedule, enforce and monitor privacy obligations. Our prototype of an obligation management system and its integration with both the PRIME system and the HP identity management solution show the feasibility of our approach: this also helped us to further understand this space and highlight a potential scalability problem that occurs when large amounts of personal data have to be processed. In this context, our current system generates a large amount of privacy obligations with a consequent management overhead.

Based on this and learnt lessons, we introduced the concept of parametric obligations as a way to drastically reduce the number of managed obligations and allow the obligation management system to scale. We described our current thoughts on parametric obligations and its implications on our current obligation management model. Our work is in progress. Further research and development is going to be done in the context of PRIME and an HP Labs project.

## References

1. Rotemberg, M., Laurant, C.: Privacy International: Privacy and Human Rights 2004: an International Survey of Privacy Laws and Developments, Electronic Privacy Information Center (EPIC), Privacy International. <http://www.privacyinternational.org/survey/phr2004/>, 2004
2. OECD: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. <http://www1.oecd.org/publications/e-book/9302011E.PDF>, 1980
3. Online Privacy Alliance: Guidelines for Online Privacy Policies. <http://www.privacyalliance.org/>, Online Privacy Alliance, 2004
4. Casassa Mont, M.: Dealing with Privacy Obligations: Important Aspects and Technical Approaches, TrustBus 2004, 2004
5. Casassa Mont, M.: Dealing with Privacy Obligations in Enterprises, HPL-2004-109, 2004
6. PRIME Project: Privacy and Identity Management for Europe, European RTD Integrated Project under the FP6/IST Programme, <http://www.prime-project.eu/>, 2006
7. Hewlett-Packard (HP): HP OpenView Select Identity: Overview and Features, <http://www.openview.hp.com/products/slctid/index.html>, 2005
8. OASIS: Extensible Access Control Markup Language (XACML) 2.0, [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=xacml](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml), 2005
9. Casassa Mont, M.: A System to Handle Privacy Obligations in Enterprises, HP Labs Technical Report, HPL-2005-180, 2005
10. IBM: The Enterprise Privacy Authorization Language (EPAL), EPAL 1.2 specification. <http://www.zurich.ibm.com/security/enterprise-privacy/epal/>, IBM, 2004
11. Karjoth, G., Schunter, M.: A Privacy Policy Model for Enterprises. IBM Research, Zurich. 15th IEEE Computer Foundations Workshop, 2002
12. Casassa Mont, M., Thyne, R., Chan, K., Bramhall, P.: Extending HP Identity Management Solutions to Enforce Privacy Policies and Obligations for Regulatory Compliance by Enterprises - HPL-2005-110, 2005
13. Agrawal, R., Kiernan, J., Srikant, R., Xu, Y.: Hippocratic Databases, <http://www.almaden.ibm.com/cs/people/srikant/papers/vldb02.pdf>, IBM Almaden Research Center , 2002

# A New User-Centric Identity Management Infrastructure for Federated Systems

Vassilis Poursalidis and Christos Nikolaou

Computer Science Department, University of Crete, Heraklion Crete, Greece  
{poursal, nikolau}@csd.uoc.gr

**Abstract.** In today's Information Systems, users present credentials with local significance, to be authenticated and gain access to internal functionality. Users have different login-password combinations for each online service, or even different credentials for different roles within a service. As a result they tend to make poor password choices that are easy to remember, or even repeat the same login-password information on different services. This poses security threats to service providers and a privacy risk for end-users. The solution is to shift to identity management systems. Such a system will issue a digital identity for every user and will be able to control the full life-cycle of these identities, from creation to termination. Another aspect of such a system is the single sign-on mechanism, whereby a single action of user authentication and authorization can permit the user to access multiple services. The benefits are improved security, accountability and privacy protection.

## 1 Introduction

The identification process allows services to authenticate users, granting them access to internal service functionality thus providing a personalized experience. But these credentials have local significance, meaning that they are only eligible for the originating service. The problem of multiple credentials is expected to become even more challenging for end-users partly because of the growth of online services, and partly because of the increasing need to utilize remote resources and services. Beyond the basic problem of users having trouble remembering multiple username and password combinations, current technology presents security risks and administrative costs to service providers. In fact, current technology drives most users into creating accounts using poor and easy to remember passwords, or even to use the same login-password combination across multiple services. On the other hand, the security conscious user will choose different passwords for every different account. The accounts held by the first group of users will be easily breached, posing security threats to service providers and exposing the privacy of account holders. The second group will increase administrative costs, due to forgotten passwords.

The inability of existing solutions to seamlessly authenticate users has drawn attention to digital identity management systems. Such systems address the problem of multiple credentials by incorporating a Single Sign-On (SSO) mechanism. The mechanism is a single point where users perform a login procedure once. If the

outcome is successful a set of assertions is returned, which can be presented to services and gain immediate access. The assertions certify the identity of the holder to the service, preventing it from issuing custom credentials. Besides the SSO mechanism, identity management systems also control the full life cycle of digital identities, from creation to termination. Identity management systems also allow users to create as many pseudonyms as they like, using a single digital identity. This eliminates the possibility of linking certain actions to the digital identity that executed them, without the express permission from the authority that issued the digital identity of the user.

Our digital *Identity Management Infrastructure (IMI)* is a multi-pseudonym identity system, shifting the control and creation of the plethora of pseudonyms to end-users. Pseudonyms are secondary identities, derived from an identifier (master identifier) that was supplied from the authority that issued the digital identity. The advantage of our scheme is twofold. On the one hand users are able to preserve their anonymity, as experienced today. On the other hand services are certain that secondary identities belong to a physical person, avoiding the custom development of authentication methods. Also, services are certain that if legal issues arise with one of the secondary identities the master identifier and the actual user can be traced back, with the help of the authority that issued his digital identity. Our approach also protects the privacy of the user, by preventing the existence of a single point where multiple digital identities are held. This single point could become a target for potential attackers and result in mass identity exposure.

The rest of this paper is organized in the following way: section 2 clarifies the terminology used, and briefly describes prior art. Section 3 describes our infrastructure specification, its components and the communication patterns between them, for the provision of the desired functionality. Finally, section 4 presents directions for future work.

## 2 Related Work

Identity is a collection of characteristics which are either inherent or are assigned by another [7]. A digital identity comprises electronic records that represent network principals, including people, machines, and services [5], [10]. To be able to create, maintain and use digital identities the deployment of a digital identity management system is required. This infrastructure uses identities in the process of authentication and maps identifiers to the information needed for identification and authorization [7], [8]. The functionality described sets the basis for SSO solutions, where a system attempts to capture identification and authentication information once, and provide it to services accessed by a user automatically. However, a unique set of credentials that can identify us presents major privacy threats. To address the problem one can use pseudonyms, where the user has the ability to prove the identity without revealing oneself. Pseudonymity combines many of the advantages of having a known identity with the advantages of anonymity. The main difference between anonymity and pseudonymity is that while in anonymity the identity is not known, in pseudonymity, there exists a separate persistent “virtual” identity which cannot be linked to a physical person [5], [9]. The unique digital identity issued is simply used to create

multiple and dependable secondary identities that can be used in different services, preserving the users privacy but still holding them accountable [11]. The most active field in identity management systems is identity federation. The idea is that multiple organizations form a federation and authentication tokens from one organization in the federation are considered valid to the remaining of the organizations.

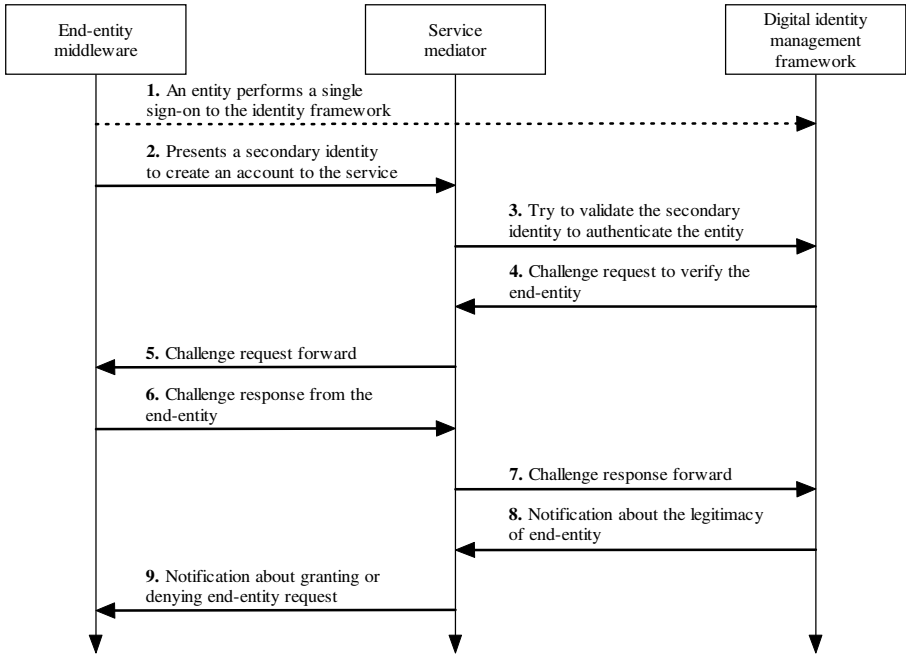
IBM, Microsoft, RSA Security, BEA Systems and VeriSign are working on a set of specifications known as the WS-Roadmap. Our interest focuses on the WS-Federation [1] specifications, where mechanisms are defined to allow different realms or domains to federate by allowing and brokering trust of identities, attributes and authentication between participating Web Services. Similar work is performed by Liberty Alliance, a consortium of 150 companies and organizations. Its purpose is to develop open standards in federated identity management that supports all current and emerging network devices [2]. Our research work is very similar to work presented at [3] and [4], which indicates the strong interest of both the research and the industry community towards our approach.

### 3 Our Approach: The Identity Management Infrastructure

Our infrastructure defines a set of entities and the interaction between them to provide a global scale single sign-on system. The software modules that we have deployed can be grouped into a *Digital Identity Management (DIM)* framework, the *service mediator* and the *end-entity middleware*. The DIM framework creates and revokes identities, and provides to services proof of the legitimacy of the end-entities. The service mediator handles the authentication of the entity's identity, outsourcing the burden of accounting from the digital identity management framework. The outsourcing also benefits the service in terms of mapping an identity to a real world entity, since the digital identity management framework assures the existence of a real world entity behind every digital identity.

The end-entity middleware preserves the user's privacy by creating different secondary identities for each service and its anonymity by creating different secondary identities for the same service. By creating secondary identities that cannot be correlated and at the same time can be validated by the DIM framework, the middleware achieves both strong validation and end-entity protection. The content of a secondary identity is just a unique identifier, a partial identity containing the number of this secondary identity and the issuer of the identity. By shifting the creation and management of secondary identities to the end-entity middleware, the user is responsible for the information that is shared amongst services and of the extent of information given. The main advantage from moving this information to the edges of our infrastructure is the absence of a single repository containing a list of user profiles thus protecting users from exposure of sensitive information.

The next figure [Figure 1], presents the basic steps for performing SSO and subsequently using the DIM framework to authenticate users to services. While most of the details about the exchange of messages are hidden, for clarity reasons, the basic idea remains the same: we use services without the requirement of creating a separate account for each service.



**Fig. 1.** Sequence of interactions to perform a single sign-on and authenticate secondary identities without using another set of credentials or end-entity interaction

### 3.1 DIM Framework

One could think of a centralized approach, where a single authority or pool of authorities is responsible for creating and managing digital identities. The main advantage of this system is that it operates under a single administration domain, so centralized security measures can be taken to minimize the possibility of a successful attack. Services also benefit as they have to trust a single system for providing legitimate user information. Users are given a set of credentials in order to activate their account on the system; no sensitive information is handled by users, who are considered the security weak link. But this approach defines a single point of failure, or under heavy load a bottleneck. There are also privacy concerns for a single system vouching digital identities for large number of users.

Another approach would be to define several independent identity management authorities. In that case, the amount of information available to every authority will decrease and continue on decreasing as the number of authorities grows. The advantage here is that every authority has a smaller number of users to support and less sensitive information in case of a security breach. Moreover the users can select the authority that they want to handle their digital identity information. This will drive the authorities to be competitive and apply stronger security policies, further benefiting the user. The drawback is that services must know all available independent



authorities, as users' credentials can originate from any of the available authorities. Another problem is that some authorities may misbehave, issuing false credentials and creating problems for services.

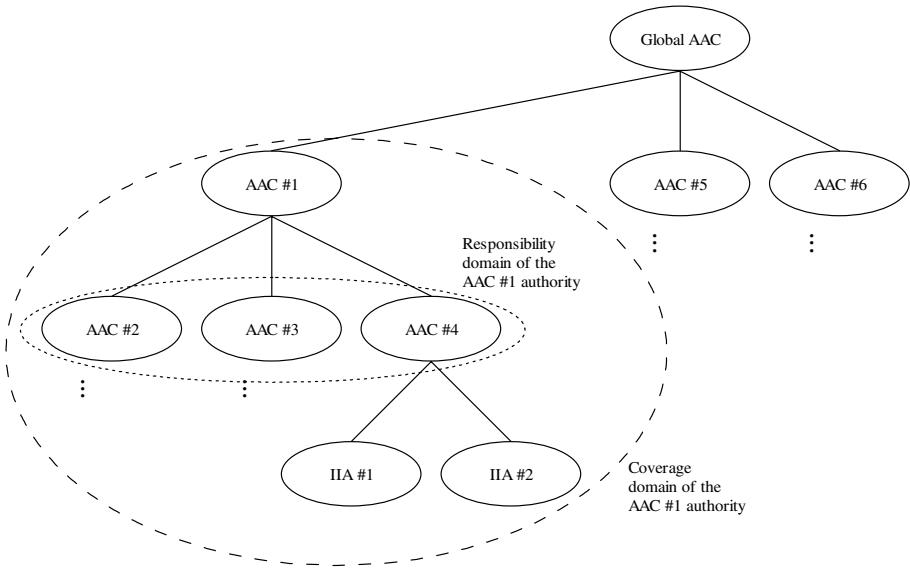
Our DIM framework uses a hybrid solution that merges the advantages of both approaches, while eliminating the disadvantages imposed by either approach. By creating a tree hierarchy with every non-leaf node acting as a catalogue and the leaf nodes being the identity providers. The administration burden is thus broken down to a small set of children nodes, while validation of authorities is trivial by traversing the tree to its root. Here services are required to know only the root node of the tree; they are able to verify the validity of any authority by simply examining if they belong to the tree hierarchy.

Our approach goes a step further, by completely transferring the management of identities and relative information to the user. The identity provider is thus left with all the information needed to verify the authenticity of the user and perform the SSO protocol. All the information about secondary identities and the mapping of identities to services is held and managed by the user. Therefore users are solely responsible for their security and how their personal information is being used (privacy concerns). The advantage of this approach is that even if a catalogue or an identity provider authority is attacked successfully, no information can be gained on how the digital identities were used.

### 3.2 DIM - Authorities Aggregation Catalogue

In the DIM framework there are catalogues holding information about the authorities that belong to the hierarchy. The full name we have assigned to every catalogue is *Authorities' Aggregation Catalogues' (AAC)*. The special AAC of our framework is the root node of the tree, called Global AAC. The identity providers are called *Identity Issue Authorities (IIA)*. [Figure 2] shows a sample hierarchy, with the root node being the Global AAC, all intermediate nodes being AACs and leaf nodes being IIAs. The root node is the authority that is considered to be well known, and this helps define a start point for authority trustworthiness. Subsequent authorities are trusted only if a path exists between the authority in question and the Global AAC.

When engaging in communication with an unknown party, the parent authority can be contacted to verify the party's trustworthiness. This procedure can be repeated for the parent authority as well, until a trusted or a well-known AAC is reached. To establish independent administration domains every AAC is limited to know only the list of direct children authorities, ignoring the presence of a possible sub-tree defined by its children. The selection procedure is essential, since the trust from parent to children authorities is implicit. This type of trust is unconditional, much as blind confidence is transferred from the parents to their children. Parent authorities must vouch for the behavior of their children. A schematic example is shown in [Figure 2] where the responsibility domain of the AAC #1 consists of its children authorities, while the corresponding coverage domain includes every descendant authority.



**Fig. 2.** Example of the responsibility and coverage domain of an authority in the digital identity management hierarchy

### 3.3 DIM - Identity Issue Authority

The IIA is considered a *Trusted Third Party (TTP)* working with the service provider and the users, in order to vouch for digital identities held by the latter. Its primary role is to supply digital identities to individuals and perform the SSO process. A digital identity contains information about the holder of the identity (master identifier, public key and personal information), the authority that issued it and finally a digital signature. The public key and personal information is the information presented by the user to the IIA when creating a new digital identity, while the master identifier is assigned by the IIA and is used to uniquely identify the end-entity. The IIA information block in the digital identity contains details about the authority that issued the digital identity and how to contact it. The digital signature block is created by the issuing IIA and is used as non-repudiation proof as well as for integrity check of the digital identity. For the personal information block, it contains private information and the password supplied by the individual for which the digital identity was created. Along with the digital identity, a secret key is exchanged between the IIA and the individual. We will call this, the identity key, as it is used to encrypt secondary identities. The encryption prevents different digital identities from being associated and at the same time the IIA can validate its authenticity.

Besides creating digital identities the IIA provides two public interfaces, one for end-users and the second for services. The first performs the SSO of the individual, allowing him to use services without needing any other form of credentials. The interface tries to bind the holder of the digital identity with the digital identity itself. To succeed, individuals are required to present their digital identity along with the password (only known to the holder of the digital identity) that was chosen when

creating it. After completing the validation procedure a mapping scheme is exchanged between the IIA and the end-entity middleware, which in fact is a secret key. We call this key validation key and it is used to verify and protect individuals for the current session. To achieve this, the IIA encrypts challenge information with the validation key that only the target individual can decrypt and respond to it.

The second public interface is used by the service provider (in particular by the service mediator) to authenticate an individual requesting access to privileged information. Individuals must first connect to the IIA that issued their digital identities and execute the identity validation procedure. Afterwards individuals can be authenticated to services simply by providing a secondary identity and the IIA contact information. The service mediator will connect to the IIA and present the identity of the individual. The IIA responds informing the service mediator about the legitimacy of the secondary identity. The key aspect of this approach is the ability for an individual to be uniquely identified (proof of legitimacy) and at the same time retaining his anonymity (preserve of privacy). In addition the user is protected from impersonation since he is the only one able to respond to messages encrypted with the validation key.

### 3.4 Service Mediator

The service mediator is one of the two software components that are not part of the DIM framework. It provides an abstract layer to core service functionality; through the mediator services are able to use the advantages of the IMI, keeping their internal structure and logic intact. The functionality it provides includes the receipt of secondary identities from individuals and the validation of this information through the DIM framework. In addition, information about the outcome of the validation procedure is forwarded to the core module of the service, as specific actions may have to be taken depending on the context of the service.

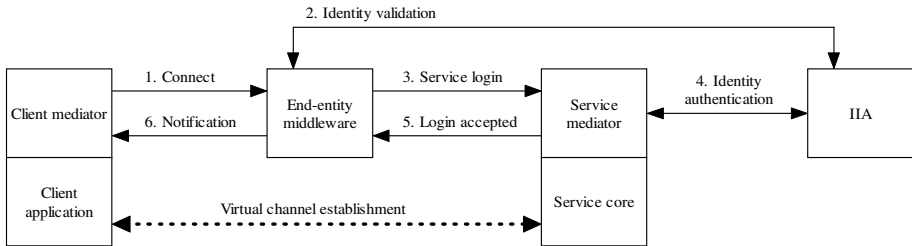
A problem we had to solve was to decide how authority identity information is handled by the service mediator. As we have mentioned the mediator needs to know about the authorities that belong in the DIM framework, in an effort to block digital identities originated from illegitimate IIAs. This information is gathered through the authority validation process, where a valid path must exist between the authority being checked and an established member of the DIM hierarchy. During the process the authorities encountered are stored locally, to avoid revalidating them in future requests. But a question remains regarding the actions to be taken when an authority that was previously a member of the DIM framework leaves the hierarchy. Such actions make all descendant authorities illegitimate, and requests from individuals holding digital identities from those authorities should be rejected. One solution would be to use authority revocation lists, where the service mediator is informed about deletions in the hierarchy in order to prevent further interactions with that branch. The scheme resembles the certificate revocation lists (CRLs). The second solution requires the local list held by the service mediator to act as a cache. Authorities added to this list will be discarded after a certain time span and have to be revalidated. This scheme is simpler, but leaves a period where invalid credentials are

considered legitimate. If we exclude the caching this solution is similar to the online certificate status protocol (OCSP), proposed as an alternative approach to CRLs. This second approach was deployed in the IMI.

### 3.5 End-Entity Middleware

The second software component that is not part of the DIM framework is the end-entity middleware. The main task of this component is to control and handle how digital identity information is published to services. To achieve this, the middleware has an internal storage mechanism, where the digital identity and secondary identities information is stored. The secondary identities are not just piled in a list, but organized according to the services in which they are used. When a connection to a specific service is required, the secondary identity can be retrieved and used. The mapping scheme allows multiple identities to be used in one service, and the same identity to be used across different services. The first permits the user to create multiple and independent profiles in a single service. These profiles cannot be correlated by the service provider, giving the user the flexibility of creating independent accounts at will. The second allows the integration of several services under a single account. Thus information between these services can be exchanged, in order to help the user in his everyday tasks. An example is the case of a car rental and airline company; flight booking information can be exchanged with the car rental company, which in turn can make an offer about renting a car for the same period.

The communication patterns and “virtual channel establishment” between the client application and the service core is shown in [Figure 3], where every transaction is given a sequence number to easily follow the timing of events. For a detailed description we encourage the reader to refer to [12].



**Fig. 3.** Steps for establishing a virtual channel between the service client and core service functionality. The sequence of the steps is given by the numbers preceding the description.

We will now focus on how secondary identities are created and validated. In section 3.3 we mentioned both the identity key and the master identifier. The first is exchanged between the IIA and the end-entity middleware during the digital identity creation, while the second is derived from the personal information presented by the user. The procedure to create a secondary identity starts with the collocation of the master identifier with a sequence of random bits. This creates the secondary identifier, which must be unique if we are creating a new secondary identity. Next the secondary

identifier is passed through a cipher function using the identity key as the encryption key. The outcome is the secondary identity, which can be used to services without privacy concerns as they are opaque to service providers. A secondary identity created by the end-entity middleware will eventually reach the IIA. The task of the IIA is to check the validity of this secondary identity, namely if it matches to the master identifier of the user. The reason we require this matching is because the IIA does not collect information about the secondary identities of the user. So, if a user tries to repudiate that he is the holder of a certain account, the secondary identity can be deciphered, extracting the user's master identifier.

For lack of space we omit discussion of a number of topics regarding the design and implementation of IMI. The interested reader is referred to [12] for a thorough discussion of the design of the communication protocols that we use in IMI, the authority lookup and validation, the identity creation, validation and authentication and the service login. IMI takes advantage of current approaches and technologies in applied cryptography. We make use of both symmetric and asymmetric ciphering techniques, while utilizing digital signatures and certificates [6]. In [12], we also describe the software libraries that were developed, in order to provide a demonstration platform for our global scale SSO system.

### 3.6 Discussion

We now compare the other identity management systems with IMI. To avoid repeating information about similar system we have aggregated Liberty Alliance and WS-\* projects under the federated identity management approach (Federated IM). The only centralized identity management approach is Passport .NET which we will refer to as centralized identity management approach (Centralized IM). As we can see in [Table 1] we use six different traits to compare the different solutions. For additional information and explanation of the table please refer to [12].

**Table 1.** Comparison of the approaches in identity management systems

		Approaches		
		Federated IM	Centralized IM	IMI
Traits	Identity to services	Multiple	Single	Multiple
	Identities to service	Single	Single	Multiple
	Storage of ID info	Distributed	Centralized	Individual
	User aware	No	No	Yes
	Verify ID provider	Point-to-point	Single point	Hierarchy
	Sign-on	Single	Unique credentials	Single

In [12], a detailed analysis is provided about the extra load on the infrastructure that our IMI imposes, mostly in terms of messages exchanged and conclude that this load is not unreasonable. Also in [12] there is a table describing how our solution performs in various attack patterns, and a brief description showing IMI's behavior and information exposure from such attacks.

## 4 Conclusions and Future Work

It is desirable to avoid a single hierarchy for the management of digital identities. Several independent hierarchies should be supported with every hierarchy applying its own specific policies on how authorities are included, or even how digital credentials are supplied. Furthermore we should include the notion of “trust” among authorities and among services and identity providers. The structure would then self-regulate without incorporating complex monitoring tools, dropping out malicious or faulty authorities. The inclusion of trust would also benefit services in spotting misbehaving IIAs and denying access to credentials originating from them.

Another important direction for future work is to incorporate standards proposed by organizations such as OASIS, W3C and make use of well established Internet technologies such as the XML Encryption, XML Signature, SAML and SSL socket connections. In the current IMI we implement our own certificate documents, rather than using version 3 of the X.509 standard. The disadvantage of the X.509 certificates is their binary format, when compared with our self-explained certificates in XML format.

## References

1. Bajaj S. et al. “Web Services Federation Language (WS-Federation)”. IBM Corporation et al, Specification Document: July 2003
2. Hodges J., Wason T. “Liberty Architecture Overview”. Liberty Alliance, White Paper: January 2003
3. Brown K. “Security Briefs: Step-by-Step Guide to InfoCard”. <http://msdn.microsoft.com/msdnmag/issues/06/05/securitybriefs/default.aspx>
4. “idemix”. <http://www.zurich.ibm.com/security/idemix/>
5. Windley P. “Digital Identity”. Sebastopol, California: O'Reilly, 2005
6. Gladman B., Ellison C., Bohm N. “Digital Signatures, Certificates and Electronic Commerce”. April 1999
7. Pfitzmann B., Waidner M. “Anonymity, Unobservability, Pseudonymity, and Identity Management - A proposal for terminology”. Tu Dresden, Department of Computer Science Technical report, 2004
8. Buell A. D., Sandhu R. “Identity Management”. IEEE Internet Computing, November 2003: 26-28
9. Hansen M., Berlich P., Camenisch J., Claub S., Pfitzmann B., Waidner M. “Privacy-Enhancing Identity Management”. Information Security, Elsevier Science Press, 9.1 (2004): 35-44
10. Marsh S. “Identity and Authentication in the E-economy”. Information Security, Elsevier Science Press, 7.3 (2003): 12-19
11. Damiani E., Vimercati S., Samarati P. “Managing Multiple and Dependable Identities”. IEEE Internet Computing, December 2003: 29-36
12. Poursalidis V. “Identity Management Infrastructure for the Digital World”, Master's Thesis, University of Crete, 2005

# Information Security Risk Assessment Model for Risk Management

Dariusz Wawrzyniak

University of Economics, ul. Komandorska 118/120, 53-345 Wrocław, Poland  
`dariusz.wawrzyniak@ae.wroc.pl`

**Abstract.** The article presents a simple model for the information security risk assessment. There are four main elements of the model: security threats, their business impact, security measures and their costs. The *security measures – threats* relationship matrix is the fundamental quantitative tool for the model. The model bases on well known methods like ALE, ROSI and ISRAM but allows for establishing more flexible and more precise metrics supporting the security management process at different organizational levels<sup>1</sup>.

## 1 Introduction

Computer security plays an increasingly important role in nowadays business. More to say, business use of internet has exposed security as one of the key factors for successful e-business competition. Today we need a better understanding and better management of computer security in order to make the e-business secure and effective. Security of information systems is becoming a part of core business processes in every organization involving not only computer specialists but also managers who take the information security responsibility. The problem is not only the question of computer science - it has become the interdisciplinary task taking the advantage of many sciences including the organizational and economical ones as well as statistics and mathematics.

## 2 Computer Security Management Process

Computer security management process deals with a set of fundamental steps carried out periodically in a closed circle as shown in figure 1. These steps base upon the security policy implemented in the company. However, there are also different approaches pointing out the information security risk management as the most general area of the whole business security process. Such approaches mostly base on the Basel Committee documents [9, 12] contents that emphasize

---

<sup>1</sup> The proposed approach is a part of the 1 H02B 016 30 project (Quantitative methods for information security risk management in banking - theory and practical implementations in SAP R/3 system) sponsored by Polish State Committee for Scientific Research.



**Fig. 1.** The general framework for information security management process

the growing importance of information security risk especially in financial business. The security risk management definitions given in the literature sources vary. However, they all point out the significance of the risk assessment procedures aiming at the quantity output values describing the risk level. In fact they take the advantage of the early concepts of risk analysis defined as the systematic categorization of threats to the system and of the counteractive measures and conception of a plan of action which will direct the majority of the resources against the most probable threats and against the greatest risk [14]. To make correct risk analysis it is necessary - in accordance with several criteria - to establish a system of priorities as regards various threats. The factors that should be taken into consideration concern the potential frequency of occurrence of given threat, the size of loss, the degree of difficulty and the costs of introduced security measures as well as the number of potential intruders. The risk analysis is not aimed to design a plan of total protection but merely to ensure a degree of security proportional to the importance of the protected information and so the quantity of the used resources.

### 3 Risk Assessment Problem

There are two types of risk analysis methods. Quantitative risk analysis methods use mathematical and statistical tools to represent risk. In qualitative risk analysis methods, risk is analyzed with the help of adjectives instead of using mathematics. Risk analysis methods that use very intensive quantitative measures are not easy to use for information security risk managers and so they are not commonly used in business practice. On the other hand, qualitative methods do not offer enough information outputs to be useful tools for the risk management process. An example for such method can be found in the Risk Management Guide for Information Technology Systems [8] from the National Institute of Standards and Technology which defines risk management as the



process of identifying risk, assessing risk, and taking steps to reduce risk to an acceptable level. The risk must be systematically assessed to effectively manage it. Security risk assessment determines the level of security risk that exists within the organization. Quantitative risk analysis attempts to assign numeric values to the components of the risk (likelihood and potential damage) and to security controls (cost and effectiveness) so it can also help to calculate the cost effectiveness of the risk management process. However, justifying investments can be problematic because information security often delivers non-financial benefits, rather than an increase in revenue or a reduction in costs. Thus there are few problems that founded the basis for the model proposed below. The most important is: How to develop the information risk assessment model easy to use and allowing for comprehensive quantitative approach alike? The general idea for the proposed model comes from well known approaches dealing with risk analysis and assessment. They are ALE, ISRAM and ROSI<sup>2</sup>.

### 3.1 ALE

Annual Loss Expected (ALE) is a simple and common measurement for risk which can be pointed by three sub models [13]:

$$ALE = (expected\_rate\_of\_loss) \cdot (value\_of\_loss) \quad (1)$$

$$ALE = (impact\_of\_event) \cdot (frequency\_of\_occurrence) \quad (2)$$

$$ALE = \sum_{i=1}^n I(O_i)F_i \quad (3)$$

where  $O_1, \dots, O_n$  is the set of harmful outcomes;  $I(O_i)$  the impact of outcome in monetary value and  $F_i$  the frequency of outcome  $i$ .

Apart from which model will express the choice to formulate ALE, the meaning remains the same, to estimate the loss in order to minimize to an acceptable amount. Besides, ALE can also be utilized as the base for other common metrics as shown in table 1.

---

<sup>2</sup> The methods described in the article are easy to use and most commonly used but the problem is obviously not limited to them. There are a lot of similar approaches taking the advantage of simple quantitative tools. They are (among the others): FMEA Method, BITS Calculator [1], Marion Method, the approaches proposed in ISO/TR 13569 and Basel Committee recommendations. There are also lots of very comprehensive methods using intensive statistics and mathematics allowing for different security aspects measurement like GASSATA algorithm [7], many intrusion detection methods basing on immunological algorithms [4], Markov Models [15], Bayesian Models and fuzzy logic models as well as different economic analysis. These comprehensive models are being used in many software solutions (mainly intrusion detection ones) but most often are limited to certain security system areas and rather do not meet all the requirements of security risk management identified by e-business risk managers.

**Table 1.** Common metrics used by security risk managers [10]

Metric name	Metric symbol	Value calculation
Annual Loss Expected	$ALE$	$(rate\ of\ loss) \cdot (value\ of\ loss)$
Savings (reduction in ALE)	$S$	$ALE_{baseline} - ALE_{with\_new\_safeguards}$
Benefit	$B$	$S + (profit\ from\ new\ ventures)$
Return on Investment	$ROI$	$\frac{B}{cost\_of\_safeguards}$
Internal Rate of Return	$IRR$	$C_0 = \sum_{t=1}^n \frac{B_t - C_t}{(1+IRR)^t}$

### 3.2 ISRAM

The underlying risk model of ISRAM is based on the following formula [6].

$$Risk = Probability\ of\ OSB \cdot Consequence\ of\ OSB \quad (4)$$

where *OSB* - occurrence of security breach.

The risk model of ISRAM, which is deduced from formula (4), is given by formula (5) consists of two main parts, which are the projections of two fundamental parameters in formula (4).

$$Risk = \left( \frac{\sum_m T_1 \left( \sum_i w_i p_i \right)}{m} \right) \left( \frac{\sum_n T_2 \left( \sum_j w_j p_j \right)}{n} \right) \quad (5)$$

where  $i$  - the number of questions for the survey of probability of occurrence,  $j$  - the number of questions for the survey of consequences of occurrence,  $m$  - the number of participants who participated in the survey of probability of occurrence,  $n$  - the number of participants who participated in the survey of consequences of occurrence,  $w_i, w_j$  - weight of the question  $i$  ( $j$ ),  $p_i, p_j$  - numerical value of the selected answer choice for question  $i$  ( $j$ ),  $T_1$  - risk table for the survey of probability of occurrence,  $T_2$  - risk table for the survey of consequences of occurrence<sup>3</sup>.

ISRAM is basically a survey preparation and conduction process to assess the security risk in an organization. Two separate and independent survey processes are being conducted for two risk parameters in formula (5). The preparation and conduction of survey, so as the analysis of its results are defined according to the well defined steps to yield the risk.

### 3.3 ROSI

A simple equation for calculating the Return on Investment for a security investment (ROSI) is as follows[11]:

<sup>3</sup> It has to be noted that the method bases on the seven step procedure and the values used in cited formula are being determined in different steps. For further details please refer to the original article [6].

$$ROSI = \frac{(RiskExposure \cdot \%RiskMitigated) - SolutionCost}{SolutionCost} \quad (6)$$

As emphasized in the cited article, if the method for determining ROSI produces repeatable and consistent results, ROSI can serve as a useful tool for comparing security solutions based on relative value.

## 4 The Risk Assessment Model

All above approaches take some advantage of quantitative methods, however they do not seem to meet all the requirements defined for standalone solutions extensively supporting information security risk management process. Such complex problem requires a bit more sophisticated and flexible methods. However, it must be emphasized that the mathematical tools should never overwhelm the approach transparency not to make it practically unusable. The proposed model takes the advantage of basic matrix processing. Three matrixes has been used representing the relationships between the security threats and their business impact as well as the security measures and their costs. The general idea of the model is presented below.

**T** matrix represents the relationship between security measures and security threats. The higher matrix value the higher measure impact on the threat mitigation. In the other words, the lower value, the higher threat realization probability (dealing with given security measure). The values of the matrix come from

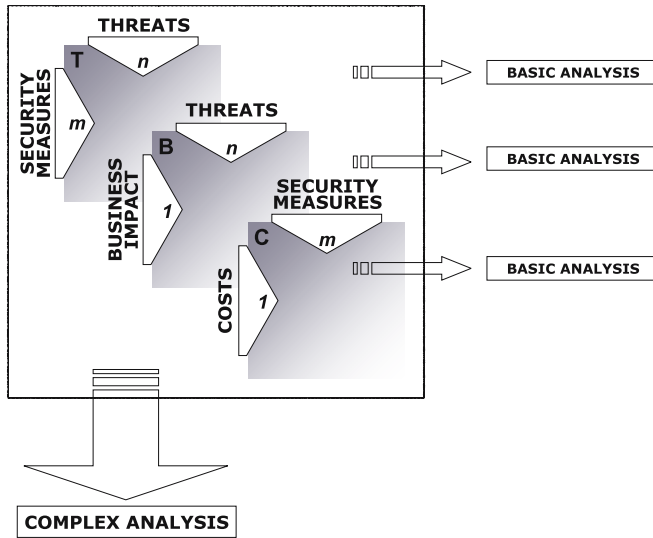
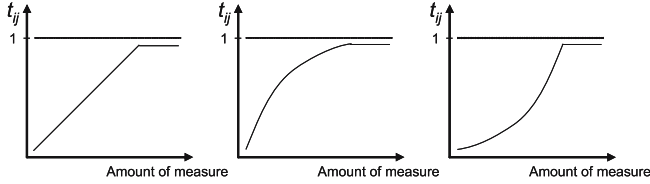


Fig. 2. The model framework

within the interval  $\langle 0 ; 1 \rangle$ . Zero value deals with the case of complete lack of relationship between security measure and the threat. The value equal to 1 is not allowed since it would mean that the measure eliminates the threat completely<sup>4</sup>. So the value 1 can be treated like a asymptote despite the general concept for the matrix values determination that can be seen in the figure 3.



**Fig. 3.** Possibilities for  $t_{ij}$  values assessment

The X axis represents the amount of measure implemented against given threat. The Y axis represents the effect of the measure. Every threat measure relationship generates its own function, and thus every real system takes another  $t_{ij}$  value for every measure on every threat.

The **T** matrix is the central element of the model. It constitutes one of the most important security management relationship between identified security threats and the security measures implemented in the system. The relationship definition is essential for all security management activities since it has a great influence on the process of risk assessment. The **b** vector presents the relationship between the threats realization and their business impact. The threats set is the same as in the **T** matrix. The business impact is being described by the value from within the interval  $\langle 0 ; 1 \rangle$ . The lower value, the lower threat impact. Zero value is not allowed since every identified threat has to feature some business impact (otherwise it is not a threat at all). The value 1 can occur but should be used only in specific cases. The **c** vector presents an additional model information allowing also for some financial analysis. Its concept is very simple - to present the security measurements costs. However, it is recommended that it should contain the standardized costs values. The formulas below present the formal model definition.

$$\mathbf{T} = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ \dots & \dots & \dots \\ t_{m1} & \dots & t_{mn} \end{bmatrix} \quad \text{where} \quad t_{11}, \dots, t_{mn} \in \langle 0 ; 1 \rangle \quad (7)$$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \dots \\ b_n \end{bmatrix} \quad \text{where} \quad b_1, \dots, b_n \in \langle 0 ; 1 \rangle \quad (8)$$

<sup>4</sup> Even if it would be practically possible the threat should not occur in the matrix at all.

$$\mathbf{c} = \begin{bmatrix} c_1 \\ \dots \\ c_n \end{bmatrix} \quad \text{where} \quad c_1, \dots, c_m \in (0 ; 1) \quad (9)$$

Of course the question is how to define such matrixes and how to determine their values. There are few main problems:

- threats identification,
- security measures identification,
- definition of business impacts,
- definition of security measures costs and their standardized values,
- the ways of relationship values assessment.

#### 4.1 Threats Identification

The system threats identification always depends on the security policy fundamentals. For the purpose of proposed model it is essential to rely the identification process also on the management requirements concerning the expected model information output. In the other words, the threats being identified have to accomplish the targets defined for the risk analysis. Is it also very important to choose the required level of particularity. The identified set of threats should consist of the objects of comparable level.

#### 4.2 Security Measures Identification

The system security measures are easy to identify. However, the rules of their identification in the model should meet the chosen level of threats particularity. In the other words, the identified security measures have to correspond with chosen threats. There is no general rule for identifying the set of measures since such identification has to be the result of the analysis purposes and the risk management principles implemented in given case.

#### 4.3 Definition of Business Impacts

The business impacts can be identified only by the use of expert decision strongly supported by the historical data. It is very difficult to measure it since the impact always consists of two main elements. The first one is the financial loss dealing with the threat realization and the second is the moral loss that can impact the business functionality in a very long term. Thus, the expert decision has to take into account both elements as well as the kind of business for given case. For example, financial institutions are especially vulnerable to certain threats realizations since their success depends also on avoiding the moral losses. The business impact level has to meet the threat realization influence on the business continuity. For example, the level value 0,5 would mean the threat realization causes very serious problems while the value very close to 1 would mean the threat realization causes the business collaps.

#### 4.4 Definition of Security Measures Costs

The security measures costs should be identified according to given case analysis requirements and taken into account in their standardized form. The recommended standardization method is the one that outputs with the values from within the interval  $\langle 0; 1 \rangle$ <sup>5</sup>. The values 1 and 0 calculated after standardization should be replaced by the second highest and the second lowest ones respectively in order to avoid 1 and 0 in the vector  $\mathbf{c}$  what is expected in the model.

#### 4.5 Relationship Values Assessment

The relationship values can be assessed by the expert in two ways:

- as a result of the historical data output,
- as a result of experts knowledge.

There is no general rule allowing for the optimal values assessments. Both methods features the advantages and disadvantages alike. The historical data in the case of computer security always feature some kind of inaccuracy due to the rapid security threats and measures development. On the other hand, the arbitral expert decision can be too subjective or can be a result of misunderstanding or lack of knowledge. However, there are few methods for many experts assessments integration and it seems that such solutions should be implemented in the case.

The proposed model is flexible that means the above problems can be solved according to given case requirements as a result of security policy rules and security management requirements.

### 5 Model Output Values and Their Interpretation

The matrixes described above constitute the fundamental basis for the model and can be interpreted at different levels as shown in the figure 2. The first level (Basic analysis) takes the advantage of formalizing the information essential for the risk management. The matrixes show the values that can be treated as the model output values and compared with historical data. It allows for the analysis aiming at identifying the crucial gaps in the security system and - what is even more important - the changes in the security system effectiveness. The second level (Complex analysis) is similar to the first one but gives more information thanks to  $\mathbf{b}$  and  $\mathbf{c}$  vectors. A risk manager can use these vectors optionally. The most comprehensive level concerns both projection of the matrixes and their mathematical transformations. The key analysis method of the model deals with the threats assessment taking into account their probability of realization and their business impact. The assessment procedure requires following steps:

---

<sup>5</sup> For example:  $z_{ij} = \frac{x_{ij} - \min_i x_{ij}}{O_j}$ , where  $z_{ij}$  is the variable value after the standardization and  $O_j$  is the variable range.

i. Determining the threat probability of occurrence vector (**tp**) basing on the **T** matrix. **tp** values come from the highest values in **T** columns describing the highest level of given threat risk mitigation.

$$\mathbf{tp} = \begin{bmatrix} tp_1 \\ \dots \\ tp_n \end{bmatrix} \quad \text{where} \quad tp_i = 1 - \max_{j=1..m} (t_{ji}) \quad (10)$$

ii. Determining the vector presenting the threats probability of occurrence multiplied by their business impact.

$$\mathbf{tx} = [tx_1 \dots tx_n] \quad \text{where} \quad tx_i = tp_i \cdot b_i \quad (11)$$

iii. Calculating the output risk value  $r_x$  - the most synthetic model metrics.

$$r_x = \mathbf{tx} \circ \mathbf{b} \quad (12)$$

The  $r_x$  value comes from within the interval  $\langle 0; n \rangle$ . The higher  $r_x$  value, the higher information security risk level.

Similar procedure can be implemented also in the security measures costs analysis. It can be conducted by the use of following steps:

i. Determining the vector **sx** presenting the security measures effectiveness.

$$\mathbf{sx} = \begin{bmatrix} sx_1 \\ \dots \\ sx_m \end{bmatrix} \quad \text{where} \quad sx_i = 1 - \max_{j=1..n} (t_{ij}) \quad (13)$$

ii. Determining the **cx** values as a result of multiplication of **sx** values and **c** values

$$\mathbf{cx} = \begin{bmatrix} cx_1 \\ \dots \\ cx_m \end{bmatrix} \quad \text{where} \quad cx_i = sx_i \cdot c_i \quad (14)$$

Thus the **cx** vector presents the security measures cost-efficiency relationship emphasizing the measures that feature high costs and low efficiency. The values comes from within the interval  $(0; 1)$  and the point is that the security measures featuring high values should be carefully checked since most probably they do not return their investments optimally.

The above operations enhance the **T** information capability due to the emphasizing the threats featuring high business impact as well as the security measures that are not efficient enough. The output **tx**,  $r_x$ , **sx** and **cx** should constitute the fundamental basis for the risk management process.

## 6 Conclusions

The proposed model does not aim at the general decomposition of information security risk assessment problem. The purpose for the approach is to propose

the model easy to business use, flexible and taking the advantage of some basic quantitative methods. There is also a possibility for using the model in multiple mode dealing with more than one set of matrixes. It must be also emphasized that the model effectiveness and practical functionality depend mostly on the right expert decisions. The crucial point is that these decisions have to base mainly on the historical data dealing with the security threats realizations and their business impact and not only on the experts knowledge.

## References

- [1] BITS Key Risk Measurement Tool for Information Security Operational Risks, BITS Financial Services RoundTable (2004)
- [2] Cavusoglu H., Mishra B., Raghunathan S.: A Model for Evaluating IT Security Investments. *Communications of the ACM*. Vol. 47, No. 7 (2004)
- [3] Davis A.: Return on security investment - proving its worth it. *Network Security* 11 (2005), 8-10
- [4] Dhaeseleer P., Forrest S., Helman P.: An Immunological Approach to Change Detection: Algorithms, Analysis and Implications. *IEEE Symposium on Security and Privacy* (1996)
- [5] Gordon L.A., Loeb M. P., Lucyshyn W.: Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy* 22 (2003) 461485
- [6] Karabacak B., Sogukpinar I.: ISRAM: information security risk analysis method. *Computers & Security* (2005) 24, 147-159
- [7] Me L.: GASSATA, a Genetic Algorithm as an Alternative Tool for Security Audit Trails Analysis.
- [8] Risk Management Guide for Information Technology Systems. National Institute of Standards and Technology. Special Publication 800-30 (2001)
- [9] Risk Management Principles for Electronic Banking. Basel Committee on Banking Supervision (2003)
- [10] Schechter E.: Computer Security Strength & Risk: A Quantitative Approach. Thesis presented to The Division of Engineering and Applied Sciences. Harvard University (2004), 29
- [11] Sonnenreich W.: Return On Security Investment (ROSI): A Practical Quantitative Model. A summary of Research and Development conducted at SageSecure (2002)
- [12] Sound Practices for the Management and Supervision of Operational Risk. Basel Committee on Banking Supervision (2003)
- [13] Tsiakis T., Stephanides G.: The economic approach of information security. *Computers & Security* (2005) 24, 105-108
- [14] Wawrzyniak D.: Organizational Aspects of Data Security in Banking Computer Systems. *Business Information Systems Proceedings*. W. Abramowicz (ed.) (1998) 237-245
- [15] Wawrzyniak D.: Zarządzanie bezpieczeństwem systemów informatycznych w bankowości. *Wydawnictwo Zarządzanie i Finanse* (2002).



# On the Limits of Cyber-Insurance

Rainer Böhme<sup>1</sup> and Gaurav Kataria<sup>2</sup>

<sup>1</sup> Institute for System Architecture, Technische Universität Dresden  
`rainer.boehme@tu-dresden.de`

<sup>2</sup> Heinz School of Policy and Management, Carnegie Mellon University  
`gauravk@andrew.cmu.edu`

**Abstract.** It has been argued that cyber-insurance will create the right kind of security atmosphere on the Internet. It will provide incentive (through lowered premiums) to firms to better secure their network thus reducing the threat of first party as well as third party damage, promote gathering and sharing of information security related incidents thus aiding development of global information security standards and practices, and finally, increase the overall social welfare by decreasing the variance of losses faced by individual firms via risk pooling as in other kinds of insurance. However, a unique aspect of cyber-risks is the high level of correlation in risk (e.g. worms and viruses) that affects both the insurer and the insured. In this paper, we present a discussion on the factors that influence the correlation in cyber-risks both at a global level, i.e. correlation across independent firms in an insurer's portfolio, and at a local level, i.e. correlation of risk within a single firm. While global risk correlation influences insurers' decision in setting the premium, the internal correlation within a firm influences its decision to seek insurance. We study the combined dynamics of these two to determine when a market for cyber-insurance can exist. We address technical, managerial and policy choices influencing both kind of correlations and welfare implications thereof.

## 1 Introduction

The usual approach to managing information security risk is similar to other business risks, i.e. first eliminate, then mitigate, absorb and then, if possible, transfer. Since eliminating security risks in today's environment is not possible, managers deploy protection technologies like firewall, antivirus, encryption, and instate appropriate security policies like passwords, access control, port blocking etc. to mitigate the probability of a break-in or failure. If the residual risk is manageable it is absorbed, otherwise, transferred by either outsourcing security or buying insurance.

Though this approach seems appropriate, it creates a widening rift between security experts who propose employing standardized best practices and deploying homogeneous software to enhance system manageability thereby reducing vulnerabilities, versus those, who propose using cyber-insurance as a means of

transferring risks associated with system vulnerabilities. This is because insurance relies on the principle of independent risks while standardized system environments by themselves create a global monolithic risk manifested in virtually every standardized system. Unlike in physical world where risks are geographically dispersed, in information world, network exploits, worms and viruses span all boundaries. All systems that run standardized software and processes are vulnerable, because bugs in them, once discovered, are common knowledge and can be exploited anywhere. This potentially creates a situation where not only *all* systems within an organization could potentially fail by virtue of their being identical and vulnerable to same exploits, but all similar systems worldwide could fail affecting many organizations simultaneously as seen in case of worms like *SQL Slammer*, *Code Red* etc. Ironically, most techniques for security risk mitigation could themselves induce correlated failures as they too are standardized. For instance, antivirus updates, IDS attack signatures and software patches are all downloaded from web sources, which, if compromised can in turn compromise millions of systems that depend on them for their security [1]. Such possibilities should surely cross the mind of an insurer who plans to offer cyber-insurance to only those businesses which “responsibly” manage their information system by “timely” updating their antivirus, firewall, IDS etc.

The existence of high correlation in breach or failure of information systems adds a new dimension to risk management that has rarely been looked at in the context of information security [2,3]. Information security risk management has been studied by Soo Hoo [4], Schechter and Smith [5], Arora et al. [6] and Gordon et al. [7,8]. Majuca et al. [9] propose cyber-insurance as an effective strategy for security risk management. They study the evolution of cyber-insurance market citing *moral hazard* and *adverse selection* as the primary concerns. Ogut et al. [10] and Kunreuther et al. [11] discuss interdependent risks between firms and their suppliers. Yet, most studies in this area have not explicitly modeled correlated risks and the impediments they cause to cyber-insurance except Böhme [12] and Geer et al. [2]. In insurance and actuarial literature, research on aggregation of correlated risks and extreme value theory (EVT) is abundant [13]. However, the research in that area has not focused on modeling correlated risks within a single firm seeking insurance.

While global risk correlation influences insurers’ decision in setting the premium, the internal correlation within a single firm influences its individual decision to seek insurance. A risk-averse firm prefers low variance of loss and hence low correlation of failure amongst its internal systems. This paper is, to the best of our knowledge, the first attempt to simultaneously analyze the causes of internal (within a single firm) and global (across multiple firms) correlation of cyber-risks and to estimate their combined effect on the presence of cyber-insurance market.

The remainder of this paper is organized as follows: Section 2 further motivates the rationale behind a two-step risk arrival process with correlation on either stage. Sections 3–5 describe the model, where Sect. 3 deals with the supply-side, Sect. 4 covers the demand-side, and Sect. 5 formulates the market equilibrium

conditions. The model has been employed in large-scale Monte Carlo simulations to identify conditions where a market for cyber-insurance can exist. The results thereof are presented in Sect. 6 before we conclude in Sect. 7.

## 2 Security Risk Correlation

Due to significant homogeneity and presence of dependencies in computer systems their failure is highly correlated. Recent spate of Internet worms like *MS-Blaster* and *Sasser* have highlighted this very threat. These worms exploited vulnerabilities present in ubiquitous Microsoft Windows operating system to infect millions of computers worldwide. Although worms and viruses receive maximum media attention, there are many other factors that can cause damage to a firm's information system, e.g. insider attacks, configuration errors, targeted hacker attack, hardware failure, software bugs, and defective patches among others.<sup>1</sup>

Unlike individual firms that care about correlated failure of systems only within their own network, the insurance companies are concerned about global correlation in their entire risk portfolio because that affects the risk premium they charge individual firms. Interestingly, the factors that influence security outcomes exhibit different correlation properties (see Table 1).

**Table 1.** Examples for different kinds of cyber-risk correlation

	Low $\rho_G$	High $\rho_G$
High $\rho_I$	Insider attack	Worms and viruses
Low $\rho_I$	Hardware failure	Spyware/phishing

The failure of a computer within a firm due to a hardware problem is likely neither influenced by, nor is it expected to influence failure of other computers in the same firm or other firms. This incident can therefore be considered to exhibit low intra-firm correlation (henceforth  $\rho_I$ ) and low global correlation (henceforth  $\rho_G$ ). Insider attacks exhibit high  $\rho_I$  but low  $\rho_G$  because an insider who is abusing his privileges like admin password can affect almost all computers within his domain but cannot compromise computers outside his administrative domain [14]. In contrast, software attacks involving user interaction, such as phishing or spyware, have high  $\rho_G$  and low  $\rho_I$  because a few careless employees in many different firms may respond to a phishing email or install a new game at work thereby infecting or compromising their system. But all such employees are likely not clustered within a single firm. Finally, worms and viruses normally exhibit both high  $\rho_I$  and  $\rho_G$  because they are seldom contained within a single network.

<sup>1</sup> Due to space limitation, a detailed explanation of each factor is not provided in this version.

The research in network security area is striving to develop techniques to contain spread of worms and viruses by automatic generation and distribution of attack signatures [15,16,17]. As these techniques make use of the concurrence of malicious traffic to identify pattern and extract signatures, global correlation may be reduced by the maturing of those technologies, but it is unlikely to vanish completely. On the other hand, internal correlation is unlikely to reduce much by use of such reactive techniques as the response time associated with them can be too high. Chen et al. [3] propose using software diversity to limit internal correlation.

### 3 Supply-Side: Two-Step Risk Arrival with Correlation

In this paper, we propose to address the particularities of cyber-risks in a two-step risk arrival process. The first step models the aggregation of cyber-risks within a single firm's network represented by  $n$  nodes. The second step aggregates the risks in the portfolio of an insurer issuing coverage to  $k$  similar firms. We allow for correlation on both steps, whereas the extent of dependence may vary between the portfolio level (global correlation  $\rho_G$ ) and the firm level (internal correlation  $\rho_I$ ).

We model correlated failure of computers within firms using Beta-binomial distributions [3]. The Beta-binomial distribution is a randomized Binomial distribution where the prior for the underlying Bernoulli trials is Beta distributed. This lends Bayesian subjectivity to the correlation of individual Bernoulli trials, which can be estimated by security analysts based on the technical and managerial set up within a firm. Hence,  $\rho_I$  is the correlation parameter of the Beta-binomial distribution. The Beta-Binomial distribution has previously been proposed in computer science literature to model correlated failure of backup systems [18] and to model failure across multiple versions of software [19].

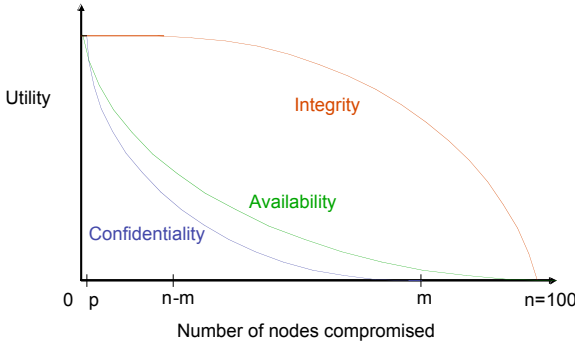
As mentioned above, the insurer has  $k$  firms in its risk portfolio. The losses and thus claims at firms are correlated due to presence of global correlation  $\rho_G$ . We model the distribution of these correlated risks in the overall portfolio using copulas [13]. Copulas are sophisticated statistical tools to model dependence of arbitrary probability distributions. In this paper, we use the  $t$ -copula because of its property to model correlation of extreme events [20]. Details on the mathematical formulation of correlation structure are omitted here for the sake of brevity. We therefore refer the reader to our working paper [21].

### 4 Demand-Side: Information Security Risk Management

The supply-side model outlined in the previous section allows an insurer to calculate appropriate premiums for cyber-risks with given correlation profiles  $(\rho_I, \rho_G)$ . A demand-side model is needed to analyze when and whether it is optimal for firms to buy insurance coverage at a given premium. In the following we are going to introduce a stylized model of the business value of information technology and then discuss operable compensation schemes for cyber-insurance contracts with regard to the intangible nature of information assets and their difficulty to value and substantiate.

#### 4.1 Information Assets

The efficacy of a firm's information system is determined by its ability to store, process and retrieve information in an efficient manner. While some industries like e-commerce depend completely on their information systems, other industries depend on them to a varying degree, to carry out their business. Failure of information system due to an attack or malfunction can severely limit certain business functions that depend on information storage, processing or retrieval.<sup>2</sup> Therefore, most systems are designed to incorporate some level of redundancy or fault-tolerance at both communication and storage level. In a typical network setting, clients store information on servers which distribute it among other servers for consistency, load-balancing and fault-tolerance. Performance and security are generally competing goals when dealing with information [22]. No redundancy implies higher performance and low security, while backups and consistency-checks enhance security at the cost of lowering performance [23]. Numerous threshold schemes for the design of storage systems have been proposed [24]. These schemes have three parameters:  $n$ ,  $m$  and  $p$  (where  $n \geq m \geq p$ ). We assume that the information asset of a firm is divided among  $n$  nodes on its network. Due to presence of some redundancy in the network the entire information can be recreated with help of any  $m$  nodes. Assuming that some dependencies exist among them, at least  $p$  nodes need to be compromised to breach any useful information (where  $p$  can also be equal to 1 in case of no dependency).



**Fig. 1.** The fall in utility as a function of nodes compromised

Under this setup we observe the impact of node failure on the firm for each of the three common protection goals (Figure 1):

**Confidentiality:** To steal complete information an adversary needs to compromise at least  $m$  nodes. It can steal some information if the number of nodes breached is  $\geq p$ .

<sup>2</sup> Even if fall-back plans exist, continuing core business without IT is accompanied by productivity losses.

**Integrity:** Information can be restored if number of node failures is no greater than  $n - m$ .

**Availability:** Due to dependencies and interconnection of nodes on the network, the failure of one node affects other nodes. The degrading effect can be high for nodes which have high dependencies like print servers, file servers, routers etc, while a stand alone desktop has only minimal effect.

## 4.2 Transition from Protection Goals to Loss Amounts

From the above shown relationship between the number of failed nodes and the enforcement of security properties, specific loss functions  $\ell(x)$  can be derived. A loss function maps the physical state (number of node failures) to disutility a firm faces due to that physical loss. For sake of simplicity in our preliminary analysis we assume a linear loss function and a CRRA<sup>3</sup> utility function that maps failure/breach of systems to the firm's utility. A risk-averse utility function is one where firms prefer low variance of loss even when expected loss remains unchanged. In a competitive insurance market, firms pay a premium that is marginally greater than the expected loss in order to avoid exposure to the risk. Due to the unique correlation structure of cyber-risks it is not certain that the premiums are always economically reasonable. In the next section we investigate how our models for supply-side (Sect. 3) and demand-side (this section) interact and identify cases where cyber-insurance is practical.

## 5 Existence of a Cyber-Insurance Market

Due to the very nature of information assets it becomes extremely difficult to objectively quantify confidentiality, integrity or availability of information and the loss caused to the firm due to breach in any/all of them. For instance, breach of two megabytes out of ten megabytes of trade secret does not necessarily translate into a 20 % breach of confidentiality. However, for insurers to come up with practical policies it is essential that the risk be objectively and unambiguously defined, therefore, we believe if claims are linearly dependent on the number of computer/system failure then a policy can be unambiguously offered and objectively monitored. Based on this simple setup we explore the interaction between the demand side and the supply side of cyber-insurance.

Given an insurance premium of  $\gamma$  per node, the firm chooses the fraction  $\lambda^*$  as the amount of insurance coverage bought for each node on its network, which maximizes its expected utility. In the limit case  $\lambda^* = 0$ , the firm decides to buy no insurance at all and bear all risks internally (self-insurance, see [26]). The firm thus pays a net premium of  $\lambda^* \cdot \gamma \cdot n$  to the insurance company, and in case of loss due to failure of  $x$  computers it receives a compensation of  $x \cdot \lambda^*$ . However, premiums are not determined exogenously, they depend on the expected expenditure of insurance companies to settle all claims in a given period. The insurers' costs  $C$  in a single period can be expressed as a sum of three components

---

<sup>3</sup> Constant Relative Risk Aversion, see [25].

$$C = E(L) + A + i \cdot c$$

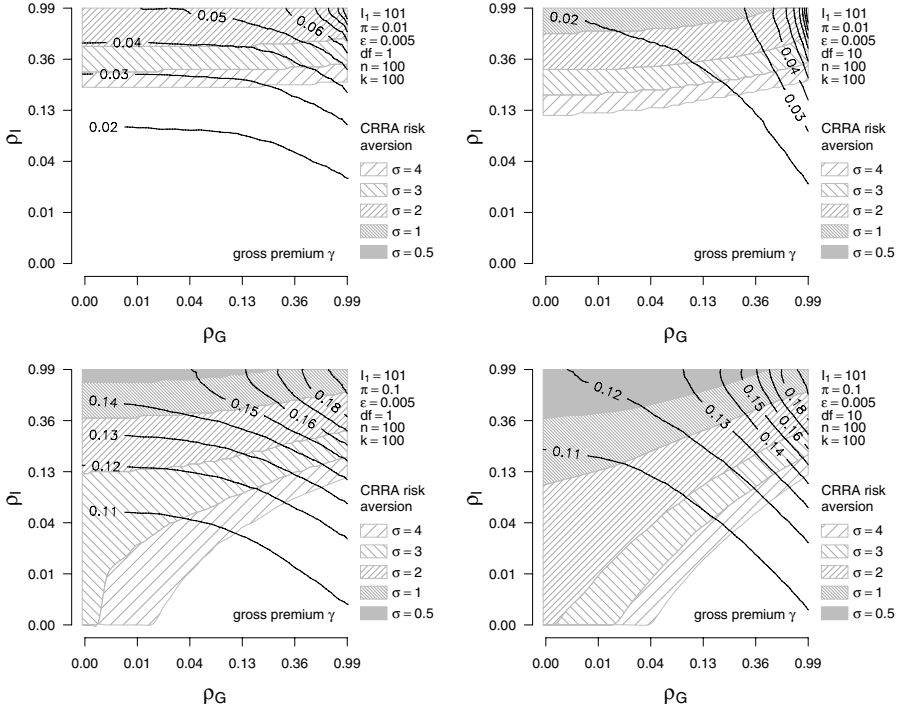
Where,

- $E(L)$  is the expected loss amount, with  $L$  being a random variable.
- $A$  is the sum of all administrative costs, which we assume to be negligible.
- $c$  is the safety capital required to settle all claims if the realization of  $L$  turns out to be the  $\epsilon$ -worst case ( $\epsilon$  is the probability of ruin for the insurer).
- $i$  is the interest rate to be paid for the safety capital  $c$ . The rate should reflect the risk associated with the business in general and the choice of  $\epsilon$  in particular.

Parameters  $\epsilon$  and  $i$  are exogenous in our model, and we use values of  $\epsilon = 0.005$  and  $i = 0.1$  (similar to [12]). Since  $E(L)$  solely covers the average case, the importance of safety capital to avoid the ruin of the insurance company is evident. Determining the right amount of  $c$ , however, is difficult because it depends on the tails of the loss distribution  $L$ .  $L$  itself is a sum of  $k$  correlated random variables, modeling the loss amount of each individual firm in the insurers' portfolio, which is again a sum of the correlated random variables modeling the risk arrival process at each individual node in the firm. The shape of the p.d.f. after each of the convolution steps depends on the amount of correlation, so both  $\rho_I$  and  $\rho_G$  appear in the calculation of premium  $\gamma$ , which makes the derivation of  $L$  in closed form intractable. Consequently, we resort to Monte Carlo simulation methods to determine the regions in the  $\rho_I - \rho_G$  plane, where a sound business model to offer cyber-insurance at reasonable premiums exists. This is equivalent to identifying regions with non-negative consumer and supplier surplus, therefore yielding positive welfare effects.

## 6 Results from a Simulation Study

We first calculated the premium that the insurers need to charge firms to insure risks with certain correlation properties. Then, taking the premium as given, we calculated the firm's utility both with and without insurance to determine when a firm would opt for insurance. The plots in Figure 2 show the premiums, and indicate which regions satisfy the conditions for insurance market to exist. We notice that with increase in risk-aversion firms prefer insurance. However, they prefer not to insure risks if both  $\rho_I$  and probability of failure  $\pi$  are low. This is so, because firms already achieve a kind of portfolio balancing within their own network and thus do not need to buy external risk balancing. Insurers, on the other hand, demand higher premium in presence of high global correlation  $\rho_G$ , which is required to balance a clustered portfolio. Therefore, we see that only firms with higher risk-aversion demand insurance when  $\rho_I$  is low and premium is high. Finally, the insurable region deteriorates for small shape parameters ( $df$ ) of the  $t$ -copula, which reflects a stronger dependency in the tails of the distribution. Since the entire joint distribution determines insurability, empirical research is needed to find the most appropriate copula and the parameterization for different classes of cyber-risks.



**Fig. 2. Insurable regions:** Contour lines indicate the minimum gross insurance premium  $\gamma$  to cover a normalized risk of par value 1 for varying level of  $\rho_I$  and  $\rho_G$ . White areas are “uninsurable”, hatched areas indicate regions where cyber-insurance is practical for risk aversion equal or above a given level  $\sigma$ ;  $\pi$  = prob. of computer failure;  $\epsilon$  = prob. of ruin for insurer;  $I_1$  = initial wealth of firm;  $df$  = shape of the  $t$ -copula;  $n$  = no. of computers per firm;  $k$  = no. of firms in insurer’s portfolio. Results obtained from Monte Carlo simulation with 20,000 trials per parameter setting.

Lack of research in this area could be a reason for why cyber-insurance market has not matured yet. *Mi2g*, a reputed security trend analysis and consulting firm, estimates global loss due to security incidents in upwards of US \$200 billion, while the current cyber-insurance market is worth only about US \$2 billion [9]. We believe that a more detailed analysis of security outcomes following the correlation among component factors, as we describe, will be helpful in preparing market friendly coverage policies.

In addition, technical, managerial and policy approaches could be developed that can favorably alter the inherent correlation structure of the market. On the technical side, a stronger emphasis on diversity of system platforms might be an appropriate measure to counter both internal [3] and global [12] correlation. Techniques for automatic worm signature generation and distribution should be perfected, while at the same time, the current practice of unreserved auto-updates of system or application software should be reconsidered (see also [1]). On the managerial level, the recent trend to standardization, through outsourcing or other



means, may create latent liabilities that have not yet appeared on the horizon of risk management and thus are not reported on the balance sheet. Finally, policy makers can address correlation via diversity in several ways. They have indirect control of the market structure in software markets via competition policy, and/or by making cyber-insurance compulsory for certain businesses. A direct stimulus with less regulatory burden can also be given by assigning diversity a higher priority in public procurement. The exact measures and its likely outcomes, however, are to be evaluated in more targeted research and on the basis of empirical data.

## 7 Conclusion

We have shown how the correlation structure in cyber-risks can be incorporated in an economic model that takes into account the specific properties of both information assets and IT risks, namely systemic interdependence of loss events within and between firms. This model has been employed in simulation analyses to infer parameter constellations where a market for cyber-insurance can exist in theory and where it cannot. We have shown that cyber-insurance is best suited for classes of risk with high internal and low global correlation. This is so, because low internal correlation allows firms to realize self-insurance in their own network and thus limits demand for cyber-risk transfer. High global correlation, in turn, causes imperfect risk-pooling in the insurers' portfolios. Consequently, insurers have to add high safety loadings to the premiums and thus limit the supply for cyber-insurance. Scholars of cyber-risk management should incorporate these findings in future work on the optimal mixture of instruments. This includes regarding all measures that may influence the correlation between cyber-losses as relevant to risk management.

## References

1. Beattie et al., S.: Timing the application of security patches for optimal uptime. In: *Proceedings of LISA 2002: 16<sup>th</sup> Systems Administration Conference*, Berkeley, CA, USENIX Association (2002) 233–242
2. Geer et al., D.: CyberInsecurity – The cost of monopoly (2003) . <http://www.ccianet.org/papers/cyberinsecurity.pdf>
3. Chen, P.Y., Kataria, G., Krishnan, R.: Software diversity for information security. In: *Workshop on the Economics of Information Security (WEIS)*, Harvard University, Cambridge, MA (2005) <http://infosecon.net/workshop/pdf/47.pdf>.
4. Soo Hoo, K.J.: *How Much Is Enough? A Risk-Management Approach To Computer Security*. PhD thesis, Stanford University, CA (2000) . <http://cisac.stanford.edu/publications/11900/>
5. Schechter, S.E., Smith, M.D.: How much security is enough? A risk management approach to computer security. In Wright, R.N., ed.: *Financial Cryptography (7th Int'l Conf.)*. LNCS 2742, Berlin Heidelberg, Springer Verlag (2003) 73–87
6. Arora, A., Hall, D., Pinto, C.A., Ramsey, D., Telang, R.: Measuring the risk-based value of IT security solutions. *IEEE IT Professional Magazine* **6** (2004) 35–42
7. Gordon, L.A., Loeb, M.P.: The economics of information security investment. *ACM Transactions on Information and System Security* **5** (2002) 438–457

8. Gordon, L.A., Loeb, M.P., Sohail, T.: A framework for using insurance for cyber-risk management. *Communications of the ACM* **46** (2003) 81–85
9. Majuca, R.P., Yurcik, W., Kesan, J.P.: The evolution of cyberinsurance. In: *ACM Computing Research Repository (CoRR)*, Technical Report cs.CR/0601020 (2006)
10. Ogut, H., Menon, N., Ragunathan, S.: Cyber insurance and IT security investment: Impact of independent risk. In: *Workshop on the Economics of Information Security (WEIS)*, Harvard University, Cambridge, MA (2005) <http://infosecnet.net/workshop/pdf/56.pdf>.
11. Kunreuther, H., Heal, G.: Interdependent security. *Journal of Risk and Uncertainty* **26** (2003) 231–249
12. Böhme, R.: Cyber-insurance revisited. In: *Workshop on the Economics of Information Security (WEIS)*, Harvard University, Cambridge, MA (2005) <http://infosecnet.net/workshop/pdf/15.pdf>.
13. Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extremal Events for Insurance and Finance*. Second edn. Springer Verlag, Berlin Heidelberg (1999)
14. Schultz, E.E.: A framework for understanding and predicting insider attacks. In: *Proc. of Compsec*, London, UK (2002) 526–531
15. Kreibich, C., Crowcroft, J.: Honeycomb - creating intrusion detection signatures using honeypots. In: *Proceedings of the Second Workshop on Hot Topics in Networks (HotNets-II)*. (2003)
16. Singh, S., Estan, C., Varghese, G., , Savage, S.: Automated worm fingerprinting. In: *Proceedings of the 6th ACM/USENIX Symposium on Operating System Design and Implementation (OSDI)*. (2004)
17. Newsome, J., Karp, B., Song, D.: Polygraph: Automatic signature generation for polymorphic worms. In: *Proceedings of the IEEE Security and Privacy Symposium*. (2005)
18. Bakkaloglu, M., Wylie, J., Wang, C., Ganger, G.: On correlated failures in survivable storage systems (2002) Technical Report CMU-CS-02-129, Carnegie Mellon University, School of Computer Science.
19. Nicola, V.F., Goyal, A.: Modeling of correlated failures and community error recovery in multiversion software. *IEEE Transactions on Software Engineering* **16** (1990) 350–359
20. Demarta, S., McNeil, A.J.: The  $t$  copula and related copulas. *International Statistical Review* **71** (2005) 111–129
21. Böhme, R., Kataria, G.: Models and measures for correlation in cyber-insurance. In: *Workshop on the Economics of Information Security (WEIS)*, University of Cambridge, UK (2006) <http://weis2006.econinfosec.org/docs/16.pdf>.
22. Wylie et al., J.J.: Survivable information storage systems. *IEEE Computer* **33** (2000) 61–68
23. Shamir, A.: How to share a secret. *Communications of the ACM* **22** (1979) 612–613
24. Rabin, M.O.: Efficient dispersal of information for security, load balancing and fault tolerance. *Journal of the ACM* **32** (1989) 335–348
25. Pratt, J.W.: Risk aversion in the small and in the large. *Econometrica* **32** (1964) 122–136
26. Ehrlich, I., Becker, G.S.: Market insurance, self-insurance, and self-protection. *Journal of Political Economy* **80** (1972) 623–648

# Towards a Risk Management Perspective on AAI

Christian Schläger and Thomas Nowey

University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany  
{christian.schlaeger, thomas.nowey}@wiwi.uni-regensburg.de

**Abstract.** Authentication and Authorisation Infrastructures (AAIs) support service providers on the internet to outsource security services. Motivations for their usage stem from software engineering and economics. For the latter an assessment of inherent risks is needed. In this work the authors deduct an appropriate, formalistic risk assessment method for AAI and analyse outsourceable security services in comparison to traditional – non AAI involved – service providing. To achieve the assessment of risks various methods for risk management have been analysed and finally a suitable qualitative method has been chosen. As AAI differ in their potential to cover security services, combinations of these services are compared. The given risk assessment method enables providers to decide on a special infrastructure for their purpose and lets users of AAI determine if given advantages surpass the immanent risks. This work also enables service providers to estimate costs for such an infrastructure and calculate potential savings.

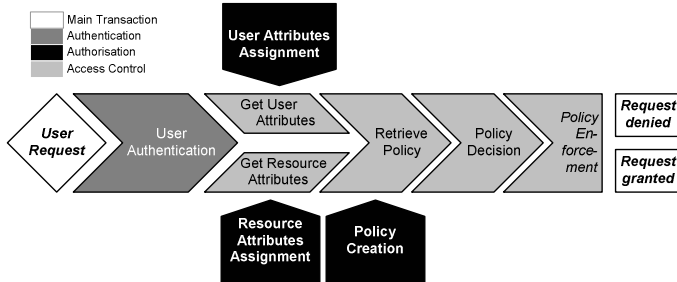
## 1 Introduction

Service providing on the internet has been a huge success story. Although ease of use is proclaimed in many advertisements, the usage of a service on the internet – e.g. to buy a book or to use a geographic routing service – is not trivial at all. The purchase of a book is not simply a link to click on but it stands at the end of a sequel of security and data intensive processes – most of them hidden from the user. The complexity of doing business over the internet has increased both for customers and vendors. Concentrating on security connected processes we find a chain of distinctive services linking the user's request with the service providing as shown in Fig. 1. The given chain of security services is enhanced by an attribute infrastructure like deduced from OASIS' XACML and SAML standards in [14].

Risk is an omnipresent factor in internet transactions. Risks have to be identified and valued to decide upon appropriate controls and monitoring mechanisms. Basically, one has the options to either avoid, reduce, shift, or accept a certain risk.

In [14] and [15] it is argued that Authentication and Authorisation Infrastructures (AAIs) can be used to source out security services in order for the service provider to concentrate on core competencies, raise the overall level of security, provide new, flexible, and more powerful access control services like ABAC (attribute-based access control), and strengthen the usability through user's Single Sign-On (SSO). However, the usage of a new architecture, especially if not entirely under the control of a service provider, raises questions about risk assessment in comparison to

traditional methods of service providing on the internet. The authors show that different approaches to AAIs are available each with inherent benefits and shortcomings. Differences result from the architecture, the level of outsourcing, and specific use cases.



**Fig. 1.** Security services for accessing a resource in an attribute enhanced infrastructure

In this paper we identify and measure risk factors in traditional e-commerce surroundings versus e-commerce applications with an AAI. Although AAIs are by nature generic architectures, e-commerce has been taken as an example.

## 2 Related Work

The topic of AAIs as a tool for service providers on the internet has been discussed on a technical level by [9], making a comparative survey, and in more detail by [15] in 2005. Various architectures of research projects and products have been analysed and motivation for the parties has been given: e.g. [14] proposed a reference architecture for an AAI respecting privacy and flexibility. The idea of Single Sign-On has been discussed in the field of identity management. A classification of architectures can be seen in [5]. A quite technical paper by [7] has analysed the risks in the Microsoft Passport protocol. All work that can be found today on AAIs, the most recent given here, has so far neglected the risk assessment in these architectures in comparison to traditional service providing.

That risk in e-commerce is immanent is being reported regularly by intelligence agencies, other governmental institutions, or the media. The interested reader is pointed to [6] for a survey of general risks in e-business.

Risk management techniques have become a vital element of modern security management. A risk is an unwanted event that has negative consequences [12] and can be described as the “combination of the probability of an event and its consequence”.

Systematic risk assessment is especially helpful when it comes to the economic evaluation of information security investments [11]. In literature and in practice numerous methodologies and frameworks for conducting risk analyses exist [17]. Virtually all of today's existing approaches use qualitative metrics to assess risks. Quantification is regarded an important issue but due to many challenges in this field there is, to our knowledge, no methodology for our purpose up until now. Thus we are

going to use a qualitative scale. The concept of Annual Loss Expectancy has a long tradition in risk management and is the basis for the Return on Security Investment.

[2] and [4] pointed out that a holistic cost-benefit evaluation of security investments should also consider the motivation and possible return for an attacker.

### 3 Methodology

To correctly and completely assess risks in e-commerce or internet transactions, one has to follow a structured approach to fully comprehend and expose all relevant aspects. As shown in section 2, several procedures are known. The authors have opted for [13] with slight adaptations. Risks distinguish themselves from other events due to a loss associated with the event, a measurable frequency of the occurrence, and by a chance to change the outcome of the event. Consequently we are going to divide the holistic process of accessing a resource into separate steps, distinguishing between several forms of implementation for different architectures. We measure the impact for the stakeholders and the according frequency. Finally, we evaluate architectural decisions on their impact and suggest, based on the risk assessment on the pros and cons, the usage of an AAI. The Return on Security Investment (RoSI) is used to economically justify an AAI usage.

Let  $l_i$  be the frequency of a successful attack on  $i$  in one year.  $L_i$  is defined as the expected loss for  $i$  in the case of a successful attack. Consequently, the Annual Loss Expectancy for  $i$  is defined as

$$ALE_i = l_i \cdot L_i \quad (1)$$

### 4 AAI Architectures

Usually, the usage of AAI is motivated from a software engineering point of view – outsourcing non functional activities into an infrastructure [16], and from an economic point - outsourcing security services to concentrate on core competencies gaining competitive advantages and raising the security level through third party know-how [15]. Referring back to Fig. 1 we see that several security services build on each other to compute an access decision. Again, taken the SAML and XACML termini as a guideline, we can deduct four separate steps of services: Authentication Assertion, Attribute Assertion, Policy Decision Assertion, and Policy Enforcement. One, all of them, or combinations can be outsourced by a Service Provider into an AAI.

The characteristics of an AAI can be determined with the help of the given four sub-services in combination with the two prevailing architectural paradigms. AAI are to this day build either centrally with a central database or provider in the middle or as a federation where service providers act as AAI providers themselves. The best examples of these two archetypes are of course Microsoft's centralised .NET Passport versus the distributed Liberty Identity Federation Framework.

In this paper we restrict the introduction of current AAI to four representatives, each enhancing the other by or specialising in one of the given sub-services. For a more detailed analyses see [14, 15].

#### **4.1 Microsoft .NET Passport**

Microsoft .NET Passport, although often criticised, was the first and the largest commercial AAI so far. Passport concentrates on Single Sign-On (SSO) for the user who gets his passport account with every hotmail account, using a central database to keep all client information. Passport relies heavily on the usage of cookies imitating to some extent Kerberos's ticketing functionalities. The login to a SP is redirected to Passport requiring his username and password. The SP's ID is transmitted via URL encoding enabling Passport to redirect the client and storing several cookies. At the SP a software agent is needed – the so called Passport-Manager. This software reads URL encoded data and stores additional cookies into the SP's domain permitting an access control decision. At another vendor the passport cookies are used to enable a SSO [10]. The vendor decides about access of resources using his authorisation and access control mechanism of choice. Passport is a centrally organised SSO system meaning that it only asserts the user's authentication.

#### **4.2 Liberty Alliance Identity Federation Framework (ID-FF)**

Liberty was the open source community's answer to Microsoft Passport in 2001. In Liberty a Circle of Trust (CoT) establishes a Liberty system [8]. Each partner provides the authentication for his users with his own methods while they themselves can login to all other partners with a SSO. The user authenticates at his IdP and, if he wishes, a cookie is stored under a common domain where every member hosts a server so they all can access the cookie. If a user moves to a CoT member the cookie is read, the IdP asks for appropriate authentication, and an assertion is awaited. Communication is based on the SAML protocol. A CoT has to decide on the implementations. The SAML assertions can carry any attribute the CoT agrees upon. Liberty's architecture is distributed. The IdP is not fixed like in Shibboleth or centralised like in Microsoft .NET Passport. It is possible to login at different points of the CoT thus resulting e.g. in different user names or attributes that are transferred. The identity of the user is not revealed in the process of requests and assertions. For risk assessment purposes we call Liberty identity and attribute federated.

#### **4.3 PERMIS**

The EU project PERMIS [3] is closely integrated into the target system. This can be e.g. an apache web server. Instead of using the apache security functionality PERMIS is used to derive the user's role names and a PERMIS policy used to control access. The target application is responsible for user authentication. PERMIS uses X.509 attribute certificates (AC) binding the user's distinguished name to a role. An XML policy authorises roles and targets. If a user desires access the PERMIS access control enforcement function will delegate his request to the access control decision function which determines the correctness of the AC and its compliance with the policy. If access is granted the decision is given back to the enforcement function which grants the access or not. The centrally stored ACs can contain any information an Attribute Authority has assigned. Of course different authorities can work together creating an attribute storage LDAP. The decision and enforcement functions have to be implemented into the web server at the SP.

#### 4.4 PAPI

PAPI (Point of Access to Providers of Information), developed in 2001 by RedIRIS, a Spanish research network, could be regarded as a maximised AAI. It forms a distributed access control system for digital resources accessible over an institution's intranet or the internet. The user has to authenticate at the authentication server (AS) of his home domain. As PAPI is agnostic to the form of authentication the user's domain is responsible to supply a distinguish name. After successful authentication a website is given back to the user containing all accessible digital resources. Clicking on a link, the user is redirected to the Point of Access (PoA) taking with him an encrypted key identifying the AS. The PoA fetches the resource and delivers it to the user. PAPI acts as a proxy server and handles all interaction for the associated clients and servers. Consequentially, PAPI forms the maximal AAI [1].

### 5 Risk Identification

Assets under risk are the identities of clients and service providers, attributes about resources and users, the service or the good requested, as well as the system itself. All assets are prone to loose the three major security goals: Integrity, confidentiality, and availability. [13] has shown the types of vulnerabilities one might find for hardware, software, and data. Adopting that notion, the vulnerabilities are interruption, for example via a Denial-of-Service-attack, interception of the communication, for example via a Man-in-the-Middle-attack, the modification of the asset, for example attributes granting access to the resource only if the user is over 18 could be changed to access under 16, and finally fabrication of new identities. Fabrication would occur if a bogus merchant is created luring the client to log-in with his SSO credentials.

To assess the risk of each asset we make use of the introduced frequency for a successful attack. The frequency is affected twofoldly – firstly by the technical barrier  $T$  one raises to prevent an attack and secondly by the motivation of the attacker, the so called “return of attack” –  $RoA$ .

$$l_i = f(T_i, RoA_i) \quad (2)$$

$$ALE_i = f(T_i, RoA_i) \cdot L_i \quad (3)$$

The higher  $T$  the less likely a successful attack occurs; the higher  $RoA$  the higher the attacker's motivation and the resources employed and consequently the likelier an attack. The  $RoA$  can be seen as more or less stable as a service provider per se is doing business by offering something of value. He will not stop providing services to minimise risks. However,  $T$  is completely in the hands of the service provider. Outsourcing security services to an AAI can inflect on  $T$  and therefore on the frequency of a successful attack.

If the outsourcing of security services inflects the  $ALE$  the question remains which security services should be outsourced and to what extent. Different AAI approaches and architectures are able to perform one, all, or combinations of these services. We will take each sub-service and analyse the risks associated as can be seen in Table 1. Each sub-service can be interrupted via a Denial of Service or the deletion of its data.

**Table 1.** Security sub-services with associated risks and consequences

Service	Risks for user	Risks for provider
<i>Authentication Assertion</i>	<b>Identity theft:</b> Identity is intercepted and/or misused. Provider's identity is forged and the user lured into signing-in or paying for services never to be received.	<b>Identity theft:</b> User identity is forged or intercepted resulting in delivery without access rights. The theft of the provider's identity results in a loss of reputation.
<i>Attribute Assertion</i>	<b>Attribute forging or modification:</b> If resource attributes are modified, not complete, or added, the following decision can't be trusted. It might be that access or privileges are not granted. <b>Attribute interception:</b> A bogus merchant could use the attributes to misuse credentials like a credit card number, conduct illegal profiling, or sell the information.	<b>Attribute forging or modification:</b> If user attributes are modified, not complete, or added, the following decision can't be trusted. False denies result in loss of business or user motivation to change the provider. False access can be used for fraud. <b>Attribute interception:</b> an attacker could gain secret knowledge about processes or products.
<i>Policy Decision Assertion</i>	<b>Decision forging or modification:</b> Access could be wrongly denied.	<b>Decision forging or modification:</b> Access could be wrongly denied or granted.
<i>Policy Enforcement</i>	<b>Enforcement modification:</b> Access could be wrongly denied.	<b>Enforcement modification:</b> Access could be wrongly denied or granted.

As the effect is devastating but trivial - no provider or user can conduct business – it is not shown explicitly.

### 5.1 General Implications of AAI Usage

With the usage of an AAI various changes occur in the business surrounding. For once, the potential number of customers for one provider enlarges. The number of users of an AAI that merges  $N$  service providers is at most the sum of all users (4).

$$n_{AAI} \leq \sum_{i=1}^N n_i \quad (4)$$

The technical barrier for an attack  $T$  is no longer just one single  $T_i$  but has to be seen as a combination of all barriers for the given sub-services, each potentially outsourced:  $T_i^{AuthN}$  - for the Authentication,  $T_i^{Attrib}$  - for the Attribute Services,  $T_i^{PD}$  - for the Policy Decision, and  $T_i^{PE}$  - for the Policy Enforcement.  $T_i$  can't be computed by the sum of all barriers but is determined by the minimum: the weakest link in the chain determines its overall strength (5).

$$T_i = \min(T_i^{AuthN}, T_i^{Attrib}, T_i^{PD}, T_i^{PE}) \quad (5)$$



## 5.2 AAI Architectures and Their Implications

If using an AAI like Microsoft's .NET Passport the authentication of the user is relayed to Passport. The provider uses Passport's technical barrier to prevent misuse of the authentication sub-service for his business. His  $ALE_p$ , consequently, depends on the following equation (6):

$$ALE_i = f(\min(T_{AAI}^{AuthN}, T_i^{Attrib}, T_i^{PD}, T_i^{PE}), RoA_i) \cdot L_i \quad (6)$$

Using PERMIS  $T^{Attrib}$  and  $T^{PD}$  depend on the AAI.  $T^{AuthN}$  has to be managed by the SP or another AAI providing SSO. The enforcement needs to be handled by the target system.

For one single provider the loss and supplied return of attack stays the same. However, in the case of a fully developed AAI – like in PAPI – where all security services are outsourced and the AAI provider acts as a proxy for all  $N$  service providers, a successful attack on one security service results in a breach of all  $N$  vendors.  $T_i$  is substituted by  $T_{PAPI}$ . The AAI resembles a middleman. Consequently, the  $RoA$  is the sum of all returns (7).

$$ALE_i = f(T_{PAPI}, \sum_{i=1}^N RoA_i) \cdot L_i \quad (7)$$

(6) is true if the barrier  $T$  is set by one AAI provider like Passport, PERMIS, or PAPI. However, if the AAI is distributed like the Liberty ID-FF the technical barrier can't be estimated as easily. As  $N$  SPs act also as identity and attribute providers for other SPs in a CoT and use their own means of authentication the notion of the weakest link once more takes effect (8):

$$ALE_i = f(\min(\min(T_1^{AuthN}, \dots, T_N^{AuthN}), \min(T_1^{Attrib}, \dots, T_N^{Attrib}), T_i^{PD}, T_i^{PE}), RoA_i) \cdot L_i \quad (8)$$

Please note that although no AAI is introduced in detail here having a federated policy decision of this type is also possible.

## 6 Towards Risk Assessment in AAI

To correctly assess the risk of the usage of an AAI the authors make use of a qualitative method. As the Annual Loss Expectancy  $ALE$  in an AAI for  $SP_i$  is, with the exception of a proxy approach, independent of his  $L_i$  and  $RoA_i$ , one can narrow the effect to the technical barrier of the security sub-service (5), (6). The technical barriers of the four sub-services, when provided by  $SP_i$  himself, are taken as the normalised value. The outsourced value  $T_{AAI}$  is either more (+), less (-), or equal (~). Table 2 states the risk assessment. A distinction is being made if the sub-service is federated or centralised.

**Table 2.** Risk assessment for service providers in AAI

	$T_{AAI, \text{centralised}}$	$T_{AAI, \text{federated}}$
$T^{\text{AuthN}}$	+: Although $n_{AAI}$ exceeds $n_i$ , the potential of granting SSO with a hard password in a controlled environment argues for a stronger authentication. The usage of complex identification methods like a PKI is more preferable.	-: As $SP_i$ is in no control of all other SP the weakest link in the chain dictates the barrier for identity theft and alike. A controlled, standardised approach for each SP is not mandatory.
$T^{\text{Attrib}}$	+: Merging all attributes balances modified or forged information. With a pattern of the user's behaviour suspicious behaviour can be detected.	+: Same as centralised approach.
$T^{\text{PD}}$	+: Centralised policy decision enables complex, flexible, and specialised access control like XACML policies through synergies and a broader information base.	~/-: As policy decision making has to be provided by every SP no synergies can be utilised. The weakest method sets the highest barrier for attacks.
$T^{\text{PE}}$	-: The usage of a central proxy strongly affects the potential $RoA$ resulting in a higher $l_i$ (PAPI, (7)).	Not feasible.

Identity theft and fraud are the user's two main concerns in e-commerce. Assuming the SP himself is acting trustworthy, an attacker could only harm the user if a technical barrier fails. Consequently, the user has a strong interest in high security but is in no position of influencing the barriers directly. An exception has to be made as far as the authentication is concerned. Using weak passwords is making identity theft easy. SPs usually shy at demanding strong passwords or the usage of a PKI, fearing increased help desk costs or shrinking user acceptance. With a SSO these disadvantages could be reduced. However, the user has to trust the IdP not to misuse his data. The discussion about .Net Passport and the development of the Liberty ID-FF shows an interest in privacy and missing user acceptance. A user has to evaluate privacy aspects versus the ease of use through SSO and a potentially higher and transparent security system.

## 7 Methods for Deciding on AAI Usage

In section 6 we have assessed risks depending on different AAI structures and services. The question remains whether an AAI is economically useful.

To determine the cost effectiveness of security investments the RoSI approach has been widely accepted.  $ALE_{old}$  depicts the expected loss without additional security investment.  $C$  are security costs to reach  $ALE_{new}$ ,  $R$  is additional revenue in cause of the membership in an AAI Federation, e.g. through wider adoption of the service, a larger customer base, or a better corporate image.

$$ALE_{old} - ALE_{new} - C + R = RoSI \quad (9)$$

For economic reasons the RoSI must be at least positive. Taking into consideration that  $ALE$  is defined by the weakest point of the security sub-services  $T^{\min}$  and that the cost for sub-service  $x$  at  $SP_i$  is  $c_i^x$  we can deduct two reasons for outsourcing security sub-services.  $r_i^x$  defines the additional revenue for  $SP_i$  when outsourcing  $x$  due to the reasons mentioned above.

First, if the outsourced sub-service  $T_i^x$  is not  $T^{\min}$  but  $T_i^x \leq T_{AAI}^x$ , then

$$\begin{aligned} ALE_{old} - ALE_{new} &= \Delta ALE = 0 \text{ from (9)} \\ -C + R &\geq 0 \rightarrow C \leq R \rightarrow c_{AAI}^x \leq c_i^x + r_i^x \end{aligned} \quad (10)$$

Meaning that if no strengthened barrier against an attack results out of the decision to use the AAI's service it can be economically reasonable to use the AAI if cost-savings are higher or additional revenue is gained for example through a larger customer base.

Second, if the outsourced sub-service  $T_i^x$  is  $T^{\min}$  and  $T_i^x \leq T_{AAI}^x$ , then

$$\begin{aligned} ALE_{old} - ALE_{new} &= \Delta ALE > 0 \text{ from (9)} \\ c_{AAI}^x &< c_i^x + r_i^x + \Delta ALE \end{aligned} \quad (11)$$

The amount to be invested in an AAI is the sum of reduced costs through outsourcing, additional revenue through a larger customer base, and the saved  $ALE$ . Please note, when changing more than one sub-service in (11) the additional revenue  $r_i^x$  is not affected proportionally.

Using AAI services does not automatically change the risk assessment of a business process. As seen in Table 2 the decision has to be carefully evaluated if additional risks are worth the enhancements. Furthermore, the decision of outsourcing doesn't have to depend on risk but can be seen as an entirely economic decision (10).

However, as shown in (11) the implication of fewer risks – or lesser  $ALE$  – motivates higher investments for the AAI usage as well as sums up to potential savings.

## 8 Conclusion and Future Work

Unfortunately, empirical data about the risks of AAI is missing. Therefore, our approach stays conceptual and follows the qualitative methods by [13]. However, our approach permits, for the first time, the analysis of risks in each sub-service in authentication, authorisation, and access control deducting formally the factors which are influencing an outsourcing decision. Exclusively motivating AAI from a technical perspective is not sufficient. It is of high importance to identify the four security sub-services for a system and measure its costs and risks. Accordingly, a service provider can decide on a suitable AAI. Next steps have to comprise the search for empirical data.

## References

- [1] Castro-Rojo, R., Lopez, D. R.: The PAPI system: point of access to providers of information. In: *Computer Networks: The International Journal of Computer and Telecommunications Networking*, Volume 37. Elsevier, Amsterdam (2001) 703-710
- [2] Cavusoglu, H., Mishra, B., Raghunathan, S.: A Model for Evaluating IT Security Investments. In: *Communications of the ACM*, Volume 47. ACM Press, New York (2004) 87-92
- [3] Chadwick, D., Otenko, A.: The PERMIS X.509 role based privilege management infrastructure. In: *Proceedings of the 7th ACM Symposium on Access Control Models and Technologies (SACMAT '02)*. ACM Press, New York (2002) 135-140
- [4] Cremonini, M., Martini, P.: Evaluating Information Security Investments from Attackers Perspective: the Return-On-Attack (ROA). In: *Proceedings of the Fourth Workshop on the Economics of Information Security*. Harvard (2005)
- [5] Jøssang, A., Pope, S.: User Centric Identity Management. In: Clark, A., Kerr, K., Mohay, G. (eds.): *Proceedings of AusCERT Asia Pacific Information Technology Security Conference 2005*. Gold Coast (2005) 77-89
- [6] Katsikas, S. K., Lopez, J., Pernul, G.: Trust, Privacy and Security in E-business: Requirements and Solutions. In: *Proc. of the 10th Panhellenic Conference on Informatics (PCI'2005)*. Lecture Notes in Computer Science. Springer-Verlag, Berlin Heidelberg New York (2005) 548-558
- [7] Kormann, P., Rubin, A.: Risks of the Passport single sign-on protocol. In: *Computer Networks: The International Journal of Computer and Telecommunications Networking*, Volume 33. Elsevier, Amsterdam (2000) 51-58
- [8] Liberty ID-FF Bindings and Profiles Specification, Liberty Alliance Project, 2003. Accessible at <http://www.projectliberty.org/specs/liberty-idff-bindings-profiles-v1.2.pdf>
- [9] Lopez, J., Oppinger, R., Pernul, G.: Authentication and authorization infrastructures (AAIs): a comparative survey. In: *Computers & Security*, Volume 23. Elsevier, Amsterdam (2004) 578-590
- [10] Microsoft Passport Review Guide. Accessible at [http://download.microsoft.com/download/a/f/4/af49b391-086e-4aa2-a84b-ef6d916b2f08/passport\\_reviewguide.doc](http://download.microsoft.com/download/a/f/4/af49b391-086e-4aa2-a84b-ef6d916b2f08/passport_reviewguide.doc)
- [11] Nowey, T., Federrath, H., Klein, C., Plössl, K.: Ansätze zur Evaluierung von Sicherheitsinvestitionen. In: *Proc. 2. Jahrestagung des GI-Fachbereichs Sicherheit*, Lecture Notes in Informatics, P-62, Köllen-Verlag, Bonn (2005) 15-26
- [12] Pfleeger, S.L.: Risky Business: what we have yet to learn about risk management. *Journal of Systems and Software*, Volume 53. Elsevier, New York (2000) 265-273
- [13] Pfleeger, C.P., Pfleeger, S.L.: *Security in Computing*. 3rd edn. Prentice Hall, New Jersey (2002)
- [14] Schlaeger, C., Nowey, T., Montenegro, J.A.: A Reference Model for Authentication and Authorisation Infrastructures Respecting Privacy and Flexibility in b2c eCommerce. In: *Proc. of the First International Conference on Availability, Reliability and Security (ARES '06)*. IEEE Computer Society, Los Alamitos (2006) 709-716
- [15] Schlaeger, C., Pernul, G.: Authentication and Authorisation Infrastructures in b2c e-commerce. In: Bauknecht, K., Pröll, B., Werthner, H. (eds.): *Proc. of the Sixth International Conference on Electronic Commerce and Web Technologies - EC-Web '05*. Lecture Notes in Computer Science, Vol. 3590. Springer Verlag, Berlin Heidelberg New York (2005) 306-315
- [16] Tanenbaum, A.S., van Stehen, M.: *Verteilte Systeme. Grundlagen und Paradigmen*. Prentice Hall, München (2003)
- [17] Vidalis, S.: A Critical Discussion of Risk and Threat Analysis Methods and Methodologies. School of Computing Technical Report CS-04-03, University of Glamorgan (2004)

# Towards a UML 2.0 Extension for the Modeling of Security Requirements in Business Processes

Alfonso Rodríguez<sup>1</sup>, Eduardo Fernández-Medina<sup>2</sup>, and Mario Piattini<sup>2</sup>

<sup>1</sup> Departamento de Auditoría e Informática, Universidad del Bio Bio,  
La Castilla S/N, Chillán, Chile  
alfonso@ubiobio.cl

<sup>2</sup> ALARCOS Research Group, Information Systems and Technologies Department,  
UCLM-Soluziona Research and Development Institute,  
University of Castilla-La Mancha, Ciudad Real, Spain  
{Eduardo.FdezMedina, Mario.Piattini}@uclm.es

**Abstract.** Security is a crucial issue for business performance, but usually, it is considered after the business processes definition. Many security requirements can be expressed at the business process level. A business process model is important for software developers, since they can capture from it the necessary requirements for software design and creation. Besides, business process modeling is the center for conducting and improving how the business is operated. This paper contains a description of our UML 2.0 extension for modeling secure business process through activity diagrams. We will apply this approach to a typical health-care business process.

## 1 Introduction

The new business scene, where there are many participants and an intensive use of communications and information technologies, implies that enterprises not only expand their businesses but also increase their vulnerability. As a consequence, with the increase of the number of attacks on systems, it is highly probable that sooner or later an intrusion can be successful [19]. This security violation causes losses. For this reason, it is necessary to protect computers and their systems in the best possible way. Best possible security does not necessarily mean absolute security, but a reasonable high security level in relation to the given limitations [25].

On the other hand, business processes are key to maintain competitiveness. Since, they are the ability of an enterprise to describe, standardize, and adapt the way it reacts to certain types of business events, and how it interacts with suppliers, partners, competitors, and customers [21].

Regardless of the importance of the security notion for companies, this is often neglected in business process models, which usually concentrate on modeling the process in a way that functional correctness can be shown [2] mainly due to the fact that the expert in the business process domain is not an expert in security [9]. Typically, security is considered after the definition of the system. This approach often leads to problems, which most of the times are translated into security

vulnerabilities [17], which clearly justify the need of increasing the effort in the pre-development phases, where fixing the bugs is cheaper [14].

If we consider that empirical studies show that it is common at the business process level that customers and end users are able to express their security needs [14], then it is possible to capture at a high level, security requirements easily identifiable by those who model business processes. Besides, requirements specification usually results in a specification of the software system which should be as exact as possible [1], since, effective business process models facilitate discussions among different stakeholders in the business, allowing them to agree on the key fundamentals and to work towards common goals [5].

For business process modeling, there are several languages and notations [8], however, UML (Unified Modeling Language) is a widely accepted standard notation. The most important change of UML 2.0 version with respect to the previous ones has been that of the activity diagrams which improve the business process representation. Our work considers a UML 2.0 extension that allows us to incorporate security requirements into activity diagrams from the perspective of the business analyst. We have considered the security requirements identified in the taxonomy proposed in [7].

Our proposal is based on the MDA (Model Driven Architecture) approach. We will define early requirements identification using UML and this will make it possible to perform independent specifications of the implementation. Moreover, we believe that it is possible to have two different perspectives about security requirements at a high level of abstraction. One of them related to business analysts and the other associated with security experts. In this paper we have deepened in the first perspective.

The structure of the rest of the paper is the following: in Section 2, we will summarize the main issues about security in business processes. In Section 3, we will present a brief overview of UML 2.0 activity diagrams and extensions. In Section 4, we will propose a UML 2.0 extension to represent security requirements. Finally, in Section 5, we will present an example and in Section 6 our conclusion will be drawn.

## 2 Security in Business Process

In spite of the importance of security for business processes, we have found out two problems. The first one is that modeling has not been adequate since, generally, those who specify security requirements are requirements engineers that have accidentally tended to use architecture specific restrictions instead of security requirements [6]. And in the second place, security has been integrated into an application in an ad-hoc manner, often during the actual implementation process [2], during the system administration phase [13] or it has been considered like outsourcing [16].

An approach to model security considering several perspectives is presented in [9]. Authors take into consideration the following perspectives: *static*, about the processed information security, *functional*, from the viewpoint of the system processes, *dynamic*, about the security requirements from the life cycle of the objects involved in the business process, *organizational*, used to relate responsibilities to acting parties within the business process and the *business processes* perspective, that provides us with an integrated view of all perspectives with a high degree of abstraction. Moreover, capturing the security requirements of a system is a hard task that must be established at the initial stages of system development, and business spruces offer a

view of business structure that is very suitable as a basis for the elicitation and specification of security requirements. Business process representations may in this way present in all stages of system development different levels of abstraction appropriate for each stage [14]. Consequently, we believe that business analysts can integrate their view about business security into the business process perspective.

On the other hand, functional security requirements tend to vary depending on the kind of application. This cannot be said about security requirements since any application at the highest level of abstraction will tend to have the same basic kinds of valuable and potentially vulnerable assets [7].

The research works related to security specifications carried out by business domain experts are; (i) scarce [2, 9, 15], (ii) oriented to transaction security [20], (iii) directly oriented to information systems in general [23] or (iv) thought for security and software engineers [16]. Moreover, several works [10, 13, 14, 24] have used UML to perform the specification of security requirements. In these works, activity diagrams have not been used to capture security requirements. However, we believe that it is possible that business analysts can express their security requirements through activity diagrams.

### 3 UML 2.0 Activity Diagrams and UML 2.0 Extensions

UML 2.0 is divided into structural and behavioral specifications. Behavior models specify how the structural aspects of a system change over time. UML has three behavior models: activities, state machines, and interactions. Activities focus on the sequence, conditions, and inputs and outputs for invoking other behaviors, state machines show how events cause changes of object state and invoke other behaviors, and interactions describe message-passing between objects that causes invocation of other behaviors [4].

Activity diagrams are the UML 2.0 elements used to represent business processes and workflows [11]. In UML previous versions, expressivity was limited and this fact confused users that did not use the orientation to objects as an approach for modeling. Now, it is possible to support flow modeling across a wide variety of domains [3]. An activity specifies the coordination of executions of subordinate behaviors, using a control and data flow model. Activities may form invocation hierarchies invoking other activities, ultimately resolving to individual actions [18]. The graphical notation of an activity is a combination of nodes and connectors that allow us to form a complete flow.

On the other hand, the Profiles package contains mechanisms that allow meta-classes from existing meta-models to be extended to adapt them for different purposes. The profiles mechanism is consistent with the OMG Meta Object Facility (MOF) [18]. UML profiles consist of Stereotypes, Constraints and Tagged Values. A stereotype is a model element defined by its name and by the base class to which it is assigned. Constraints are applied to the stereotype with the purpose of indicating limitations (e.g. pre or post conditions, invariants). They can be expressed in natural language, programming language or through OCL (Object Constraint Language).





The stereotypes derived from «SecurityRequirement» can be *added* to activity diagrams elements. Any security requirement (NR, AD, I, P or AC) can be added to activity diagram elements (see Table 1). For example, an «Integrity» requirement can be specified over data store, control flow or object flow.

«SecurityRole» and «SecurityPermissions» are related in different ways, because both can be *obtained* from the UML 2.0 element of activity diagrams (see Table 1). For example, «SecurityRole» can be obtained from activities, partitions or regions specifications, but it is not specified in an explicit way over these activity diagrams elements. «SecurityPermission» is a special case, because, permissions depending on each activity diagram element which they are related to. For example, for Actions object, Execution or CheckExecution operations must be specified (see Table 3).

**Table 1.** Security Requirements and Activity Diagram Elements

Stereotypes for secure activity specification	UML 2.0 element for containment in activity diagrams					
	Activity	Activity Partition	Interruptible Activity Region	Action	Data StoreNode	Object Flow
Nonrepudiation (NR)						✓
AttackHarmDetection(AD)	✓	✓	✓	✓	✓	✓
Integrity (I)					✓	✓
Privacy (P)		✓				
AccessControl (AC)	✓	✓	✓			
Security Role	✓	✓	✓			
SecurityPermissions				✓	✓	✓



In addition, we need the definitions of some new data types to be used in tagged value definitions. In Table 2, we will show the new data type stereotypes definitions. All new data types have been derived from the Enumeration Class.

**Table 2.** New data types





Name	Description	Values associated
SecReqType	It represents a type of security requirement. It must be specified for Non Repudiation, Attack/Harm Detection, Integrity, Privacy or Access Control.	NR, AD, I, P, AC
PerOperations	It is an enumeration for possible operations over objects in activity diagrams. These operations are related to permissions granted over the object	Execution, CheckExecution, Update, Create, Read, Delete, SendReceive, CheckSendReceive
ProtectDegree	It is an abstract level that represents criticality. This degree can be low (l), medium (m) or high (h).	l, m, h
PrivacyType	It consists of anonymity (a) or confidentiality (c).	a, c
AuditingValues	It represents different security events related to the security requirement specification in business processes. They will be used in later auditing	ElementName, SourceName, DestinationName, DateTimeSend, DateTimeReceive, Date, Time, RoleName

Next tables will show the stereotypes for secure activity specifications extensively. Each stereotype specification contains: name, base class, description, notation, constrains and tagged values.

**Table 3.** Security activity and security requirement stereotypes

Name: SecureActivity Base Class: Activity		Description: A secure activity contains security specification related to requirements, role identifications and permissions
Constrains	<ul style="list-style-type: none"><li>It must be associated at least with one SecurityRequirement <b>context</b> SecureActivity <b>inv:</b> self.SecurityRequirement-&gt;size()&gt;=1</li></ul>	
Name: SecurityRole Base Class: Actor (from UseCases)		Description: It contains a role specifications. This roles must be obtained from access control and/or privacy specifications
Constrains	<ul style="list-style-type: none"><li>The role in the security role stereotype can be derived from: Activity, ActivityPartition and/or InterruptibleActivityRegion (see Table 1)</li><li>It must be associated with an access control specification and can be associated with privacy and security permissions <b>context</b> SecurityRole <b>inv:</b> self.AccessControl -&gt; size() &gt;= 1 <b>context</b> SecurityRole <b>inv:</b> self.Privacy -&gt; size()&gt;= 0 <b>context</b> SecurityRole <b>inv:</b> self.SecurityPermission -&gt; size()&gt;= 0</li></ul>	
Name: SecurityPermission Base Class: Element (from Kernel)		Description: It contains permission specifications. A permissions specification must contain details about the objects and operations involved
Constrains	<ul style="list-style-type: none"><li>It must be associated with security role specification <b>context</b> SecurityPermission <b>inv:</b> self.SecurityRole -&gt;size()&gt;= 1</li><li>It must be associated with Actions, DataStoreNode or ObjectFlow <b>context</b> SecurityPermissions <b>inv:</b> self.Actions.size+self.DataStoreNode.size+self.ObjectFlow.size=1</li><li>It must be specified such as Objects and Operations pairs. <b>context</b> SecurityPermissions <b>inv:</b> if self.Actions-&gt;size()=1 then self.SecPerOperations="Execution" or self.SecPerOperations="Checkexecution" endif if self.Datastorenode-&gt;size()=1 then self.SecPerOperations="Update" or self.SecPerOperations = "Ceate" or self.SecPerOperations="Read" or self.SecPerOperations = "Delete" endif if self.Objectflow-&gt;size()=1 then self.SecPerOperations="Sendreceive" or self.SecPerOperations="Checksendreceive" endif</li></ul>	
Tagged Values	SecurityPermissionOperation: SecPerOperations	
Name: SecurityRequirement Base Class: Element (from Kernel)		Description: Abstract class containing security requirements specifications. Each security requirement type must be indicated in some of its subclasses.
Constrains	<ul style="list-style-type: none"><li>A security requirement must be associated with a secure activity <b>context</b> SecurityRequirement <b>inv:</b> self.SecureActivity -&gt;size()=1</li><li>The notation must be completed in the subclass specification for each security requirement. It must be used one security requirement type.</li></ul>	<b>Notation</b> 
Tagged Values	SecurityRequirementType: SecReqType	
Name	Nonrepudiation	<b>Notation</b> 
Base Class	SecurityRequirement	
Description	It establishes the need to avoid the denial of any aspect of the interaction. An auditing requirement can be indicated in Comment	
Constrains	<ul style="list-style-type: none"><li>It can be only specified in the diagram elements indicated in Table 1.</li></ul>	
Tagged Values	AvNr: AuditingValues <b>context</b> Nonrepudiation <b>inv:</b> self.AvNr="ElementName" or self.AvNr="SourceName" or self.AvNr="DestinationName" or self.AvNr="DateTimeSend" or self.AvNr="DateTimeReceive"	

**Table 4.** Stereotypes specifications for security requirements

Name	AttackHarmDetection	<div>Notation</div> <div></div>	
Base Class	SecurityRequirement		
Description	It indicates the degree to which the attempt or success of attacks or damages is detected, registered and notified. An auditing requirement can be indicated in Comment		
Constrains	It can be only specified in the diagram elements indicated in Table 1.		
Tagged Values	AvAD: AuditingValues <b>context</b> AttackHarmDetection <b>inv:</b> self.AvAD="ElementName" or self.AvAD="Date" or self.AvAD="Time"		
Name	Integrity	<div>Notation</div> <div></div>	
Base Class	SecurityRequirement		
Description	It establishes the degree of protection of intentional and non authorized corruption. The elements are protected from intentional corruption. An auditing requirement can be indicated in Comment.		
Constrains	It can be only specified in the diagram elements indicated in Table 1. The Protection Degree must be specified by adding a lower case letter according to PDI tagged value.		
Tagged Values	PDI : ProtectDegree AvI: AuditingValues <b>context</b> Integrity <b>inv:</b> self.AvI="ElementName" or self.AvI="Date" or self.AvI="Time"		
Name	Privacy	<div>Notation</div> <div></div>	
Base Class	SecurityRequirement		
Description	It indicates the degree to which non authorized parts are avoided to obtain sensitive information. An auditing requirement can be indicated in Comment.		
Constrains	It can be only specified in the diagram elements indicated in Table 1. A privacy requirement has one security role specification <b>context</b> Privacy <b>inv:</b> self.SecurityRole -> size() = 1 The Privacy Type must be specified adding a lower case letter according to Pv tagged value. If privacy type is not specified then anonymity and confidentiality are considered.		
Tagged Values	Pv: PrivacyType AvPv: AuditingValues <b>context</b> Privacy <b>inv:</b> self.AvPv="RoleName" or self.AvPv="Date" or self.AvPv="Time"		
Name	AccessControl	<div>Notation</div> <div></div>	
Base Class	SecurityRequirement		
Description	It establishes the need to define and/or intensify the access control mechanisms (identification, authentication and authorization) to restrict access to certain components in an activity diagram. An auditing requirement can be indicated in Comment.		
Constrains	It can be only specified in the diagram elements indicated in Table 1. It is valid only if it is specified at least one security role. <b>Context</b> AccessControl <b>inv:</b> self.SecurityRole -> size() >= 1		
Tagged Values	AvAC: AuditingValues <b>context</b> AccessControl <b>inv:</b> self.AvAC="RoleName" or self.AvAC="Date" or self.AvAC="Time"		

## 5 Example

Our illustrative example (see Figure 2) describes a typical business process for the admission of patients in a health-care institution. In this case, the business analyst identified the following Activity Partitions: Patient, Administration Area (which is a top partition that is divided into Admission and Accounting middle partitions), and the Medical Area (divided into Medical Evaluation and Exams).

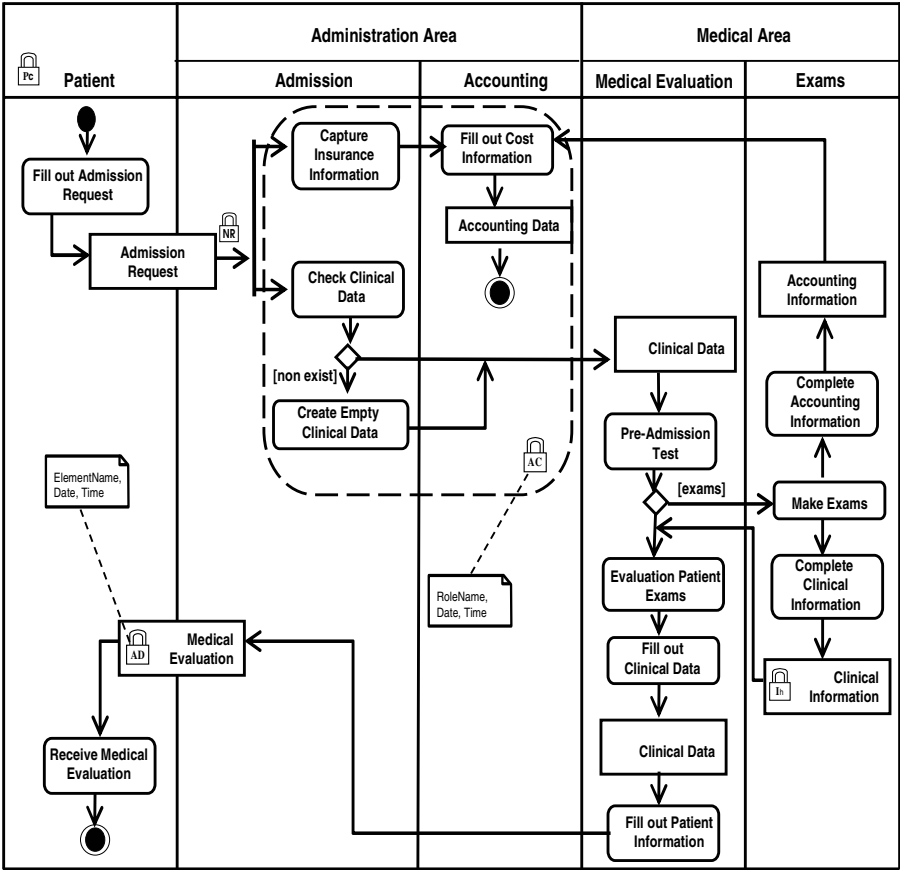


Fig. 2. Admission of Patients in a Medical Institution

The business analyst has considered several aspects of security. He/she has specified «Privacy» (confidentiality) for Activity Partition “Patient”, with the aim of preventing the disclosure of sensitive information about Patients. «Nonrepudiation» has been defined over the control flow that goes from the action “Fill Admission Request” to the actions “Capture Insurance Information” and “Check Clinical Data” with the aim of avoiding the denial of the “Admission Request” reception. «AccessControl» has been defined over the Interruptible Activity Region. A «SecurityRole» can be derived from this specification. Admission/Accounting will be a role. All objects in an interruptible region must be considered for permissions specification (see Table 5). Access control specification has been complemented with audit requirement. This implies that it must register role name, date and time of all events related to the region interruptible. Integrity (high) requirement has specified for Data Store “Clinical Information”. Finally, the business analyst has specified Attack Harm Detection with auditing requirement. All events related to attempt or success of attacks or damages are registered (names in this case are clinical information, date and time).

**Table 5.** «SecurityRole» and «SecurityPermission» specifications

Role	Permissions		
	Objects		Operations
Admission/Accounting	Action	Capture Insurance Information Fill out Cost information Check Clinical Data Create Empty Clinical Data	Execution CheckExecution Execution Execution
	DataStoreNode	Accounting Data	Update

## 6 Conclusions and Ongoing Work

The UML 2.0 version, particularly improved for business process representation through activity diagrams, opens an opportunity to incorporate security requirements that allow us to increase this aspect of the systems from early stages in software development. In this paper, we have presented a UML 2.0 extension that allows us to incorporate security requirements into activity diagrams that will increase the scope of the expressive ability of business analysts.

The next step should be that of applying an MDA approach to transform the model (including the security requirements) into most concrete models (i.e. execution models). Therefore, future work must be oriented to enrich the security requirements specifications, improving the UML extension specification to complement it with Well-Formedness Rules and OCL.

## Acknowledgements

This research is part of the following projects: DIMENSIONS (PBC-05-012-1), supported by FEDER and the “Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha”, COMPETISOFT (granted by CYTED) and RETISTIC (TIC2002-12487-E) granted by the “Dirección General de Investigación del Ministerio de Ciencia y Tecnología” (Spain).

## References

1. Artelsmair, C. and Wagner, R.; *Towards a Security Engineering Process*, The 7th World Multiconference on Systemics, Cybernetics and Informatics. Vol. VI. Orlando, Florida, USA. (2003). pp.22-27.
2. Backes, M., Pfitzmann, B. and Waider, M.; *Security in Business Process Engineering*, International Conference on Business Process Management. Vol. 2678, LNCS. Eindhoven, The Netherlands. (2003). pp.168-183.
3. Bock, C.; *UML 2 Activity and Action Models*, Journal of Object Technology. Vol. 2 (4), July-August. (2003). pp.43-53.
4. Bock, C.; *UML 2 Activity and Action Models, Part 2: Actions*, Journal of Object Technology. Vol. 2 (5), September-October. (2003). pp.41-56.

5. Eriksson, H.-E. and Penker, M., *Business Modeling with UML*, OMG Press. (2001).
6. Firesmith, D.; *Engineering Security Requirements*, Journal of Object Technology. Vol. 2 (1), January-February. (2003). pp.53-68.
7. Firesmith, D.; *Specifying Reusable Security Requirements*, Journal of Object Technology. Vol. 3 (1), January-February. (2004). pp.61-75.
8. Giaglis, G. M.; *A Taxonomy of Business Process Modelling and Information Systems Modelling Techniques*, International Journal of Flexible Manufacturing Systems. Vol. 13 (2). (2001). pp.209-228.
9. Herrmann, G. and Pernul, G.; *Viewing Business Process Security from Different Perspectives*, 11th International Bled Electronic Commerce Conference. Slovenia. (1998). pp.89-103.
10. Jürjens, J.; *Towards Development of Secure Systems Using UMLsec*, Fundamental Approaches to Software Engineering, 4th International Conference, FASE 2001 at ETAPS-2001 Genova, Italy, April 2-6, 2001, Proceedings. Vol. 2029. (2001). pp.187-200.
11. Kalnins, A., Barzdins, J. and Celms, E.; *UML Business Modeling Profile*, Thirteenth International Conference on Information Systems Development, Advances in Theory, Practice and Education. Vilnius, Lithuania. (2004). pp.182-194.
12. List, B. and Korherr, B.; *A UML 2 Profile for Business Process Modelling*, 1st International Workshop on Best Practices of UML (BP-UML 2005) at ER-2005. Klagenfurt, Austria. (2005).
13. Lodderstedt, T., Basin, D. and Doser, J.; *SecureUML: A UML-Based Modeling Language for Model-Driven Security*, UML 2002 - The Unified Modeling Language, 5th International Conference. Vol. 2460. Dresden, Germany. (2002). pp.426-441.
14. Lopez, J., Montenegro, J. A., Vivas, J. L., Okamoto, E. and Dawson, E.; *Specification and design of advanced authentication and authorization services*, Computer Standards & Interfaces. Vol. 27 (5). (2005). pp.467-478.
15. Maña, A., Montenegro, J. A., Rudolph, C. and Vivas, J. L.; *A business process-driven approach to security engineering*, 14th. International Workshop on Database and Expert Systems Applications (DEXA). Prague, Czech Republic. (2003). pp.477-481.
16. Maña, A., Ray, D., Sánchez, F. and Yagüe, M. I.; *Integrando la Ingeniería de Seguridad en un Proceso de Ingeniería Software*, VIII Reunión Española de Criptología y Seguridad de la Información, RECSI'04. Leganés, Madrid. España. (2004). pp.383-392.
17. Mouratidis, H., Giorgini, P. and Manson, G. A.; *When security meets software engineering: a case of modelling secure information systems*, Information Systems. Vol. 30 (8). (2005). pp.609-629.
18. Object Management Group; *Unified Modeling Language: Superstructure*, version 2.0, formal/05-07-04. In <http://www.omg.org/docs/formal/05-07-04.pdf>. (2005).
19. Quirchmayr, G.; *Survivability and Business Continuity Management*, ACSW Frontiers 2004 Workshops. Dunedin, New Zealand. (2004). pp.3-6.
20. Röhm, A. W., Herrmann, G. and Pernul, G.; *A Language for Modelling Secure Business Transactions*, 15th. Annual Computer Security Applications Conference. Phoenix, Arizona. (1999). pp.22-31.
21. Roser, S. and Bauer, B.; *A Categorization of Collaborative Business Process Modeling Techniques*, 7th IEEE International Conference on E-Commerce Technology Workshops (CEC 2005). Munchen, Germany. (2005). pp.43-54.
22. Stefanov, V., List, B. and Korherr, B.; *Extending UML 2 Activity Diagrams with Business Intelligence Objects*, 7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK2005). Copenhagen, Denmark. (2005).

23. Tryfonas, T. and Kiountouzis, E. A.; *Perceptions of Security Contributing to the Implementation of Secure IS*, Security and Privacy in the Age of Uncertainty, IFIP TC11 18th International Conference on Information Security (SEC2003). Vol. 250. Athens, Greece. (2003). pp.313-324.
24. Vivas, J. L., Montenegro, J. A. and Lopez, J.; *Towards a Business Process-Driven Framework for security Engineering with the UML*, Information Security: 6th International Conference, ISC 2003, Bristol, U.K. (2003). pp.381-395.
25. Zuccato, A.; *Holistic security requirement engineering for electronic commerce*, Computers & Security. Vol. 23 (1). (2004). pp.63-76.

# A Framework for Exploiting Security Expertise in Application Development

Theodoros Balopoulos, Lazaros Gymnopoulos, Maria Karyda,  
Spyros Kokolakis, Stefanos Gritzalis, and Sokratis Katsikas

Laboratory of Information and Communication Systems Security (Info-Sec-Lab),  
Department of Information and Communication Systems Engineering,  
University of the Aegean, Samos, GR-83200, Greece  
{tbalopoulos, lazaros.gymnopoulos, mka, sak, sgritz,  
ska}@aegean.gr  
<http://www.icsd.aegean.gr/Info-Sec-Lab/>

**Abstract.** This paper presents a framework that employs security ontologies and security patterns to provide application developers with a way to utilize security expertise. Through the development of a security ontology, developers locate the major security-related concepts relevant to their application context. Security patterns are then integrated with these concepts to provide tested solutions for accommodating security requirements.

## 1 Introduction

Incorporating security features in the development of applications is an issue that has been attracting the attention of both researchers and developers. To address this issue many solutions have been proposed; some of which are described in section 3 of this paper. A more detailed and complete review of security design methods for information systems was presented by Baskerville [19]. However, these solutions are either not always easily applicable in practice, or limited in scope, since many security requirements are intrinsically difficult to deal with, and software developers are not usually security experts.

We believe that the main drawback of existing security design methods is successfully brought out by M. T. Siponen who extends Baskerville's work in [20]: modern security design approaches cannot be integrated into the information systems development process. Siponen supports that the aforementioned drawback should be addressed by increased attention to the integration aspect and a shift to more socio-technical and social approaches. Therefore, incorporating security requirements in the application development process still remains an open issue. This paper proposes a framework that allows developers to make use of the available security expertise by employing ontologies and security patterns that enable the capture, articulation and reuse of designated solutions to known security issues and requirements.

The paper is structured as follows. Section two reports on the method of work followed, which resulted in the framework proposed in this paper. Section three describes the related work. Section four provides a detailed description of the



framework and section five compares the proposed framework with related approaches. Finally, the last section provides our overall conclusions and the directions for future research.

## 2 Method of Work

The framework proposed in this paper is the outcome of a research project exploring ways for the effective introduction of security attributes in the process of application development. Within the research process, security ontologies were first employed in order to explore how they can help developers better understand the application context and communicate with security experts. Results of these efforts have already been published in ([1], [2], [3]). Following this, our research indicated that security patterns would be an appropriate tool for capturing security expertise, and that this can be formalized by employing security ontologies. Thus, based on the ontologies developed, we explored the use of security patterns in the specific application contexts: we designed an appropriate structure for security patterns and a security patterns repository [4]. This paper presents a holistic framework that was constructed by bringing creatively together elements of our previous work. We believe that the proposed framework can provide a useful solution for developers, especially those involved in the development of security critical applications.

## 3 Related Approaches

In this section we describe some indicative security design approaches in order to demonstrate some of their weaknesses and to define better our own approach. This is not an exhaustive list or a complete review. An analytical review of existing methods can be found in [19] and [20].

### 3.1 UMLsec

UMLsec [5] is a standard UML extension. It allows for the incorporation of security-related information in UML diagrams and supports mechanisms that verify that the security requirements are indeed fulfilled. However, it does not provide step-by-step instructions for reaching this end. The security requirements that can be expressed and validated using UMLsec include confidentiality, integrity, secure information exchange and access control. The major UML diagrams that UMLsec builds upon are the following [6]:

- Class diagrams, which are used to assure that information exchange satisfies the security level stated in the requirements.
- Statechart diagrams, which are used to avoid covert information paths between higher and lower security level entities.
- Interaction diagrams, which are used to verify secure interaction between entities.
- Deployment diagrams, which are used to deal with the security of the physical layer.

### **3.2 The Ontology-Driven Approach to Information Security**

Raskin et al [7] advocate an ontological semantic approach to information security. Both the approach and its resources, the ontology and lexicons, are borrowed from the field of natural language processing and adjusted to the needs of the security domain. This approach pursues the following goals: (i) the inclusion of natural language data sources as an integral part of the overall data sources in information security applications, and (ii) the formal specification of the information security community know-how for the support of routine and time-efficient measures to prevent and counteract computer attacks.

### **3.3 The Tropos Approach to Modeling Security**

Mouratidis et al [8] have presented extensions to the Tropos ontology to enable it to model security issues of agent-based systems. They have introduced the concept of security constraints that allow functional, nonfunctional and security requirements to be defined together, yet being clearly distinguished. They argue that their work makes it easy to identify security requirements at the early requirements stage and propagate them until the implementation stage.

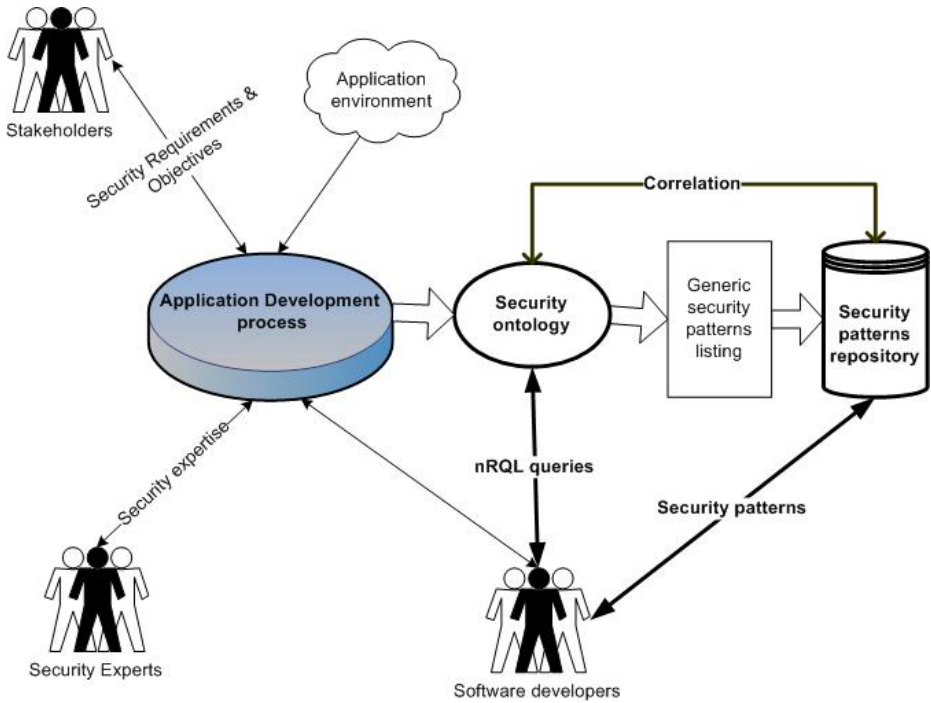
### **3.4 A Reuse-Based Approach**

Sindre et al provide in [21] a reuse-based methodology for misuse case analysis. Their methodology has two main processes: the development for reuse and the development with reuse. The “development for reuse” process actually describes the methodology for choosing which development artifacts should be reused and also how their storage should be carried out. The later includes guidelines for the construction and organization of appropriate repositories. The “development with reuse” process describes an activity diagram with five steps (Identify Assets, Determine Security Goals, Specify Threats, Analyze Risks, Specify Requirements). The authors analyze how steps 3 and 5 can be reused, while they also suggest that the first two steps could also be reusable ones.

## **4 The Framework for Secure Applications Development**

This section presents a holistic framework, depicted in Figure 1, for incorporating security characteristics and accommodating security requirements in application development.

In [2] we have proposed a methodology for developing security ontologies that can be used to support the process of applications development. In [3] we have presented the use of the developed ontologies in two different application contexts in the area of electronic government. Furthermore, in [4] we have elaborated on the use of security patterns for secure application development and presented the security patterns repository that has been developed throughout this research project.



**Fig. 1.** The framework for secure application development

The framework proposed in this paper constitutes an integrated approach that is addressed to developers that face the need for employing specialized knowledge, and helps them to make use of recorded solutions to known security issues. This holistic framework for application development builds on the use of ontologies and security patterns and combines elements from our previous work in a creative way. The contribution of this paper lies mainly in our effort to combine existing elements and describe their interrelations and interactions.

Key actors in this framework include (a) the information system stakeholders, i.e. the application users, the administrators and the management, (b) security experts whose knowledge and expertise is needed to enhance the application development process by successfully introducing security features in applications, and (c) the application developers. The latter are the ones that can use this framework for accommodating all different requirements and objectives with regard to security.

Information system stakeholders along with security experts and the software developers set the business and security objectives for the specific application. Existing security expertise is used along with the knowledge of the environment in which the specific application is going to be deployed in order to introduce environment-specific security requirements. To achieve this, the basic concepts populating the application context need to be captured and articulated; this is done through the development of the corresponding ontology. Developers can alternatively use existing ontologies or ones that have been developed for similar contexts.

## 4.1 The Security Ontology

An ontology is a description of the entities and their relationships and rules within a certain domain [9]. Ontologies have been widely used within the fields of artificial intelligence, expert systems and the semantic web, mainly for knowledge representation and sharing. Computer programs can use ontologies for a variety of purposes including inductive reasoning, classification, a variety of problem solving techniques, as well as to facilitate communication and sharing of information between different systems. Ontologies are a great tool for defining and communicating the different ways in which people perceive a specific domain. Security ontologies are ontologies covering the domain of security [10].

The Security Ontology depicted in Figure 1 aims at capturing and recording available knowledge regarding business and security objectives of a specific application development environment. The process followed for developing the security ontology, based on the method proposed in [11], was iterative and included four phases: determining competency questions, enumerating important terms, defining classes and the class hierarchy, and finally, the instantiation of the hierarchy.

The competency questions, which guided the security development process, were loosely structured security oriented questions that the developed security ontology should be able to answer. These questions were taken from typical situations developers face when confronted with security requirements. Next, the most important terms with regard to security were enumerated; the most important of them formed ontology classes; others formed properties of classes and some were not used at all.

In the next phase, the class hierarchy was developed. There are three different approaches in developing a class hierarchy: (a) the top-down approach, where one starts with the definition of the most general concepts of the domain and then goes to the more specialized ones, (b) the bottom-up approach, which starts with the definition of the most specific classes that constitute the leaves of the hierarchy while grouping of these classes into more general concepts follows, and (c) a combination of the two. To develop the security ontologies presented in [2] and [3] we followed the third of the strategies; our rich set of competency questions fitted well with the top-down approach and resulted in a class hierarchy close the final. Then the bottom-up approach was employed to fit in the remaining concepts.

To examine the rigor of the Security Ontology developed we used queries expressed in the new Racer Query Language (nRQL). This language can be directly used with databases produced by instantiated ontologies through the use of the Protégé software [12] and its Racer interface engine [13]. Further details concerning nRQL queries can be found in [14].

## 4.2 Security Patterns Repository

Patterns are characterized as solutions to problems that arise within specific contexts [15]. The concept was first used in architecture, but it gained wide acceptance in software engineering with the book “Design Patterns” [16]. The motivation behind the introduction and use of patterns can be summarized as a wish to exploit the possibility of reusability. Thus, patterns are used as “a basis to build on, and draw from, the collective experience of skilled designers” [15].

Security patterns were first introduced by Yoder and Barcalow [16] who based their work on [17]. A security pattern can be defined as a particular recurring security problem that arises in a specific security context, and presents a well-proven generic scheme for its solution [18].

In the proposed framework patterns are used for the same reasons expressed above. Moreover, having in hand the respective ontology – that is a generic description of the security context – developers can easily choose patterns that correspond to that context from a generic list of patterns.

**Table 1.** Security Patterns Comprising the Repository

Pattern Name	Description of the pattern
Authentication	This pattern allows users to access multiple components of an application without having to re-authenticate continuously. It incorporates user authentication into the basic operation of an application.
Password authentication	This pattern concerns protection against weak passwords and automated password guessing attacks.
Credentials propagation	This pattern requires that users' authentication credentials are verified by the database before access is provided.
Cryptographic storage	This pattern uses encryption for storing sensitive or security-critical data in the application.
Encrypted Communications	This pattern uses encryption for the secure transmission of sensitive or security-critical data over a network.
Session Management (protection of specific session)	This pattern provides that users cannot skip around within a series of session regarding a specific function (task) of an application. The system will not expose multiple functions but instead will maintain the current task that the users desire.
Hidden implementation	This pattern limits an attacker's ability to discern the internal workings of an application—information that might later be used to compromise the application.
Partitioned application	This pattern splits a large, complex application into two or more simpler components. Thus, dangerous privileges are restricted to a single, small component. Each component has tractable security concerns that are more easily verified than in a monolithic application
Patching	During the application lifetime, bugs and vulnerabilities are discovered; patches must be provided to address these issues.
Logging - auditing	Applications and components offer a variety of capabilities to log events that are of interest to administrators and other users. If used properly, these logs can help ensure user accountability and provide warning of possible security violations.

Employing an ontology for the specific application context enables developers to deal with security requirements more effectively. To make use of existing knowledge however, a more concrete solution is needed. Security patterns provide this solution, as they contain both the description of security issues (which can correspond to the requirements) and the indicated method or tool that addresses these issues.

Not all patterns have the same granularity or address security requirements at the same level. For designing the Security Patterns Repository depicted in Figure 1, we have adopted the categorization proposed in [18]. The different categories include:

1. Architectural patterns that refer to the high-level software development process.
2. Design patterns that refer to the medium level and refine the components of an application as well as the relationships between them.
3. Idioms are patterns at the lowest level and are related and affected by the programming language that is used each time.

In [4] we have presented a detailed description of the security patterns comprising the Repository built. Table 1 presents a detailed description of each one of them while Table 2 indicates the category they belong to.

**Table 2.** Patterns' Categorization

Pattern Name	Pattern category
Authentication	Architectural
Password authentication	Architectural
Credentials propagation	Architectural
Cryptographic storage	Design
Encrypted Communications	Design
Session Management	Idiom
Hidden implementation	Architectural
Partitioned application	Architectural
Patching	Design
Logging - auditing	Design
SandBoxing	Idiom

### 4.3 Comparison with Other Approaches

The proposed framework has the following features:

- It captures the knowledge of security experts, and aims to use it to address the needs of the software developer. Other approaches, such as [5] and [7] are not focused on the software developer, but on the security expert.
- It employs an ontology to model information. This ontology deals with objects of higher structure than other approaches (such as [7] or [8]), namely security patterns, thus being able to suggest solutions and promote reusability more effectively.
- It proposes a different instantiation of the ontology per security context. This allows it to model the fine details that a general ontology such as the one proposed in [7] is much more difficult to capture.

- It is not limited in context, unlike approaches such as [4], which is dedicated to agent-based systems.
- It can be utilized to search among the possible solutions for the one that best fits the context, unlike approaches such as [5] that are utilized to validate an already chosen solution.

## 5 Conclusions and Further Research

This paper presents a combined approach to incorporating security knowledge and expertise in the application development process. It advocates the development and employment of security ontologies, which (a) can facilitate the communication among the different parties involved, i.e. developers, security experts and the application stakeholders and (b) provide a way to capture and describe the basic security-related concepts, e.g. the security requirements the application should comply with. Moreover, the use of ontologies helps aggregate different views on the security features of the application. However, the use of ontologies has some limitations, since their construction is hard and time-consuming, and there is no standardized procedure to follow. Finally, the use of security patterns, through the creation of a repository, enables developers use standard solutions for accommodating these requirements.

Up to now, this research project has produced a set of security ontologies for similar application environments that mostly relate to electronic government, as well as a series of patterns, covering different aspects of security requirements in applications, that correspond to the ontologies. The next steps in the research process include covering different domains (e.g. the domain of health applications), as well as designing the basic mechanisms for adding functionality to the security patterns repository, and more specifically enabling their management (adding, comparing, deleting, association etc.).

## Acknowledgments

This work was co-funded by 75% from the European Union and 25% from the Greek Government, under the framework of the “EPEAEK: Education and Initial Vocational Training Program—Pythagoras”.

## References

1. Balopoulos T., Dritsas S., Gymnopoulos L., Karyda M., Kokolakis S., and Gritzalis S., “Incorporating Security Requirements into the Software Development Process”, in *Proceedings of the 4<sup>th</sup> European Conference On Information Warfare And Security (ECIW '05)*, July 2005, University of Glamorgan, UK.
2. Dritsas S., Gymnopoulos L., Karyda M., Balopoulos T., Kokolakis S., Lambrinouidakis C., and Gritzalis S., “Employing Ontologies for the Development of Security Critical Applications: The Secure e-Poll Paradigm”, in *Proceedings of the IFIP I3E International Conference on eBusiness, eCommerce, and eGovernment*, October 2005, Poznan, Poland, Springer Verlag.

3. Karyda M., Balopoulos T., Dritsas S., Gymnopoulos L., Kokolakis S., Lambrinouidakis C., and Gritzalis S., "Using Security Ontologies for the development of secure e-Government applications", in *Proceedings of the DeSeGov'06 Workshop on Dependability and Security in eGovernment (in conjunction with the 1st International Conference on Availability, Reliability, and Security)*, A. Tjoa and E. Schweighofer (Eds.), April 2006, Vienna, Austria, IEEE Computer Society Press.
4. Gymnopoulos L., Karyda M., Balopoulos T., Dritsas S., Kokolakis S., Lambrinouidakis C., and Gritzalis S., "Developing a Security Patterns Repository for Secure Applications Design" in *Proceedings of the 5<sup>th</sup> European Conference on Information Warfare and Security (ECIW '06)*, 2006, Helsinki, Finland.
5. Jurjens J. (2001), Towards development of secure systems using UMLsec, Lecture Notes in Computer Science, 2029:187.
6. Stevens P. et al (2000), Using UML, Addison-Wesley.
7. Raskin V., Hempelmann C., Triezenberg K., and Nirenburg S., "Ontology in Information Security: A Useful Theoretical Foundation and Methodological Tool", In *Proceedings of the New Security Paradigms Workshop*, V. Raskin and C. F. Hempelmann (Eds), 2001, New York, USA, ACM.
8. Mouratidis H., Giorgini P., and Manson G., "An Ontology for Modelling Security: The Tropos Project", in *Proceedings of the KES 2003 Invited Session Ontology and Multi-Agent Systems Design (OMASD'03)*, 2003, University of Oxford, United Kingdom.
9. Gruber T. R., "Toward principles for the design of ontologies used for knowledge sharing," Presented at the Padua workshop on Formal Ontology, March 1993.
10. Filman R. and Linden T., "Communicating Security Agents", In *Proceedings of the 5<sup>th</sup> Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 1996, Stanford, CA, USA, pp. 86-91.
11. Noy N. F. and Mc Guinness D. L., "Ontology Development 101: A Guide to Creating Your First Ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05. 2001.
12. Protégé, <http://protege.stanford.edu/>
13. Racer Inference Engine, <http://www.sts.tu-harburg.de/~r.f.moeller/racer/>
14. The New Racer Query Language, <http://www.cs.concordia.ca/~haarslev/racer/racer-queries.pdf>
15. Schumacher M., Fernandez-Buglioni E., Hybertson D., Buschmann F., and Sommerland P. (2006), Security Patterns: Integrating Security and Systems Engineering, John Wiley & Sons.
16. Yoder J. and Barcalow J., "Architectural Patterns for Enabling Application Security", In *Proceedings of the 4<sup>th</sup> Conference on Pattern Languages of Programs (PLoP 1997)*, 1997, Monticello, IL, USA.
17. Gamma E., Helm R., Johnson R., and Vlissides J. (1995), Design Patterns – Elements of Reusable Object-Oriented Software, Addison-Wesley Professional.
18. Schumacher M. (2003), Security Engineering with Patterns: Origins, Theoretical Models, and New Applications, Paperback.
19. Baskerville R., "Information Systems Security Design Methods: Implications for Information Systems Development", *ACM Computing Survey*, 25(4): 375-414 (1993).
20. Siponen M. T., "Analysis of modern IS security development approaches: towards the next generation of social and adaptable ISS methods", *Information and Organization*, 15(4): 339-375 (2005).
21. Sindre G., Firesmith D. G., and Opdahl A. L., "A Reuse-Based Approach to Determining Security Requirements", In *Proceedings of the 9th International Workshop on Requirements Engineering: Foundation for Software Quality (REFSQ'03)*, June 2003, Klagenfurt/Velden, Austria.



# On Diffusion and Confusion – Why Electronic Signatures Have Failed

Heiko Roßnagel

Chair of Mobile Commerce and Multilateral Security  
Johann Wolfgang Goethe- University Frankfurt,  
Gräfrstr. 78, 60054 Frankfurt, Germany  
heiko.rossnagel@m-lehrstuhl.de  
<http://www.m-lehrstuhl.de>

**Abstract.** Even seven years after the directive was enacted the market share of EC-directive conforming signature cards is disappointingly low, failing to meet any involved party's expectations. Also the lack of customers discourages companies from investing in signature products and applications. As a result almost no commercial usage for qualified electronic signatures exists. Consequently no customers seek to obtain signature products. With this contribution we examine, if economic principles are responsible for the missing adoption of qualified electronic signatures in Europe. We show that their attributes related to the rate of adoption are far from optimal. We then take a look at efforts being undertaken to increase the adoption of qualified electronic signatures. We conclude the contribution with some recommendations on how to structure a future signature market in order to speed up the diffusion process.

## 1 Introduction

In the directive 1999/93/EC of the European Parliament and of the Council [5] legal requirements for a common introduction of electronic signatures in Europe were enacted. The directive sets a framework of requirements for the security of technology used for electronic signatures. Based on certificates issued by certification authorities, which certify public keys for a person registered by a registration authority, electronic signatures can be created with a so-called "secure signature creation device" (SSCD), carrying the private keys of a person. The EC-Directive distinguishes between "electronic signatures" and "advanced electronic signatures" [5]. Certification Service Providers can issue certificates for advanced signatures that will be qualified if they meet the requirements of Annex I of the directive. Those advanced signatures with qualified certificates will be referred to in this paper as qualified signatures.

Even six years after the directive was enacted the market share of EC-directive conforming signature cards is disappointingly low [4] failing to meet any involved party's expectations. Also the lack of customers discourages companies from investing in signature products and applications. As a result almost no commercial usage for qualified electronic signatures exists. Consequently no customers seek to obtain signature products.

What are the reasons for this lack of success? In [10] several reasons for the failure have been stated. The authors concentrate on technical and legal reasons like certificate management, cross certification and liability. Economic reasons are discussed as well, but not in much detail. However, it might be promising to take a more detailed look, if economic principles are responsible for the missing adoption of qualified electronic signatures in Europe.

Technical and legal reasons are often blamed to be responsible for the disappointing market situation. One common misunderstanding is that in Germany the burden of proof of misuse is placed on the signature card holder. This is not the case. The burden of proof is placed on the recipient of the document [17]. Furthermore, despite different implementations of the EC-Directive, ranging from very strict to very liberal, electronic signatures have not taken off in any member state [4]. Therefore, legal reasons do not seem to be the major hindrance. Technical reasons have already been widely discussed. Therefore, we will argue on purely economic and social grounds and ignore technical and legal reasons for the remainder of this paper, which is structured as follows: In section 2 we will present the economic basics for our analysis. In section 3 we will take a look at the attributes of qualified electronic signatures related to their rate of adoption and in section 4 we will examine the innovation- decision process and the progress of qualified electronic signatures within this process. In section 5 we present efforts to increase the diffusion rate of qualified electronic signatures and rate their potential success. In section 6 we will present some ideas to structure a future market for qualified electronic signatures that might enjoy a higher adoption rate. Section 7 concludes our findings.

## **2 Economic Basics**

### **2.1 Technology Acceptance**

In the information systems literature a variety of theoretical perspectives have been advanced to provide an understanding of the determinants of usage. From this line of research the Technology Acceptance Model (TAM) [1] has emerged as a powerful way to explain the acceptance and adoption of technology. TAM uses two factors: the perceived ease of use and the perceived usefulness of an information system [1].

### **2.2 Diffusion of Innovations**

A second line of research has examined the adoption and usage of information from a diffusion of innovation perspective [15]. This research examines a variety of factors which are thought to be determinants of IT adoption and usage [20]. Rogers defines diffusion as “the process in which an innovation is communicated through certain channels over time among the members of a social system” and as a “special type of communication, in that the messages are concerned with new ideas” [15]. An innovation is defined as an “idea, practice, or object perceived as new by an individual or other unit of adoption” [15].

### Perceived attributes of innovations

Five attributes of innovations, as perceived by the members of the social system, determine the rate of adoption:

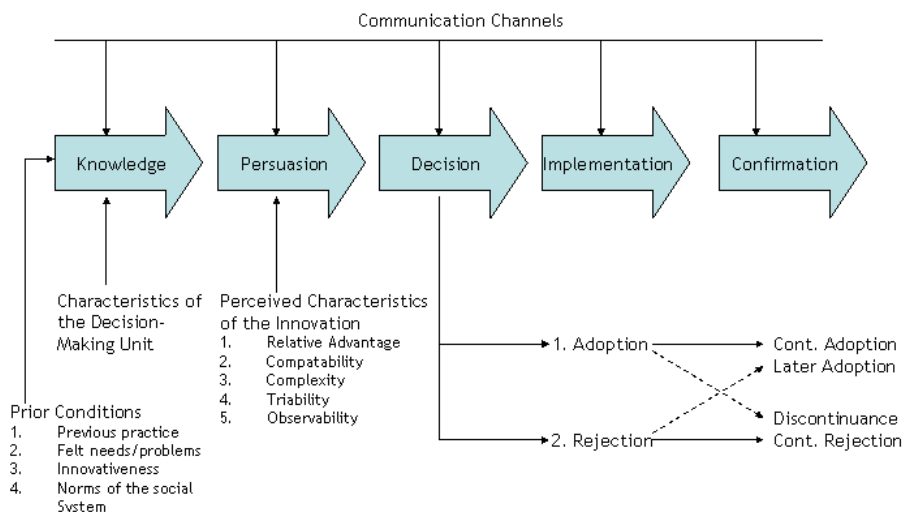
**Relative advantage** is the degree to which an innovation is perceived as better than the idea it supersedes. It is not so important if the innovation has an objective advantage, but rather if the individual perceives the innovation as advantageous. Advantages can be measured in economic terms, but social prestige, convenience, and satisfaction could also play an important role. It is analogous to the “perceived usefulness” construct in TAM [1].

**Compatibility** is the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters. An Innovation that is consistent with the existing values will diffuse more rapidly than one that is incompatible with the norms and values of the social system.

**Complexity** is the degree to which an innovation is perceived as difficult to understand and use. Innovations that are easier to understand will be adopted more rapidly than those who require the adopter to develop new skills and understandings.

**Triability** is the degree to which an innovation may be experimented with on a limited basis. New ideas that can be tried before the potential adopter has to make a significant investment into the innovation are adopted more quickly.

**Observability** is the degree to which the results of an innovation are visible to others. The easier it is for individuals to observe the results of an innovation, the more likely they are to adopt [15] [12].



**Fig. 1.** Model of the five stages in the innovation-decision process [15]

### The innovation-decision process

“The innovation-decision process is the process through which an individual passes from gaining initial knowledge of an innovation, to forming an attitude toward the innovation, to making a decision to adopt or reject, to implementation of the new idea,

and to confirmation of this decision” [15]. A model of the innovation-decision process is illustrated in figure 1.

### Adopter categories

Adopters can be classified into five categories based on their rate of innovativeness. Figure 2 shows the normal frequency distribution and the approximate percentages of the individuals included [15].

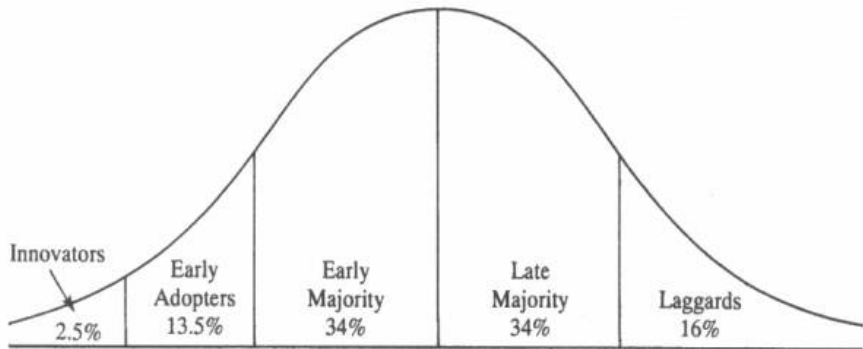


Fig. 2. Adopter categorization on the basis of innovativeness [15]

**Innovators:** Innovators play an important role in the diffusion process. They launch a new idea within the social system by importing an idea from the outside of the system boundaries. However, innovators might not be respected by other members of the social system. Innovators need the ability to understand and apply complex technical knowledge and must be able to cope with a high degree of uncertainty about an innovation at the time of adoption.

**Early adopters:** Early adopters are more integrated in the social system than innovators. This adopter category has the biggest influence and degree of opinion leadership within the system. Potential adopters look at early adopters for advice and information about an innovation. Therefore, early adopters help trigger the critical mass when they adopt an innovation.

**Early majority:** The early majority adopts innovation before the average members of a system. They do not possess a position of opinion leadership in the system, but interact with a lot of members of the social system. They are not the first to adopt an innovation, but follow with a deliberate willingness.

**Late majority:** The late majority adopts an innovation after the average member of the system. They approach innovations skeptical and cautious and adoption results because of economic necessity or increasing peer pressure.

**Laggards:** Laggards are the last members of a system to adopt. They possess almost no opinion leadership.

### Interactive innovations and network effects

An interactive innovation is an innovation that is of little use to an adopting individual unless other individuals with whom the adopter wants to communicate also adopt. Thus a critical mass of individuals has to adopt the innovation before it is of use for

the average member of the system [11]. The individuals who have adopted an innovation form a network and with each new member the overall value of the network increases [11]. This fundamental value proposition is being called network effects, network externalities, and demand side economics of scale [19]. Until a critical mass occurs in the diffusion process the rate of adoption is relatively slow [11]. After the critical mass is achieved the rate of adoption accelerates and leads to a take off in the adoption curve [13].

### **3 Attributes of Qualified Electronic Signatures Related to Their Rate of Adoption**

Having presented the economic basics, we now take a look at qualified electronic signatures and their attributes related to their rate of adoption.

#### **3.1 Relative Advantage and Perceived Usefulness**

Actually there are two ideas being superseded: manuscript signatures and electronic transactions without signatures. Qualified electronic signatures enable users to conduct legally binding contracts with relying parties that are physically at a different location at any time by communicating over the internet. However, the user is forced to make these transactions at his PC using his signature card and card reader. So while the location of the relying party becomes unimportant, the location of the user making the transaction is fixed. Therefore, qualified electronic signatures will be a supplement of manuscript signatures (when conducting transactions over the internet) and not a substitute. The perceived relative advantage will most likely be the freedom of choice with whom to conduct business, the time independence and the possibility to conduct business at home instead of the necessity to show up at a specific location as for example in dealing with public administration.

In superseding electronic transactions without signatures, qualified electronic signatures take the role of a preventive innovation. Preventive innovations are ideas that are adopted by an individual at one point in time in order to lower the probability that some future unwanted event will occur [15]. Preventive innovations usually have a very slow rate of adoption, because the unwanted event might not happen even without the adoption of the innovation. Therefore, the relative advantage is not very clear cut. Furthermore, qualified electronic signatures can only be used if they are accepted by the relying party. Therefore, the relative advantage is dependent on the size of the network of accepting parties, increasing the network effects. In order to determine the relative advantage perceived by potential adopters, it is important to take a look at the costs and benefits of qualified electronic signatures. Table 1 provides an overview of the distribution of costs and benefits.<sup>1</sup> Obviously the costs and benefits are not evenly distributed. While public administrations are the major gainers they only marginally contribute to the costs of the infrastructure. On the other hand private customers have to carry the majority of the costs, while almost not gaining any benefits. Therefore, the relative advantage will probably be perceived as very low by private customers.

---

<sup>1</sup> This only considers benefits that are not achievable without the use of electronic signatures.

**Table 1.** Distribution of costs and benefits of qualified electronic signatures [9]

	<i>Private Customers</i>		<i>Companies</i>		<i>Public Administration</i>	
	<i>Costs</i>	<i>Benefits</i>	<i>Costs</i>	<i>Benefits</i>	<i>Costs</i>	<i>Benefits</i>
<i>Electronic bid invitations</i>			■	■		■
<i>Electronic tax declaration</i>	■		■			■
<i>Access to public archives</i>	■		■	■		■
<i>Electronic elections</i>	■					■
<i>Application for public documents</i>	■					■
<i>Notifying change of residence</i>	■					■
<i>Electronic dunning procedures</i>			■	■		■
<i>Electronic marketplaces</i>	■	■	■	■	■	■
<i>Automated orderings</i>			■	■	■	■
<i>Online-Banking</i>	■		■	■	■	
<i>Alteration of contracts online</i>	■			■		
<i>Electronic billing</i>			■	■		
<i>Archiving</i>			■	■	■	■
<i>Total</i>	<i>8</i>	<i>1</i>	<i>9</i>	<i>9</i>	<i>4</i>	<i>10</i>

Table 2 shows the price strategy of the four major german trust centers. All of these trust centers are using a fixed price strategy instead of practicing price differentiation [21] for different customer groups. The prices can be regarded as being rather high if you consider that almost no applications for qualified electronic signatures exist. This leads to further reduction of the perceived usefulness.

**Table 2.** Price strategy of the four major german trust centers signatures [9]

	<b>Issue of a certificate</b>	<b>Basic fee per year</b>	<b>Sum of a 2-year usage</b>
D-Trust GmbH	<b>41 €</b>	<b>29 €</b>	<b>99 €</b>
Deutsche Post Signtrust	<b>0 €</b>	<b>39 €</b>	<b>78 €</b>
TC Trust Center	<b>8 €</b>	<b>62 €</b>	<b>132 €</b>
T-TeleSec	<b>23,57 €</b>	<b>42,95 €</b>	<b>109,47 €</b>

3.2 Compatibility

Most signature providers use a personal identification number (PIN) to authenticate the signatory. The usage of PINs has a high degree of compatibility since PINs are commonly used to authorize financial transactions for example in online banking or at Automatic Teller Machines (ATM). However, some individuals may not perceive a contract signed by means of qualified electronic signatures as a legal binding

<sup>2</sup> This offer is only for business customers.

transaction, even if this is the case. Therefore, the potential adopter should be informed about the legal consequences of using qualified electronic signatures.

### **3.3 Complexity and Ease of Use**

We cannot expect the average user to be able to understand the principles of public key cryptography [22]. This, however, might not be necessary. By using qualified electronic signatures the perceived security is rather high and a complete understanding of the underlying principles is not required. For example the use of ATMs is quite common, despite the fact that most users don't understand the underlying processes and security measures. Of course it is of utmost importance that the signature application is easy to use and to comprehend and does not allow the user to give away his private key. On the other hand, the usage of a chip card reader will likely be new to most potential adopters and installment and maintenance could lead to problems [6].

### **3.4 Triability**

With the way qualified electronic signatures are offered today, there is no triability possible. Customers are charged upfront with an initial fee and have to pay for certification services before they can create qualified electronic signatures. Therefore, potential adopters have to invest a considerable amount, before being able to test potential benefits of the innovation. However, it is possible to test electronic signatures in general by using free software like Pretty Good Privacy (PGP). But in this case different software with different look and feel, as well as a different certification structure would be tested than the one to adopt.

### **3.5 Observability**

By being able to verify the own signature the adopter can demonstrate the validity to others. However, individuals who have not obtained a qualified electronic signature themselves are not able to verify the signature leading to missing observability. Furthermore, by being a preventive innovation the unwanted prevented event, by definition, does not occur, and thus can not be observed or counted.

## **4 The State of Qualified Electronic Signatures in the Innovation-Decision Process**

So far, based on the market penetration rate of qualified electronic signatures up to now [4], we assume that only a fraction of the innovators has adopted the innovation. Furthermore, we believe that most potential adopters have not even reached the knowledge stage, meaning they are not even aware that this technology exists. A recent survey in the Czech Republik has shown that only 49% have heard the term electronic signature [3]. So far the lack of an awareness policy and missing marketing efforts, as have been undertaken for other preventive innovations like HIV prevention and seat belt usage, has hurt the diffusion process. Even worse, political signals such as allowing non-qualified electronic signatures for e-government applications are

counterproductive especially in the persuasion phase [16]. Even if potential adopters develop a favourable attitude towards qualified electronic signatures and decide to adopt, it is actually pretty hard to obtain them, because the personnel at the registration authorities is often badly informed and not aware that they even offer these products. And even for individuals who actually have adopted, the lack of applications for qualified electronic signatures and the resulting negative feedback could eventually lead to discontinuance of the innovation.

## 5 Analysis on Current Efforts to Diffuse Qualified Electronic Signatures

In the last couple of years several efforts in Europe have been started to increase the diffusion of qualified electronic signatures. Some examples are described in [14] [2] [7]. Common to these initiatives is that they focus on achieving a high penetration rate of signature cards within the complete population. As has been seen with other innovations the pure presence and availability does not necessarily lead to adoption of the innovation. One example is the German “Geldkarte”. This smart card enables small electronic payments and is included on most German ATM-cards. Despite 60 million cards being distributed in Germany only 38 million transactions have been made in 2004 (0,63 transaction per user per year) [8]. Therefore, a high penetration rate of signature cards will not automatically lead to the adoption of qualified signatures, especially if costs and benefits are not fairly distributed and prices remain as high as they are. In addition, the network for qualified signatures does not increase with the distribution of signature cards but with the adoption of the signature. So simply distributing signature cards is not enough to obtain a critical mass. One example is the Danish signature initiative OCES that started March 2003. It enables every citizen to obtain a free certificate. So far only 145.000 Danish citizens (less than 3% of the population) obtained such a certificate [7]. Therefore, it might be better to specifically target early adopters instead of trying to reach everyone. Also none of these initiatives has been able to provide some sort of triability of qualified electronic signatures.

## 6 Recommendations to Structure the Signature Market

Based on our analysis in the previous chapters we will now present some recommendations on how to structure the future market of qualified electronic signatures:

**Shift costs and benefits in order to achieve a fair distribution:** In order to increase the relative advantage of qualified electronic signatures, it is necessary to have a fair distribution of costs and benefits. Price discrimination could be used to specifically target different customer groups. Also, a new price model as proposed in [9] is necessary, that collects fees for signature verification instead of only charging the signatory, leading to reduced annual cost for the signatory. Furthermore, the acceptance of qualified electronic signatures could be increased by providing monetary benefits for its users. For example fees for public administration processes



could be omitted for users that choose to conduct these transactions online using a qualified electronic signature.

**Try to reach a critical mass:** Using dumping prices in the early phases of the diffusion process could help to reach a critical mass. These early losses can be compensated by profiting on the ensuing lock-in effects [19]. An example of such a business model is the distribution of video game consoles. Vendors of video consoles sell their product with prices below their production costs in order to increase the size of their networks and to create lock-in effects. Later on they profit from selling games to their customer base [19]. The same thing could be applied to qualified electronic signatures and the complementary product of signature verification.

**Increase the knowledge:** A large marketing campaign is essential to increase the awareness of the technology. This campaign could be financed by either the trust centers or public administration. As stated earlier the awareness of the new technology could trigger a need for it. Also, the gained benefits for public administration could finance the efforts to host such a campaign.

**Specifically target early adopters:** Early adopters are the most influential group of potential adopters. Therefore it is of utmost importance to place the product within this group in order to reach a critical mass.

**Reduce complexity:** In order to reduce complexity mobile qualified electronic signatures might be very helpful [18]. Also, for conventional signatures every effort to make the signature application as easy to use as possible, like for example including chip card reader in PCs, should be undertaken.

**Increase triability:** By, for example, issuing free 14 day certificates, certification service providers could enable potential customers to experience the product on a limited basis.

## 7 Conclusion

In this contribution we elaborated economic reasons for the slow diffusion of qualified electronic signatures. As our results show, the attributes related to the rate of adoption are far from optimal. Therefore, it is necessary to make changes how to distribute qualified signatures. Much hope has been put into efforts increasing the potential customer base, which will not necessarily lead to higher rates of adoption. In addition, we provided some possible solutions that should be undertaken in order to speed up the diffusion process and to gain a critical mass of adopters.

## References

- [1] Davis, F. D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, 13, 3, 319-340.
- [2] De Cock, D., Wouters, K. and Preneel, B. (2004) Introduction to the Belgian EID Card, in S. K. Katsikas, S. Gritzalis and J. Lopez (Eds.), *Public Key Infrastructures*, Berlin Heidelberg, Springer, 1 - 13.
- [3] Dialogin (2005) Brezen mesic internetu, <http://www.stemmark.cz/gramotnost/#f2f4a>, accessed February 28.

- [4] Dumortier, J., Kelm, S., Nilsson, H., Skouma, G. and Van Eecke, P. (2003) The Legal and Market Aspects of Electronic Signatures, Interdisciplinary centre for Law & Information Technology, Leuven.
- [5] EC-Directive 1999/93/EC. (1999) Directive 1999/93/EC of the European Parliament and of the Council on a Community framework for electronic signatures.
- [6] Fritsch, L. and Roßnagel, H. (2005) Die Krise des Signaturmarktes, Lösungsansätze aus betriebswirtschaftlicher Sicht, in H. Ferderrath (Eds.), *Sicherheit 2005*, Bonn, Köllen Druck+Verlag GmbH, 315-327.
- [7] Hvarre, J. (2004) Electronic signatures in Denmark: free for all citizens, *e-Signature Law Journal*, 1, 1, 12-17.
- [8] Koppe, V. (2005) Die Geldkarte der deutschen Kreditwirtschaft: Aktuelle Situation und Ausblick, [www.geldkarte.de](http://www.geldkarte.de), accessed February 28.
- [9] Lippmann, S. and Roßnagel, H. (2005) Geschäftsmodelle für signaturgesetzkonforme Trust Center, in O. K. Ferstl, E. J. Sinz, S. Eckert and T. Isselhorst (Eds.), *Wirtschaftsinformatik 2005*, Heidelberg, Physica-Verlag, 1167-1187.
- [10] Lopez, J., Opplinger, R. and Pernul, G. (2005) Why Have Public Key Infrastructures Failed So Far? *Internet Research*, 15, 5, 544-556.
- [11] Mahler, A. and Rogers, E. M. (1999) The diffusion of interactive communication innovations and the critical mass, *Telecommunications Policy*, 23, 719-740.
- [12] Moore, G. C. and Benbasat, I. (1991) Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation, *Information Systems Research*, 2, 3, 173-191.
- [13] Oren, S. S. and Smith, S. A. (1981) Critical Mass and Tariff Structure in Electronic Communications Markets, *The Bell Journal of Economics*, 12, 2, 467-487.
- [14] Reichl, H., Roßnagel, A. and Müller, G. (Eds.) (2005) Digitaler Personalausweis: Eine Machbarkeitsstudie, Deutscher Universitäts-Verlag, Wiesbaden,.
- [15] Rogers, E. M. (2003) Diffusion of Innovations, Free Press, New York.
- [16] Roßnagel, A. (2003) Eine konzertierte Aktion für die elektronische Signatur, *Multimedia und Recht*, 1, 1-2.
- [17] Roßnagel, A. and Fischer-Dieskau, S. (2006) Elektronische Dokumente als Beweismittel, *Neue Juristische Wochenschrift (NJW)*, 59, 806-809.
- [18] Roßnagel, H. (2004) Mobile Signatures and Certification on Demand, in S. K. Katsikas, S. Gritzalis and J. Lopez (Eds.), *Public Key Infrastructures*, Berlin Heidelberg, Springer, 274-286.
- [19] Shapiro, C. and Varian, H. R. (1999) Information Rules, Harvard Business School Press, Boston.
- [20] Taylor, S. and Todd, P. A. (1995) Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research*, 6, 2, 144-176.
- [21] R. (1985) Price Discrimination and Social Welfare, *The American Economic Review*, 75, 4, 870-875.
- [22] Whitten, A. and Tygar, J. (1999) Why Johnny Can't Encrypt: Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0, *Proceedings of the 8th USENIX Security Symposium*, August, 169-183.

# Extending P3P to Facilitate Proxies Which Pose as a Potential Threat to Privacy

Wesley Brandi and Martin S. Olivier

Information and Computer Security Architectures (ICSA) Research Group  
Department of Computer Science, University of Pretoria, Pretoria

**Abstract.** P3P allows Web sites to declare their intentions in a standard form (as a policy) in so far as privacy related matters are concerned. User agents are free to then examine P3P policies prior to engaging in normal interactions with a Web server (upon which the Web site is hosted). Unsuitable policies may result in no further interactions with the Web server. Since P3P was designed with only two parties in mind (the client and the server), the presence of a Web Proxy in the P3P framework raises privacy concerns that demand attention. What is immediately apparent is the problem of a user accessing a site with an acceptable P3P policy via a Proxy which may employ a privacy policy that is unacceptable to the user.

In this paper we discuss some of these problems within the context of a P3P environment. In discussing these problems we focus our attention on the identification of a Proxy within a P3P environment and the separation of a Proxy's policy from the policy of a site being accessed through it.

## 1 Introduction

The Platform for Privacy Preferences has been the focus of much research and criticism [1,2,3,8,10] and at its most basic level allows Web users (with their associated P3P agents) to automate the protection of their privacy. Web sites publish P3P policies clearly describing their intentions so that Web users can compare these policies to their own set of privacy preferences. Provided that the P3P policy published by the Web site is acceptable, the user may continue to make use of services offered on the Web site.

Although the way in which users and their associated P3P Agents interact with a P3P compliant Web service has received much attention, the surreptitious role played by the Web Proxy server and its impact on P3P has been neglected. In this paper we will therefore consider the scenario involving a Web Proxy that is situated between a Web user and a P3P compliant Web service.

In focusing on the role played by the Web Proxy from a P3P perspective we will discuss why it is imperative that the Web Proxy is identified as a possible privacy threat. In doing so, it will be made clear that the Web Proxy must not be excluded in so far as providing a P3P policy to the Web user. The P3P policy provided by a Proxy however, brings with it a new set of problems. We analyse these problems and discuss potential solutions.

Transparent Proxies are included in our analysis of Proxies as potential privacy threats primarily because their nature is somewhat different to that of regular Proxies and therefore deserve special attention. A Transparent Proxy functions like any other Proxy with the exception that it is not explicitly configured by a user. Usage of a Transparent Proxy is in most cases configured by a network administrator on behalf of a user. As a result, although a user may have configured his Agent to make use of a trusted Proxy, usage of the trusted Proxy may be through an unknown, untrusted and undetected Proxy i.e. the Transparent Proxy.

The problem presented by Chained Proxies presents as much of a threat to privacy as Transparent Proxies and is therefore also included in this paper. The Chained Proxy scenario arises from one Proxy acting as a client to another Proxy. A user may trust Proxy  $P_1$  and is content to access P3P services through Proxy  $P_1$ . As was the case with Transparent Proxies though, the user may be implicitly going through yet another Proxy which he does not know or trust (in the case of Proxy  $P_1$  being configured to use Proxy  $P_2$  in order to access the Web).

This paper is structured as follows: in section 2 we briefly discuss P3P and present an example that will be referred to throughout the paper. Section 3 discusses the introduction of a Web Proxy into the framework of P3P. This section will use the example introduced in section 2 to show why the Web Proxy must be considered as a threat to privacy. Section 4 discusses the problems introduced by a Proxy in more detail as well as possible solutions. This section includes a discussion of Chained Proxies, Transparent Proxies and the semantics of P3P from a Proxy perspective. This paper is then concluded in section 5 where we summarise the overall theme of our solution and highlight possible areas of future research.

## 2 P3P

P3P allows Web users to make informed decisions when determining whether usage of a Web service suits them and their needs from a privacy point of view. A Web service publishes a P3P policy via the Web which details its privacy practices. A Web user is then free to scrutinize the policy in order to determine if the privacy practices of the Web service are satisfactory. The process of perusing a P3P policy and deciding whether or not it is acceptable may be automated through the use of an agent which is familiar with the user's preferences.

In this section we define a privacy preference for a user (Bob) and the P3P policy of a Web service. Bob has adopted a somewhat draconian approach to privacy in that he has configured his browser (his P3P Agent in this case) to not allow any information relating to his internet location to be stored (in this example his IP Address serves as his internet location).

The P3P policy of the Web service that Bob will be accessing includes the `<NON-IDENTIFIABLE/>` tag which indicates that information stored by the Web service can not be traced back to the individual at all. It may be the case

that some type of anonymizing process is applied to the data. An example of the anonymizing process would be to replace the user's IP address with zeros (assuming that the user's IP address is an identifiable token to which he is bound). Due to space constraints we have not included the actual P3P policy in this paper.

P3P policies can be located via Policy Reference Files. These files are responsible for defining which P3P policies apply to certain URIs. Cranor et al [2] detail four methods that may be used in order to find the Policy Reference File that will then be used to look up the appropriate URI. Any of these mechanisms may be used in the example presented in this paper.

Since the P3P policy implemented by the Web service in our example is compliant with Bob's privacy preferences, he is able to make use of the service knowing that his privacy shall not be violated.

### 3 Dealing with Web Proxies in P3P

The P3P 1.1 Specification [2] only recognises Proxies as a cache that may be holding P3P policies belonging to the Web service that a user wishes to access. There is no discussion as to how Web Proxies should implement P3P or the way in which user agents should work with Proxies that implement P3P.

In order to consider the issues that arise when adding a Proxy to P3P we include a P3P Proxy in our example; for now this is merely a Web Proxy with a P3P policy defined for itself i.e. a P3P compliant Proxy. To minimise the impact of introducing a P3P Proxy, the associated P3P policy is accessed via a Policy Reference File which would in turn be accessed via any of the methods indicated in section 2. We discuss the finer details of a Proxy P3P policy later in this paper. The P3P Proxy we will use in this example is configured to log all Web access attempts (including the time and IP Address) indefinitely.

Assume that Bob now wishes to access a Web service that is making use of the acceptable P3P policy outlined in section 2. In accessing the Web service, Bob's P3P agent realizes that Bob's browser is configured to make use of a P3P Web Proxy (note that this is not a Transparent Proxy) which will then access the Web service on Bob's behalf. Bob's P3P agent must therefore scrutinize the policy of the P3P Web Proxy to ensure that Bob's privacy is not being violated before making use of any other service on the Web.

Since the Web Proxy's configuration stores identifiable information indefinitely, further usage of the Proxy will result in a direct conflict with Bob's privacy preferences. If Bob is serious about not wanting his IP Address logged then he will not be able to access any Web sites at all (regardless of whether or not the P3P policies on the Web sites he will be accessing are acceptable).

The consequences of introducing a third party into the P3P framework may have dire implications since P3P was designed with only two parties in mind: (1) a user on the Web with privacy preferences employing the services of (2) a service with a P3P policy on a Web server. The addition of a P3P Web Proxy can not be regarded as just an entity that needs to be scrutinized before accessing

the Web. The nature of the Proxy is such that there are many opportunities to violate the privacy preferences of an individual who uses it. For example, aside from the logging implications of a Proxy, simply caching an object may result in a violation of privacy. A cached object means an attacker using the Proxy will be able to determine whether or not a particular object on the Web has been accessed by another user of the proxy, perhaps opening the door to an inference attack of some kind.

Note that most Web Proxies issue requests to the Web for far more than just one user (consider work or academic environments). Though many users may have similar privacy preferences it is unlikely that their preferences will be exactly the same. If it is the will of the Web Proxy administrator to respect the privacy of each of the users accessing the Proxy then the administrator will either have to configure a policy that satisfies all users (whether this is feasible remains to be seen) or configure separate policies for each user (adding significant workload to the Proxy).

Alternatively, a user could choose to make use of multiple Proxies depending on his privacy expectations. It may be the case that Proxy  $P_1$  has a different privacy policy to Proxy  $P_2$ . A user could then decide to employ the services of Proxy  $P_1$  for some sites and Proxy  $P_2$  for others. Though this seems like a viable solution, this does not apply to users that have only one Proxy as a point of contact with the Web (in the case of administrators not agreeing to setup multiple Proxies).

## 4 P3P Web Proxy Problems

By including the Web Proxy in P3P, several problems arise that demand immediate attention. In this section we discuss these problems and examine ways in which they may be resolved. Briefly, the problems are as follows:

*The Semantics of P3P policies.* As mentioned in a previous section, the P3P framework was designed to address the privacy needs of a user accessing a Web service. The addition of a third party into the framework (the Web Proxy) will result in subtle changes to the semantics of P3P policies.

*Complicated Proxy policies.* Satisfying the privacy needs of each user using a Proxy may result in incredibly complex Proxy policies. Since configuring these policies will be an arduous task, we must investigate a simpler alternative.

*Transparent and Chained Proxies.* Transparent proxies are not explicitly defined to be used by the Web user. They may be the result of a sophisticated networking architecture and are therefore unseen by the Web user. Chained proxies occur when a single Proxy acts as a client to a second Proxy, this Proxy will then issue a request to yet another Proxy (or the Web) on behalf of the client Proxy. What must be investigated is how these Proxies will identify themselves to the Web user and whether or not this identification is necessary.

In considering and discussing these problems it will be evident that Web Proxies can not be ignored within the P3P architecture.

#### 4.1 P3P Semantics

The central theme of P3P centers around a client and a service. The service provides a policy, after perusal of the policy, the client may decide whether or not to continue usage of the service.

In communicating with a Proxy, a user is indirectly communicating with a Web service. It therefore makes sense that both the privacy policies of the Proxy and the Web service must be acceptable to the user before any indirect communication between the user and the Web service can begin.

A simple solution to this problem is to have the Proxy agree to implement the privacy policy of the site that is being accessed through it. If a user is content with the P3P policy of a site then it stands to reason that he should be content with the same policy being applied on the Proxy that he is using to access the site. Whilst this solution is at first attractive it has the disadvantage of the Proxy not having a say in the policy that is to be implemented. It may be the case that the Proxy does not employ the same privacy practices at all and even if it was willing to change its practices for the duration of the session, the overhead in doing so for multiple sessions and many users may be far too much to deal with.

The solution proposed in this paper allows for a Proxy to specify separate policies for individual sites. Because of the indirect means of communication between different Web services and multiple users, a Proxy must be able to provide more than one single policy which describes how it deals with all data collected. It must be able to accommodate for detailing how it intends to handle information with regards to separate entities (services) on the Web. This is best explained with an example:

Although Bob has no problems in having his details logged (both at the Proxy and at the Web service) when he accesses an online Weather service, he is not willing to compromise any privacy at all in so far as electronic payments or online banking is concerned. So, on the one hand he does not mind the Proxy keeping logs for *generic* Web access, but on the other hand he does not want any details logged at all for what he deems as *private* and *confidential*.

In order to support multiple policies we propose changes to the Policy Reference File (the `<POLICY-REF>` element in particular). The Policy Reference File, as mentioned earlier, refers to a P3P policy (or policies) and describes various attributes regarding the policy. We propose the addition of the **siteURI** parameter (as detailed in Table 1) which will typically denote the entity (Web service) for which the policy being referred to will be applied. The addition of this parameter allows a Proxy P3P policy to specify which URI the P3P policy in question refers to.

Absence of this parameter in the `<POLICY-REF>` tag of a P3P Proxy policy denotes a *generic* policy i.e. the policy that applies to all sites accessed via the Proxy. Only sites that have defined policies (via the **siteURI** parameter)

**Table 1.** An optional parameter extension to the POLICY-REF element

```
policy-ref = <POLICY-REF about="URI-reference" [siteURI="URI-reference"]>
... </POLICY-REF>
```

on the Proxy will be excluded from the *generic* policy. Essentially, the **siteURI** parameter provides a mechanism for the Proxy Policy Reference Files to indicate which P3P policies apply to certain URIs.

Unfortunately, the practical implications for Proxies that choose to make use of the **siteURI** parameter may be catastrophic. Three problems are immediately apparent:

- The Proxy may be subjected to additional stress as P3P policies are requested by users for sites that they may wish to access through the Proxy.
- There is additional overhead on the entire process of requesting an object from a Web site as the P3P policy of the Proxy for a site is looked up by the Proxy upon each initial request to a Web site.
- If an administrator can be convinced that it is imperative for the Proxy to strive towards meeting the privacy expectations of all its users then there may be considerable complexity in implementing generic proxy policies for the users, or alternatively, implementing different policies (possibly for different users) for different sites that are accessed via the Proxy.

The first problem has already been anticipated in the P3P framework and is circumvented via a simple leasing scheme through the **EXPIRY** element in the Policy Reference File. The first time a user requests a policy from the Proxy for a site then the user, in looking up the **EXPIRY** element, may gauge how long the policy will be valid for. This approach reduces the overhead of always requesting the policy from the proxy for a site each and every single time the user wishes to access the site.

A caching solution may help to alleviate stress on the Proxy in so far as addressing the second problem i.e. a Proxy can cache policies for quick retrieval at a later stage. Unfortunately, there may still be significant overhead in the initial lookup of a policy.

The third problem is not as easily tackled. We discuss this problem in detail in the next section.

## 4.2 Complicated vs Customised Proxy Policies

A silver bullet policy that addresses the privacy needs of all individuals making use of one Proxy will be a rare feat since privacy preferences of users across the Web obviously differ tremendously. Attempting to create a single privacy policy for a P3P Proxy that is a cross section of most user's privacy preferences will be an arduous task that, in most cases, will surely fail. The introduction of the **siteURI** tag into the **<POLICY-REF>** element alleviates the problem only slightly since it allows for Proxies to simplify their policies by using multiple policies for different sites.



Needless to say, there is room for improvement in addressing complex privacy policies. Large proxies may have to implement a number of privacy policies to cater for the growing demands of their users. Unfortunately, this solution may not be feasible at all for larger Proxies when one considers the sheer number of new sites that may be accessed by hundreds of thousands of users every day.

Simpler schemes may involve generic policies that are served to the general population of a Proxy and customised policies that are served to a handful of users either because they pay for the customised policy or maybe because they simply don't fit the generic mould.

### 4.3 Transparent and Chained Proxies

In order to make an informed decision in the P3P architecture, a Web user needs to be aware of all elements that may be a threat to his privacy. We have identified the Proxy as a potential threat to the Web user's privacy. A Proxy must implement a P3P privacy policy detailing its intentions to the Web user. Transparent and Chained proxies introduce another problem to the P3P architecture. In the event of a Transparent Proxy being present, the user may not be aware of its existence therefore any decision made by the user with regards to his privacy is not an entirely informed decision. It must therefore be the responsibility of a Transparent Proxy (of any Web Proxy) to identify itself as a Proxy to the P3P Agent of a Web user.

One mechanism that could be used for identification is that of the Proxy detecting when a P3P policy of a site is accessed through it and injecting its own policy into the policy of the site that is sent back to the user. This approach has two benefits. The first being that of the actual identification and the second benefit of having saved the user a trip to the Proxy to lookup the appropriate Proxy policy. A disadvantage of this approach is the overhead incurred in having to monitor all traffic accessed through the Proxy.

The identification process we propose may take place through any of the means described in the P3P Specification [2]: HTTP Response Headers, embedded link tags, etcetera. In the case of HTTP Response Headers, we propose the addition of an optional field to the P3P header, the **proxy-policy-ref-field**:

**Table 2.** Optional addition to the P3P Header

[proxypolicyref= URI-reference]
---------------------------------

The addition of the **proxy-policy-ref-field** in the header (inserted by the Proxy) will point to the URI of the P3P Policy implemented by the Proxy. Not only does the addition of this field serve as a means that can be used by the Proxy to identify itself as a part of a Web based transaction, but it also points to the P3P policy published by the Proxy.

In addition to the optional field, we propose the addition of a new element to the P3P vocabulary to enhance privacy policies for P3P Proxies. The addition

of the `<PROXY>` element as one of the root elements (an element in the `<META>` namespace - the same namespace used by `<POLICY-REF>` namespace) will allow P3P agents to immediately identify the P3P policy as one belonging to a P3P Proxy.

In identifying the policy as a policy of a P3P Proxy, the `<PROXY>` element will also give details as to any other Proxies (or sources) that may be used in retrieving Web objects. These may be Chained or Transparent proxies. Table 3 describes the `<PROXY>` element.

**Table 3.** The PROXY element

```
proxy = <PROXY about="URI-reference" [transparent=true/false]>
  [<PROXYSOURCE siteURI="URI-reference" [transparent=true/false]/>]
</PROXY>
```

The `<PROXY>` element contains an optional parameter denoting whether or not the Proxy itself is a Transparent Proxy. Elements within the `<PROXY>` element may refer to other Proxies that are used by the current Proxy. It may be the case that in accessing a site, the current Proxy will make use of another Proxy. The user will now be made aware of this process and will be able to query the privacy policy of the other Proxy used.

We allow the Proxy to specify which Proxy will be used when accessing a site by adding another extension to the `<POLICY-REF>` element. The optional `proxySource` parameter in the `<POLICY-REF>` element will typically refer to one of the `<PROXYSOURCE>` elements defined in the `<PROXY>` namespace.

**Table 4.** An optional parameter extension to the POLICY-REF element

```
policy-ref = <POLICY-REF about="URI-reference" [siteURI="URI-reference"]
  [proxySource="URI-reference"]>
  ...
</POLICY-REF>
```

These simple additions to P3P allow Proxies to achieve two important objectives that must be realised in an effort to minimise the impact of Proxies as a threat to privacy:

1. Proxies can identify themselves to a Web user, even in the case of the Proxy being transparent.
2. Proxies can list any additional Proxies that may be used (via the `<PROXYSOURCE>` element) in addition to when each of these Proxies will be used i.e. for which sites they will be used (from within the `<POLICY-REF>` element).

What must receive further attention in future research is a mechanism for handling policies that are inaccessible to the Web user. Consider the following scenario: Proxy  $P_1$  uses Proxy  $P_2$  to access a Web service. Proxy  $P_1$  has a P3P policy

and Proxy  $P_2$  has a P3P policy. In accessing Proxy  $P_1$ , a Web user is made aware of the Chained Proxy framework that Proxy  $P_1$  is a part of. The user knows that Proxy  $P_1$  uses Proxy  $P_2$  to access Web services, it is therefore essential that he scrutinizes Proxy  $P_2$ 's privacy policy before making use of any Web services.

But what will the outcome be if the user is unable to access Proxy  $P_2$ 's privacy policy (possibly due to firewall restrictions)? There may be a solution to this problem by including a cached copy of Proxy  $P_2$ 's policy on Proxy  $P_1$ . This approach however may have serious performance implications when several Proxies are used in succession, which Proxy policies should be stored where, and for how long?

## 5 Conclusion

In discussing the P3P Proxy, this paper has centered itself around two themes: Identification and Separation. It must be the responsibility of the Web Proxy to identify itself to the Web user as a Web Proxy. Identification is possible through the proposed optional field in the HTTP Response Header.

Having addressed identification of a Proxy, one can begin to address the issues that arise with Transparent and Chained Proxies. In this paper we discussed these types of Proxies as well as the problem they present in a P3P environment. A potential solution has been proposed in the form of extending policy files to specify the nature of the Proxy being accessed as well as the nature of any additional Proxies that it may make use of.

Having identified itself, the Proxy enables users to identify policies on the Proxy that will be applied when accessing specific sites through it. This is achieved via the **siteURI** parameter of the **<POLICY-REF>** element. Absence of this parameter denotes a generic policy i.e. a policy that will be applied to all sites without policies.

In recognising the Proxy as a privacy threat we have discussed several issues that are of importance when the P3P Web Proxy is introduced; in particular, we have discussed issues relating to semantics and complex policies as well as Transparent and Chained Proxies.

This paper is by no means a complete solution to the P3P Proxy problem. Issues that require further attention in future research are as follows:

- Examining whether the burden of complex policies can be addressed as discussed in section 4.2, could a feasible solution lie in the copying of policies or perhaps a simple kind of categorisation process.
- An investigation into the implications of a Proxy caching Web objects. It may be in the best interest of the user not to have any Web objects that he has requested cached on the Proxy at all since caching copies of a Web object opens the door several types of attacks, including timing attacks [7].
- Firewalls and Gateways pose just as much of a threat to privacy within the context of P3P as Web Proxies. Can the strategies proposed in this paper also be applied to them.

- Since there are several intermediaries involved in a typical Web transaction, the potential for violation of privacy preferences extends beyond the Proxy and onto different layers of the Open Systems Interconnection Reference Model (OSI [5]). Though the approach adopted in this paper may apply to some of the intermediaries (a firewall is also on the application layer) it is not as easily adopted by intermediaries that are further down in the OSI model. Since a router operates on the network layer of the OSI model it is not a candidate for the type of approach presented in this paper. Future research would do well to investigate if it is acceptable to simply accept several types of intermediaries as static privacy threats.

## References

1. K. Coyle: P3P: Pretty Poor Privacy? A Social analysis of the Platform for Privacy Preferences, June 1999, Available via the World Wide Web at <http://www.kcoyle.net/p3p.html>
2. L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall and J. Reagle, The Platform for Privacy Preferences 1.1 (P3P1.1) Specification, W3C Working Draft, 2005
3. L. Cranor and J. Reagle: The Platform for Privacy Preferences, Communications of the ACM, 4(2), 1999, pp. 48 – 55
4. B. Davidson: A survey of Proxy cache evaluation techniques, Proceedings of the 4th International Web Caching Workshop, 1999
5. J. Day and H. Zimmermann: The OSI Reference Model, Proceedings of the IEEE, 71(12), 1983, pp. 1334 – 1340
6. T. Demuth: A Passive Attack on the Privacy of Web Users Using Standard Log Information, Privacy Enhancing Technologies, 2002 , pp. 179 – 193
7. E. Felten and M. Schneide: Timing Attacks on Web Privacy, ACM Conference on Computer and Communications Security (CCS), 2000
8. H. Hochheiser: The Platform for Privacy Preference as a Social Protocol: An Examination Within the US Policy Context, ACM Transactions on Internet Technology (TOIT), 2002
9. M. Marchiori: The HTTP header for the Platform for Privacy Preferences 1.0 (P3P1.0), Network Working Group, Internet-Draft, August 2002, Available via the World Wide Web at <http://www.w3.org/2002/04/P3Pv1-header.html>
10. R. Thibadeau: A Critique of P3P: Privacy on the Web, August 2000, Available via the World Wide Web at <http://dollar.econ.cmu.edu/p3pcritique/>

# A Systematic Approach to Privacy Enforcement and Policy Compliance Checking in Enterprises

Marco Casassa Mont<sup>1</sup>, Siani Pearson<sup>1</sup>, and Robert Thyne<sup>2</sup>

<sup>1</sup> Hewlett-Packard Labs, Trusted Systems Lab  
Bristol, UK

<sup>2</sup> Hewlett-Packard, Software Business Organisation  
Toronto, Canada

{marco.casassa-mont, siani.pearson, robert.thyne}@hp.com

**Abstract.** Privacy management is important for enterprises that handle personal data: they must deal with privacy laws and people's expectations. Currently much is done by means of manual processes, which make them difficult and expensive to comply. Key enterprises' requirements include: automation, simplification, cost reduction and leveraging of current identity management solutions. This paper describes a suite of privacy technologies that have been developed by HP Labs, in an integrated way, to help enterprises to automate the management and enforcement of privacy policies (including privacy obligations) and the process of checking that such policies and legislation are indeed complied with. Working prototypes have been implemented to demonstrate the feasibility of our approach. In particular, as a proof-of-concept, the enforcement of privacy policies and obligations has been integrated with HP identity management solutions. Part of this technology is currently under productisation. Technical details are provided along with a description of our next steps.

## 1 Introduction

Enterprises that handle identities and personal information of data subjects (i.e. customers, employees and business partners) are coming under increasing pressure to improve privacy management, both to satisfy people's expectations and to comply with privacy laws and internal policies. Ultimately, the way they manage privacy aspects has implications for their reputation and brand.

Privacy laws and guidelines, such as [13, 14], dictate that enterprises should clearly state the purposes for which they are collecting personal data and should take into account the consent (or lack of consent) given by data subjects to use their data for these purposes. In addition, personal data should be deleted once its retention is not required anymore. Openness and transparency over how data is processed, manipulated and disclosed to third parties are also key requirements. Data subjects should be notified of changes affecting the management of their personal data and they should retain a degree of control over it. Compliance to all these aspects must be monitored and any violation promptly reported and addressed.

Privacy policies are commonly used to represent and describe these privacy laws and guidelines. They express *rights* of data subjects, *permissions* over usage of personal data and *obligations* to be fulfilled. These policies must be understood and refined by enterprises, deployed in their data management processes and IT infrastructures and enforced. They need to be audited and monitored for compliance. Both *operational* and *compliance* aspects must be dealt with. Current enterprise practices to privacy management are mainly based on manual processes, good behaviours and common sense. Not only are human processes prone to failure but the scale of the problem highlights the desire for additional technology to be part of the solution. The trend towards complexity and dynamism in system configurations heightens this need for automation to ensure that privacy and security properties are maintained as changes occur, and in addition to check that privacy is delivered as expected.

Enterprises are already investing in identity management solutions to automate the management of digital identities and user profiles. Most of this information is sensitive and must be managed in a privacy-aware way. To be adopted, privacy management solutions must also leverage and be compatible with these identity management solutions.

## 2 Addressed Problem

The key problem addressed in this paper is how to automate the management of *operational* and *compliance* aspects of privacy within enterprises in a systematic way, integrated with state-of-the-art identity management solutions. Currently much is done by means of manual processes, which make them difficult and expensive to comply. The introduction of automation still requires following best practice and good behaviour. However, it can help enterprises to reduce involved costs and make the overall process simpler and more effective.

Some technologies and solutions are already available in this space (see the related work section below) but they are either ad-hoc or else loosely integrated with each other and with enterprise identity management solutions.

## 3 Our Solution: Automated and Integrated Privacy Management

The main contribution of our work is automating *operational* and *compliance* aspects of privacy in a systematic way, integrated with state-of-the-art identity management solutions.

*Operational aspects* of privacy include ensuring that use of personal data (collected by enterprises) is granted by taking into account: the stated purposes for which this data was collected; the consent given by data subjects; other customisable constraints. They also include dealing with privacy obligations that dictate expectations and duties over how to handle data – such as deleting data, notifying users, transforming data, etc. Automating the management of operational aspects includes addressing how to model, deploy and enforce privacy-aware access control policies and obligations and how to achieve this whilst leveraging existing identity management solutions (specifically, in the context of access control, user provisioning and account

management [16]). *Compliance aspects* of privacy include ensuring that data is processed and handled consistently with laws, guidelines and data subjects' expectations. It must take into account the run-time behaviour of the enterprise and check for compliance at different levels of abstraction, including internal processes, applications/systems handling personal data, identity management components, systems and platforms running these components and storing personal data. Automating the management of compliance aspects includes addressing how to model all these aspects, how to gather relevant events and information, how to check for compliance and how to provide meaningful reports highlighting compliant aspects and violations.

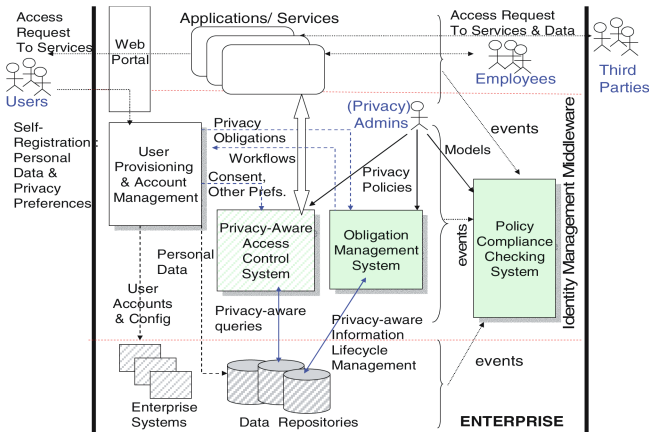
Modern identity management solutions already systematically manage identity information within enterprises [16]. They operate at a middleware level. Among other things, they allow users to self-register their personal data to enterprises and make some choices in terms of management preferences. *User provisioning* solutions deal with the creation and management of user accounts and the provisioning of related information to the involved enterprise systems; *access control* systems dictate which resources can be accessed based on access control policies and users' credentials. Our work leverages and extends these identity management capabilities to:

- Enable users to explicitly define their privacy preferences (inclusive of consent and data retention) and customise these during and after their self-registration;
- Use these users' privacy preferences (during the provisioning phase) to:
  - Configure *extended* access control systems, in order to provide privacy-aware access to personal data: this includes ensuring that these systems can keep track of stated purposes, data subjects' consent and other privacy constraints;
  - Turn parts of these privacy preferences (such as retention date, notification choices, etc.) into explicit privacy obligations on stored data to be enforced by enterprises.
- Allow enterprises to author, deploy and enforce "enterprise-side" privacy-aware access control policies and privacy obligations, derived from privacy laws and internal guidelines;
- Allow enterprises to describe via explicit models the expected behaviour of relevant enterprise processes, systems and solutions (inclusive of identity and privacy management solutions) based on privacy laws and guidelines and also to check for run-time compliance.

To achieve our aims, we have developed three R&D technologies (and related systems) to provide the required privacy management capabilities:

- **a Privacy Policy Enforcement System and an Obligation Management System** that address *operational aspects* of privacy. These two systems help enterprises to model, deploy and enforce privacy-aware access control policies and obligations with respect to managed personal data. As a significant example, we demonstrated the feasibility of integrating these systems with HP identity management solutions to handle privacy aspects;
- **a Policy Compliance Checking System** that addresses *compliance aspects* of privacy. This system helps enterprises to model privacy laws and guidelines, map them at the IT level, analyse related events and generate comprehensive compliance and violation reports.

Figure 1 shows how these technologies are integrated with relevant identity management components and how they relate to each other:



**Fig. 1.** Integration of our Privacy Management Technologies with Identity Management Solutions

The “Privacy-Aware Access Control System” component in Figure 1 is an extended access control system – via our “Privacy Policy Enforcement System” - for managing and enforcing both security-based and privacy-based access control policies. The “Policy Compliance Checking System”, amongst other things, supervises and reports on compliance aspects involving the other two privacy management systems. Our technologies are potentially self-contained and self-deployable: however only by combining them in an integrated solution we can provide a comprehensive, flexible and systematic approach to privacy management. The remaining part of this section provides more details of our technologies and their capabilities.

### 3.1 Privacy Policy Enforcement System

Privacy policies define the purposes for which data can be accessed, how to take into account data subjects’ consent and in addition the actions that need to be fulfilled at data access time, such as filtering out data, blocking access, logging, etc. Access control plays a key role in addressing these aspects. Our approach [1,3,15] to automate management and enforcement of these policies is based on a privacy-aware access control model that extends traditional access control models (based on users/groups, users’ credentials and rights, access control lists and related policies) by: (1) explicitly dealing with the stated purposes for which data is collected; (2) checking – at the access request time – the intent of requestors against these purposes; (3) dealing with data subjects’ consent; (4) enforcing additional access conditions and constraints on personal data defined by data subjects and/or enterprise administrators.



The main aspects of this model, already described in [1,3,15], are: a) a mechanism for the explicit modelling of personal data subject to privacy policies; b) an integrated mechanism for authoring privacy policies along with traditional access control policies: this is a *Policy Authoring Point (PAP)*; c) an integrated authorisation framework for deploying both *access control* and *privacy-based* policies and making related access decisions: this is an integrated *Policy Decision Point (PDP)*; d) a run-time mechanism – referred to as the “*data enforcer*” – for intercepting attempts (queries [1]) to access personal data and enforcing decisions based on privacy policies and contextual information, e.g., intent of requestors, their roles and identities, etc. This is a Policy Enforcement Point (PEP).

The “*data enforcer*” plays a key role in the process of automating the enforcement of privacy policies over personal data. At “run-time”, attempts made by users, applications and services to access personal data via queries [1] are intercepted. The data enforcer interacts with the privacy policy decision point by passing information about the request (including the intent and the types of data to be accessed) and the requestor. The privacy policy decision point makes a decision, based on available privacy policies and the context (request, requestor’s information, etc.). This decision is sent back to the data enforcer. It can be any of the following: *Access to data is denied*; *Access to data is fully granted*; *Conditional access to (part of the) data is allowed i.e. under the satisfaction of attached conditions*. Amongst other things, these conditions might require data filtering, data transformations and its manipulation. The data enforcer enforces this decision. In particular, if the decision is a “*Conditional Access*” the data enforcer might have to manipulate the query (query pre-processing) and/or transform the requested personal data (result post-processing), before returning the result to the data requestor. Data (or alternatively no data) is returned to the data requestor, based on the enforced decision.

To demonstrate the feasibility of this model, we have deployed it in a commercial identity management solution. We leveraged and extended HP Select Access [4], an HP state-of-the-art system to deal with fine-grained, policy-driven, access control management. The current commercial version of HP Select Access does not handle data as managed resources: it only deals with traditional access control policies on web resources. Based on our model, new functionalities have been added to HP Select Access in our prototype in order to explicitly deal with privacy-aware access control on personal data, as shown in Figure 2. A complete description of our extensions of HP Select Access - implemented in our prototype - is provided in [1, 3, 15].

Our recent progress in this space consists of moving our prototype and related technology towards its productisation (as a privacy extension of HP Select Access) by collaborating with the HP Software Business. We extended the capabilities of our *Data Enforcer* – for relational databases – to intercept and process a broad range of complex SQL queries, stored procedures and functions, based on stated privacy policies. We have also further extended the Policy Builder in HP Select Access to ensure that the privacy administrator can describe related privacy policies. [17] provides more details. We are also exploring and prototyping a data enforcer for LDAP data repositories – based on the same framework. Results will be published once a prototype has been fully built.

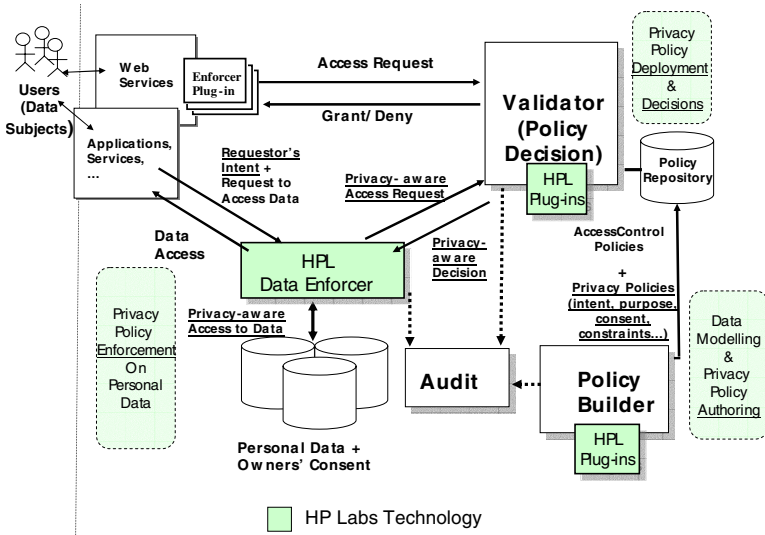


Fig. 2. Extended HP Select Access to deal with Privacy Policy Enforcement

### 3.2 Privacy Obligation Management System

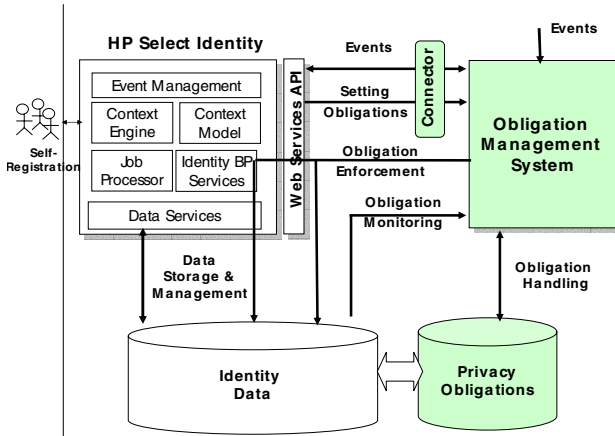
This work addresses the problem of automating the management and enforcement of privacy obligations for personal data stored by enterprises. Privacy obligations [2,12] dictate expectations and duties on how to handle personal data and deal with its lifecycle management. In our vision, the management and enforcement of privacy obligations is independent and orthogonal to the management and enforcement of privacy-aware access control policies [2]. For example, deletion of personal data has to happen independently from the fact that this data has ever been accessed. This differentiates our work from other approaches and solutions in this space (see related work).

We define an obligation management model where privacy obligations are “first class” entities [2,12]. A related *obligation management framework* is introduced to manage these privacy obligations [2,12]: (1) data subjects can explicitly define privacy preferences (e.g. on data deletion, notifications, etc.) on their personal data at the disclosure time (e.g. during a self-registration process) or at any subsequent time; (2) these preferences are automatically turned into privacy obligations; (3) enterprise privacy administrators can further associate other privacy obligations, for example dictated by laws or internal guidelines. Our *obligation management framework* handles these obligations by scheduling, enforcing and monitoring the fulfilment of privacy obligations.

Previous papers [2,12] describe the high-level architecture of our obligation management system derived from our obligation management framework. A working prototype has been fully implemented and integrated in the context of the EU PRIME

project [11], as a proof of concept, providing the specified core functionalities: scheduling, enforcement and monitoring of privacy obligations.

Our recent progress in this space consists of the integration of our obligation management system with an identity management solution, to enable privacy-aware life-cycle management of identity information, as described in Figure 1. To demonstrate how this can be achieved in a practical way, we integrated our obligation management system with HP Select Identity [5], as shown in Figure 3.



**Fig. 3.** High-level Architecture: integration of OMS with HP Select Identity

HP Select Identity [5] is a state-of-the-art solution to manage digital identities and personal data within and between large enterprises. Among other things, it automates the processes of provisioning, managing and terminating user accounts and access privileges by keeping all this information consistent and synchronised across provisioned platforms, applications and services – within and between corporate boundaries. Interactions with these third party systems (i.e. data repositories, legacy applications, services, etc.) are achieved via *Connectors*. These third parties can provide feedback to HP Select Identity (via an agent-based mechanism) about changes to their local copies of provisioned data. Changes are communicated to HP Select Identity via its Web Service API.

In our integrated prototype, HP Select Identity and our *obligation management system* interact via an ad-hoc *Connector*. As shown in Figure 3, we use HP Select Identity self-registration and user provisioning capabilities to specify and capture (at the time data is disclosed by users) privacy constraints and preferences about how personal data should be handled. These preferences are then processed by our *Connector* and sent to the obligation management system that will transform them into privacy obligations. Privacy obligations are then scheduled, enforced and monitored by our system. We leverage the workflow and user/identity management capabilities of HP Select Identity to enforce privacy obligations. This mechanism has also been used to provision privacy preferences (such as user's consent) to the privacy-aware access control system described in the previous section.

### 3.3 Policy Compliance Checking System

This work addresses the problem of automating the assessment of compliance of privacy policies within enterprises; a similar approach applies to best practice guidelines, legislation and risk analysis. The innovative aspect of our approach is that it is model-driven, where models can be authored and modified over time.

Our system verifies whether the data processing system is strong enough to automatically execute privacy policies reliably: this involves assessment of the deployment of privacy enhancing technologies and the underlying trust, security and IT infrastructure. We aim to allow enterprises to check the trustworthiness of their system components, as well as those of their business partners to whom they may transfer personal data. For example, a service may be considered trustworthy if it has been accredited by an independent privacy inspector (such as BBBOnLine or TRUSTe), or a platform may be considered trustworthy if it is judged to be in a trusted state and is compliant with standards produced by the Trusted Computing Group.

In order to automate privacy compliance the system assesses the extent to which IT controls (including privacy-enhancing technologies, such as our privacy policy enforcement system and privacy obligation management system) satisfy key privacy principles or goals. To do this the system uses a model that cascades and refines top-level properties down to specific requirements that technologies can analyse, enforce and report on. An example of technological control influence on a high level goal would be the following: a privacy related goal an enterprise could face is that data is only used for the purposes for which it was collected. This can be satisfied by the sub-goal that the enterprise uses a control that enforces role based access, where roles are associated with processes like marketing or customer support. In addition, the system should check that the control is configured correctly, the control is available, the control has not been subverted and there is proper separation of the duties defined for specific roles. There can be a many-many mapping between the goals and sub-goals: for example, it may be necessary to satisfy a combination of sub-goals in order to satisfy a higher level goal.

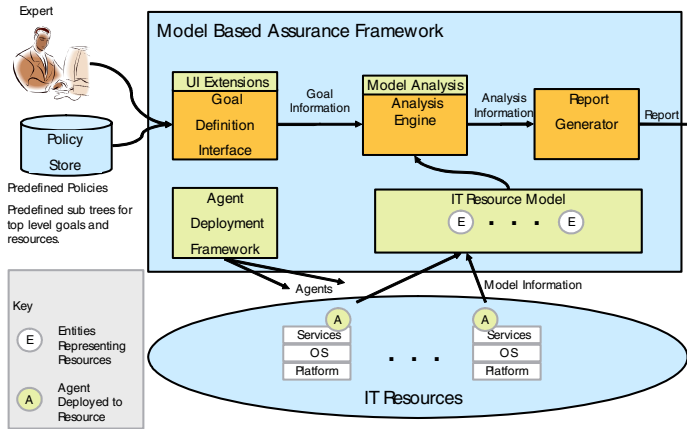
The architecture of our system is shown in Figure 4. A full working prototype, based on this architecture, has been implemented.

This system examines distributed system configurations using an agent infrastructure deployed across IT resources, feeds the findings into a reasoning engine and reports the resulting findings in a tree-like structure that can be 'drilled down' to the level of detail required. It uses functional decomposition to model privacy and model-based reasoning to carry out the analysis and generate reports. More specifically, modelling of privacy goals is combined with modelling of organisation resources and the processes around these resources. If desired, semantic web technology can be used to create a common understanding of lower level checks that are carried out.

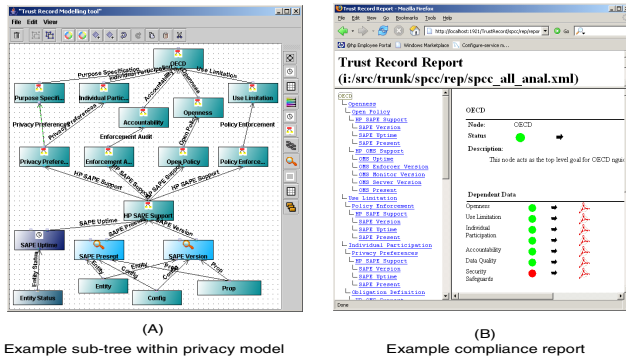
The system is intended to be used in the following way: first of all, predefined policy sub-trees would be input into our editing tool (shown in Figure 5 (A)) by a privacy expert to form a generic privacy model (this only needs doing once, but can be updated subsequently). For each specific system on which the compliance checker is to be run, a privacy officer and/or specialised administrator would tune the constraints

and deploy the system. Next, agents would be deployed to resources based on information given in the model, and would gather information over a selected time period.

Whenever desired, analysis could be triggered and a corresponding report generated: an example is shown in Figure 5 (B).



**Fig. 4. System Policy Compliance Checking: Architecture**



**Fig. 5.** (A) Example sub-tree within privacy model and (B) Example of compliance report

We focused our prototype on checking that the other two privacy management systems described in this paper are operating as expected although we also extended our models to include a much broader range of controls. Figure 5 (A) shows an example of the tool we developed to enable definition, input and customisation of models that refine and transform privacy policies from high level statements to something that can be executed automatically at a lower level. In this example, the OECD principles [14] for fair information usage were taken as the top layer within the model, there is an intermediate layer of information analysis nodes and a lower layer of technological input. In Figure 5 (A), the model focuses on assessing the deployment of the privacy policy enforcement system described in the previous section (SAPE stands for “Select

Access Privacy Enforcer”). We also developed other models, including analysis of a range of privacy and security-related IT controls and assurance information. Figure 5 (B) shows an example of the compliance report generated by our system using the model shown in Figure 5 (A). This report is targeted at company executives and auditors in order to provide information in a transparent way that can highlight areas that are a privacy concern in a dynamic and accountable way, and allow drilling down if desired to obtain further levels of detail.

A key role is played by the privacy expert(s) that is in charge of creating models. This expert must have knowledge of privacy laws, understand relevant enterprise processes, solutions and systems and author models describing the expected behaviour. It is unlikely that one person can have all this knowledge, especially in complex scenarios such as enterprises. More realistically we are looking at teams of people whose complementary knowledge can cover these aspects. In an enterprise context we believe that “auditing teams” satisfy these requirements. We are currently cooperating with an HP Internal Auditing team to explore the overall implications.

## 4 Related Work

To the best of our knowledge we are not aware of any alternative integrated and comprehensive privacy management solution covering automation of both operational and compliance privacy aspects.

Key related work, in terms of privacy-aware access control, is the Enterprise Privacy Architecture introduced and described in [18]. This approach is further refined and described in the Enterprise Privacy Authorization Language (EPAL) specification [8]. However these papers provide guidelines but do not describe how to deploy their solution within current identity management solutions. Commercial work includes IBM Tivoli Privacy Manager [6] and IBM Hippocratic databases [7]. These are vertical solutions requiring modification of the IT infrastructures in which they are deployed. They also mainly focus on data stored in relational databases. Their approach might require duplication of effort at the authoring and enforcement time.

Key differentiators of our solution are: (1) its **actual** integration with an identity management solution by extending it, without requiring duplication of administrative and management tools; (2) the fact that it is not limited to managing and enforcing privacy policies on data stored in RDBMS databases but can more generally do so for heterogeneous data repositories – including LDAP directories, virtual/meta directories, etc. These comments apply to most of the other work done on Hippocratic Databases, in addition to [7].

Paper [19] describes the concept of purpose-based access control and its application to RDBMS databases by labelling data with stated purposes within data tables. This approach is complementary to ours. Our approach does not label data and operates at a different level of abstraction: allowed purposes are defined in our privacy policies and these policies linked to relevant personal data. We will explore if we can exploit this approach by integrating it with ours – in particular in case fine-grained (at the attribute level) consent and purpose management is required.

In terms of privacy obligation management, no other significant work has been done to explicitly handle and enforce privacy obligations as we do. The EPAL [8]

language and a related Enterprise Privacy Authorisation architecture do not define obligations in detail and subordinate their enforcement to access control: this is an inadequate approach because some obligations, such as the ones involving deletion of data, are independent of access control aspects.

In terms of policy compliance checking, we are not aware of products/solutions providing this type of model-driven assurance and compliance verifications. Current products and solutions, including Synomos [9] and SenSage [10], mainly provide compliance checking process based on predefined reports (e.g. SOX-compliance report) and do not model privacy processes and IT components as we do.

The integration of our technologies with HP identity management solutions demonstrate the fact they can be deployed in real-world middleware solutions of enterprises.

Our technologies have been designed for a general purpose usage and deployment: they can be leveraged, integrated and deployed in other contexts, beyond HP identity management solutions.

## 5 Next Steps

We will further research and refine our work and related technologies. The privacy policy enforcement system will be further extended to include additional privacy constraints and more sophisticated mechanisms to process queries for additional types of data repositories (beyond RDBMS systems), such as LDAP repositories. The obligation model underpinning the privacy obligation management system needs to be further extended to be scalable and cope with large amounts of personal data. A promising research topic is to explore the management of parametric obligations that apply to a large subset of personal data subject to similar privacy preferences. The policy compliance checking system also needs to be extended in terms of modelling capabilities and to provide aspects such as data flow and a more complete assessment of the privacy enforcement technologies' ability to deliver compliance.

We will carry on our research and explore how to further exploit them in the context of HP businesses and also in the EU PRIME project [11].

## 6 Conclusions

We have described our innovative and systematic approach to address the operational and compliance requirements of privacy management, by automating core privacy management aspects: privacy-aware access control, privacy obligation management and policy compliance checking. Three related HPL R&D technologies have been implemented which can work in an integrated way by leveraging and extending state-of-the-art identity management solutions. This aspect is the main contribution of our work along with the model-based capability of our policy compliance checking system.

Working prototypes have been fully implemented to demonstrate the feasibility of our approach and integrated - as a proof-of-concept - with HP identity management solutions. The privacy policy enforcement system integrated with HP Select Access is

currently under productisation by HP Software Business Organisation. Additional work and research will be carried on both within HP Labs and in the context of the PRIME project.

## References

1. Casassa Mont, M., Thyne, R., Bramhall, P.: Privacy Enforcement with HP Select Access for Regulatory Compliance, HP Labs Technical Report, HPL-2005-10, 2005
2. Casassa Mont, M.: Dealing with Privacy Obligations in Enterprises, HPL-2004-109, 2004
3. Casassa Mont, M., Thyne, R., Chan, K., Bramhall, P. Extending HP Identity Management Solutions to Enforce Privacy Policies and Obligations for Regulatory Compliance by Enterprises - HPL-2005-110, 2005
4. Hewlett-Packard (HP): HP Openview Select Access: Overview and Features - <http://www.openview.hp.com/products/select/>, 2005
5. Hewlett-Packard (HP): HP OpenView Select Identity: Overview and Features, <http://www.openview.hp.com/products/slctid/index.html>, 2005
6. IBM Tivoli Privacy Manager: Privacy manager main web page - <http://www-306.ibm.com/software/tivoli/products/privacy-mgr-e-bus/>, 2005
7. Agrawal, R., Kiernan, J., Srikant, R., Xu, Y.: Hippocratic Databases, <http://www.almaden.ibm.com/cs/people/srikant/papers/vldb02.pdf>, IBM Almaden Research Center , 2002
8. IBM: The Enterprise Privacy Authorization Language (EPAL), EPAL 1.2 specification. <http://www.zurich.ibm.com/security/enterprise-privacy/epal/>, IBM, 2004
9. Synomos: Synomos Align 3.0, <http://www.synomos.com/>, 2005
10. SenSage: SenSage Web site, <http://www.sensage.com/>, 2005
11. PRIME Project: Privacy and Identity Management for Europe, European RTD Integrated Project under the FP6/IST Programme, <http://www.prime-project.eu/>, 2006
12. Casassa Mont, M.: Dealing with Privacy Obligations: Important Aspects and Technical Approaches, TrustBus 2004, 2004
13. Laurant, C.: Privacy International: Privacy and Human Rights 2004: an International Survey of Privacy Laws and Developments, Electronic Privacy Information Center (EPIC), Privacy International. <http://www.privacyinternational.org/survey/phr2004/>, 2004
14. OECD, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.", <http://www1.oecd.org/publications/e-book/9302011E.PDF>, 1980
15. Casassa Mont, M., Thyne, R., Bramhall, P.: Privacy Enforcement for IT Governance in Enterprises: Doing it for Real, TrustBus 2005, 2005
16. Casassa Mont, M. Bramhall, P., Pato, J.: On Adaptive Identity Management: The Next Generation of Identity Management Technologies, HPL-2003-149, 2003
17. Casassa Mont, M., Thyne, R.: Privacy Policy Enforcement in Enterprises with Identity Management Solutions, HP Labs Technical Report, HPL-2006-72, 2006
18. Karjoth, G., Schunter, M., Waidner, M. "Privacy-enabled Services for Enterprises", IBM Zurich Research Laboratory, TrustBus 2002, 2002
19. Byun, J, Bertino, E., Li, N. "Purpose based access control for privacy protection in Database Systems", Technical Report 2004-52, Purdue University, 2004



# A Generic Privacy Enhancing Technology for Pervasive Computing Environments

Stelios Dritsas, John Tsaparas, and Dimitris Gritzalis

Information Security and Infrastructure Protection Research Group  
Dept. of Informatics, Athens University of Economics and Business  
76 Patission Ave., Athens GR-10434, Greece  
{sdritsas, tsaparas, dgrit}@aueb.gr

**Abstract.** Pervasive computing is an emerging computing paradigm, which is expected to be part of our everyday life in the foreseeable future. Despite its huge potential value, one can foresee considerable drawbacks and undesirable potential uses of it in terms of privacy. In specific, the pervasive computing paradigm raises the level of the challenge to protect privacy of end-users, mainly due to the fact that devices operating in such an environment will be embedded in the fabric of the everyday life and will exhibit enhanced tracking and profiling capabilities. What is needed, inter alia, is appropriate mechanisms that are able to evolve with the needs of the users and interact with them in order to meet their privacy requirements. In this paper we suggest the foundations of a new Privacy Enhancing Technology (PET), with respect to the basic characteristics and implications introduced by pervasive environments.

## 1 Introduction

Advances in Information and Communication Technologies (ICT), especially in microprocessors, telecommunications, and sensor technology, have made Weiser's vision of pervasive computing [1] a foreseeable reality. Weiser argued that for a technology to be really ubiquitous it should become a part of the fabric of our everyday life. Thus, the main objective is to use omnipresent devices with computational and communication capabilities that function imperceptibly, unobtrusively, and which are used unconsciously by end-users.

Moreover, ubiquitous technology should contain low power and inexpensive computational devices, thus supporting diverse, autonomous, mobile, and cooperating entities through the use of the appropriate software and hardware. Pervasive Computing<sup>1</sup> refers to the emerging trend toward: numerous, casually accessible, often invisible computing devices, frequently mobile or embedded in the environment, connected to an increasingly ubiquitous network infrastructure composed of a wired core and wireless edges [2]. In this context the main characteristics of pervasive computing are:

---

<sup>1</sup> In addition to the term pervasive computing, other terms are also used to describe similar concepts (e.g. proactive/autonomic computing, ambient intelligence, nomadic computing, etc.). The term Ubiquitous Networking is herein considered as an enabling technology for the Pervasive Computing paradigm.

a) ubiquity, b) invisibility, c) sensing, d) interconnectivity and cooperation between participating devices, and e) memory amplification [3,4].

These characteristics stress the importance of privacy issues in a pervasive computing context. The most obvious issues will be given birth by the omnipresence and invisibility of the devices used in such environments. In general, the exposure of individuals will be more straightforward, as the devices participating in a pervasive environment will sense, collect, store, and share, large amounts of personal data. So, due to the fact that information collection, processing, and sharing in an imperceptible manner is a fundamental requirement for the appropriate operation of pervasive systems, the protection of privacy and pervasiveness at the same time are, by nature, somewhat conflicting.

In this context, it is necessary to develop mechanisms and techniques that would be able to deal effectively with the emerging privacy problems. Although many approaches towards the protection of privacy have been proposed whatsoever, none of them fully meets the requirements for privacy protection in pervasive systems, in a holistic and effective manner. This fact makes apparent the need for developing appropriate technical and other means, capable of ensuring the adequate protection of privacy, while preserving the revenues introduced by pervasive computing environments.

The rest of the paper is organized as follows: In section 2 we present a number of existing approaches towards the protection of privacy in pervasive computing. In section 3 we discuss the main privacy issues that appear in the pervasive computing, as well as a set of basic privacy principles (requirements) that should be met. In section 4 we present our in-progress approach for the protection of privacy by introducing a new Privacy Enhancing Technology (PET), and in section 5 we conclude by presenting our plans for further research.

## 2 Related Work

So far, several PET have been proposed, implemented, and extensively used, mainly for the Internet/Networked paradigm. However, since new threats related to privacy appear in pervasive environments, and given the fact that Internet and pervasive environment are two different paradigms (at least in their size and social impact), existing PET approaches should be now revisited.

Although mechanisms such as P3P [5] and Appel [6] are suitable for defining privacy policies, their applicability in the case of pervasive computing is questionable. The main reason for this is that these technologies cannot express policies in terms of data, destination, purpose, and contextual information, and they cannot enforce or ensure privacy through technology, since they rely on social and regular pressures only [7]. Furthermore, technologies developed mainly for the protection of anonymity (e.g. DC-Nets [8], Mix-Nets [8], Mist [9], etc.) cannot provide a holistic approach for the protection of privacy, due to their focus mainly on protecting users' identities and not to fulfill all privacy requirements. Furthermore, these mechanisms pose demanding requirements in terms of processing and power capabilities and they require special interfaces for their efficient configuration and application.

On the other hand, there exist approaches for the protection of privacy for pervasive computing environments [10,11,12,13,14]. However, these approaches suffer by not being adequately efficient and effective, because they appear to be ad-hoc and specific to the systems studied. For example, pawS system - like P3P - aims to provide users with tools that let them protect their personal privacy, while this is only based on social and legal norms, rather than rigorous technical protection of private information. On the other hand, PISA as well as other approaches is mainly focused on technological solutions to the threats to privacy, and less on non-technological solutions.

Nonetheless, the aforementioned arguments do not mean that we should start designing and implementing technologies from scratch, although for a few cases this might be proved to be the only solution. On the contrary, these approaches may be a valuable starting point and a solid basis for the new privacy protecting technologies.

3 Background

3.1 Privacy in Pervasive Computing

The increased importance of and threat to - privacy in a pervasive computing context can be easily demonstrated. For example, the omnipresence and invisibility of the participated devices can lead to persons having no idea of sensors and actuators processing their personal data. The surveillance and data collection capabilities of these devices in an unobtrusively and unconsciously manner pose a serious threat to privacy. Furthermore, the use of sophisticated data mining technologies will make the processing of personal data easier or even straightforward. Table 1 highlights the most profound characteristics of pervasive computing that pose serious threats on privacy.

Table 1. Privacy threats related to pervasive computing characteristics

1	Pervasive Computing components will be practically everywhere and affect nearly every aspect of our life.
2	Pervasive Computing components (e.g. sensors) will be invisible and potentially act transparently for many users.
3	The enhancement of storage capabilities will make easier the access and process of personal data.
4	The enhancement of sensory equipment, combined with the advances in their storage capabilities, will make feasible to perceive memory prosthesis or amplifiers, which can continuously and unobtrusively record every action, utterance, and movement of individuals and their and our surroundings.
5	The minimization of sensors, as well as the advances in data mining techniques, will increase the amount and types of personal data that are invisibly captured and analyzed.
6	The communication of the objects in Pervasive Computing will usually take place by their own initiation, in a way that might disclose personal data to other objects/users, so as to accomplish their intended purpose.

Nowadays privacy rights enjoy a constitutional status in several countries, especially - but not limited to - in the European Union and the North America. In addition, specific laws, regulations, and guidelines have been adopted, almost worldwide, in order to protect and ensure privacy.

In this context, a set of generic privacy principles, based on fair information practices, was introduced, enjoying wide acceptance in the pervasive computing research community. These principles, to be taken into consideration by technologists and systems designers that implement pervasive computing applications, are [10]:

1. Notice: Users should always be aware of the collection of their personal data.
2. Choice and consent: Users should have the choice of carrying out, or not, of their personal data.
3. Proximity and locality: The collection of data from a user's device should only occur when the user is present (proximity). Processing and access to these data should only be done within the space they were collected (locality).
4. Anonymity and pseudonymity: Whenever the user's identity is not required, or whenever the user does not consent, anonymity or pseudonymity services should be provided for.
5. Security: There should be security mechanisms, which provide adequate protection for collected data.
6. Access and resource: Access to the user's data should only be allowed to authorized persons. There should be regulatory means for the protection of a user against parties that are not complying with this regulatory framework.

In addition to the aforementioned principles, another important issue is the inherent ambiguity of people's perception of privacy. The definition of "private" is usually dealt with in the field of legal studies, so technologists often find it difficult to define a model that considers not only technical, but also social and economic implications of privacy. Therefore, a holistic perception of privacy, which synthesizes social, regulatory, and technological privacy aspects, is important in order to protect and ensure end-users' privacy in pervasive computing environments. Such a model is presented in the following section.

### 3.2 Privacy Model for Pervasive Computing

The successful and socially acceptable deployment of pervasive environments depends, *inter alia*, on how technologists and system designers understand privacy issues of intrusive technologies. Another serious concern lies with the nature of the network environments and the increased connectivity of such systems. Furthermore, it is also important to consider how users understand the privacy issues and how one can support them in determining when, how, and to what extent their personal data is communicated and sharing.

For the purpose of this paper we adopt Lederer's conceptual model for "everyday privacy" in pervasive computing [15]. The concept of everyday privacy refers to an individual end-users' ongoing exposure to, and influence over, the collection of their personal information. The model is based on the societal-scale model introduced by Lessig [16], as well as on the user perceptual model proposed by Adams [17]. The following formula provides for a qualitative abstract of this model and represents the

preferred privacy level of a user, in a given situation: *preferred\_privacy\_level* = *user* (*L, M, N, A, C, PI, IS, IR, IU*), where: L stands for Law, M for Market, N for Norms, and A for the underlying Architecture and Technology. These are the inter-dependending actors, which Lessig made use of in order to describe the profile of privacy in a given context. PI stands for the disclosed personal information, C stands for a set of contextual variables, IS stands for information sensitivity, IR stands for information receiver, and IU stands for the type of information use.

## 4 A New PET for Pervasive Computing

In this section we propose a new PET, called Privacy Enhancing Model for Pervasive Computing (PEM-PC), with an eye towards the specific privacy requirements in pervasive computing. Our approach has three main goals. The first goal is to provide the users with the necessary technological and organizational tools, in order to carry out an appropriate analysis and evaluation of the pervasive environment, which they are

**Table 2.** Basic issues examined during the suggestion of the proposed PET

Basic Issues incorporated by the proposed PET		Privacy Principles Fulfilled
User Preferences	Selection of an appropriate privacy level. Selection of a specific profile	Notice Choice and Consent
Pervasive Context	Consideration of the pervasive context. Collection of privacy-related information from the environment. Evaluation of the privacy threat posed by the pervasive environment. Management of the threats.	Notice
Privacy Boundaries	Identification of specific privacy constraints. Modification of privacy constraints through dynamic negotiation.	Choice and Consent Anonymity and Pseudonymity Proximity and Locality
Trust Implications	Identification of privacy parameters regarding the selected/required trust level. Clear description of trust issues to the user. Identification of the level of uncertainty regarding the selected/requested trust level.	Security Access and Resource
Data Sharing	Identification of the parameters regarding proactive data sharing. Identification and evaluation of data sharing regarding trust level.	Choice and Consent Security Access and Resource Anonymity and Pseudonymity

involved with. The second goal is to meet the essential generic privacy principles, using methods for context-awareness and decision-making. The last goal is to propose a specific PET, capable of controlling the increase of information asymmetry<sup>2</sup> by increasing the flow of information from the pervasive environment to the users.

Our approach incorporates as many possible aspects related to privacy in pervasive computing as possible, in accordance with the Lederer's conceptual model for "everyday privacy". Table 2 highlights the basic issues that our PET incorporates in conjunction with the privacy principles fulfilled, while a most detailed analysis is presented in the next sections.

#### 4.1 Requirements Met by the Proposed PET

In order for the proposed PET to be effective, an appropriate set of requirements should be met with. We divide this set of requirements in two categories, namely, the requirements related to users and those related to its applicability.

Initially, and from the users' perspective, a major requirement is transparency in its operation. PEM-PC will be developed so as to incorporate specific adaptability (assessing privacy constraints) and decision-making capabilities (the kind of data that can be shared or not), in a dynamic way. Another major requirement is the way in which the interface of our PET will be designed. What is specifically needed is the design of an interface, which will seriously facilitate the decision upon the values of the various parameters. These parameters are related to the value given by each user to her personal data, the definition of the desired privacy preferences, as well as the impact of possible data loss. Therefore, the implementation of the PET's interface will be based on the appropriate Human Computer Interface (HCI) principles, so as the user be able to realize, up to a fair degree, the way this PET functions, and thus be able to modify a decision tree that has been generated by the PET itself.

Another important requirement is to ensure how its interaction with the pervasive environment is taking place. The aim is that the information received from the environment be rich, distinct, and accurate, so as the PET be able to produce an adequate and reasonable level of privacy protection through the identification of possible ambiguities and through the dynamic, sensible and secure negotiation process, to take place between the PET and the privacy computing representatives<sup>3</sup>.

#### 4.2 PEMPC Components

PEM-PC should be modular, so as its underlying components are distinguishable. The generic components of PEMPC are:

- *Generic Privacy Risk Analysis and Management:* The specific component serves as a privacy-related information assessment mechanism. Its basic function is to detect possible privacy threats and to propose an appropriate trust level.
- *Proactive Data Sharing Model:* PEM-PC defines the users' personal data and the circumstances under which these data are shared with other parties [18].

<sup>2</sup> In asymmetric information situations, information is not available in the same extent or under the same conditions to all entities.

<sup>3</sup> Pervasive computing representatives can be either a physical entity or a proxy (usually an agent) acting on behalf of the pervasive system.

- *Negotiation Entities*: These are responsible for all privacy-related actions, especially for the filtering of the contextual information collected, and the appropriate setting of the privacy constraints.
- *Message Encryption*: This component is responsible for all cryptographic-related actions and aims at ensuring the security of the messages exchanged.

Figure 1 depicts the components of the proposed PET and their interactions.

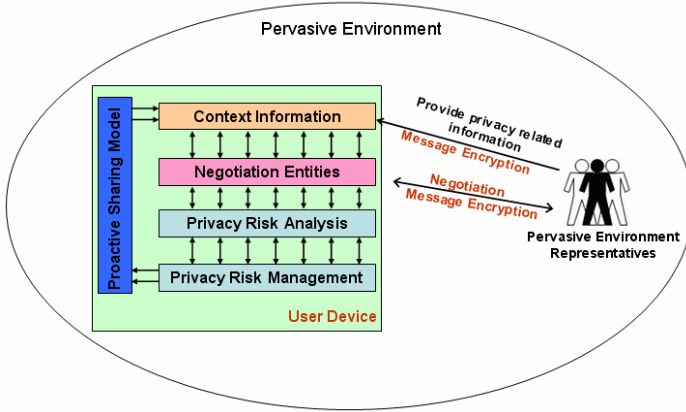


Fig. 1. PET components and their interactions

#### 4.2.1 PEMPC Prerequisites

The generic privacy risk analysis and management component is based on the development of a heuristic model, which serves for extracting privacy related information from the environment the user joins with. The model is based on a goal analysis method that examines what type of action the user wishes to take, what elements of the environment participate in the processing, and from what type of nodes this piece of information is parsed [19]. This method is considered supplemental, since it can detect additional types of threats.

Having completed this generic analysis, the evaluation of the contextual information takes place, followed by the threat management module. The evaluation is based on the application of a mathematical model on the collected information. The result of this process is the identification and management of the potential threats. The mathematical model that could be used varies and depends on the information, which is most preferable to protect, while its accuracy is a vital prerequisite of the PEM-PC in order to make more accurate decisions regarding the sharing of the data.

On the other hand, the management process requires and involves the definition of a trust level, a level of uncertainty, and an algorithm, which supports the decisions that have to be made, regarding the sharing of personal data. In this paper, the analysis of PEM-PC components is primarily focused on the conceptual level, although not neglecting specific implementation issues (i.e. the examination of specific protocols and/or programming languages).

#### 4.2.2 Privacy Analysis

The generic method is based on a series of “interactions” (“queries”) [20], posed by a user’s personal device to the pervasive environments representatives. We assume that this set of “interactions” is not predefined, but can be revised through a dynamic negotiation of PEM-PC with the specific pervasive application<sup>4</sup>. These interactions can be organized in three basic categories: Social-oriented queries, management-oriented queries, and technological queries [20], in order to incorporate the basic aspects of privacy defined by Lederer’s model. Some queries may aim at identifying who are the pervasive environment’s users, which are the data consumers, which is the current level of privacy, what type of data is usually needed to be shared and for what purpose, if there are any recognizable malicious third parties that might collect data, how the collection of data is performed, where the data are they stored, etc.

In order that a user’s device obtains an informed view of the pervasive environment, the negotiation between PEM-PC components and the system’s representatives should be defined in a sensible and comprehensible way, for both sides. In order to achieve this, we define a pervasive-oriented ontology. This ontology aims at capturing the basic issues regarding the pervasive computing domain and privacy issues, in order to facilitate the knowledge sharing in such a way that we could combine privacy related information and draw specific conclusions.

In the case of the generic privacy risk management process, we need to develop and implement a specific formal model. In specific, we need to value the collected data. Hence, we initially set a specific level of importance to each query category, and then we define a specific value to the response to each query, respectively. The overall process depends on the privacy threat entailed in each query category, so we mainly focus on queries of social and technological nature. This procedure needs to be supported by a process of quantifying the value of users’ personal data. This value could be expressed through three parameters [20], namely: a) the importance of personal data, which depends on users’ preferences, b) the risk of disclosure of these data to a third party, and c) the impact, in terms of privacy, in the case of wrong decision regarding the sharing of the data. For each of the above parameters, PEM-PC’s subsystem defines a specific value, automatically or according to what user desires, thus preserving the subjective nature of the process.

Furthermore, what is also needed is to define a specific trust level concerning the user’s environment [21]. Having in mind the above queries categories, we set in a similar way three values regarding each category, namely a) *social\_value*, b) *management\_value*, and c) *technology\_value*. The defined trust level is proportional to the aforementioned level.

Additionally, setup of a specific level of uncertainty is also needed. This level will act supportively to the level of trust defined earlier. The level of uncertainty will depend on the clarity of the answers derived from questions concerning privacy threat analysis, and will help users obtain a more distinct view of the pervasive environment, where they participate.

---

<sup>4</sup> The negotiation will be based on a previously agreed-upon set of privacy-related attributes.



### 4.2.3 Proactive Data

Proactive data sharing systems help users define personal data, the disclosure of which does not constitute a violation of the person's privacy. As a result, users can share their low-value data voluntarily. Therefore the relationship between data producers and data consumers may be theoretically transformed from antagonistic to cooperative. Such an approach can help reducing information asymmetry situations by controlling the communication between data producers and data consumers, in specific circumstances. The data shared voluntarily are characterized as "plain" type of data, and can be easily collected by data consumers [18].

In the case of PEM-PC, and in order to set the low-value data, we take into consideration the three above parameters. More specifically, data with low value in terms of the introduced risk in case of their disclosure can be considered as low value data.

### 4.2.4 Negotiation Entities

Negotiation entities are responsible for distributing users' privacy requirements and for retrieving privacy-related information from the environment the users enter. The privacy related information about the environment is obtained through the queries we mentioned previously. Therefore, the entities should have the ability to handle the information/reactions provided by the environment and track down ambiguous and vague situations. In addition, entities should possess basic negotiation skills, so that privacy requirements of the users are clearly stated. Furthermore, the entities should have the ability to filter and classify the data collected from the environment, in conformance with the specific privacy constraints defined by each user.

Privacy constraints are formulated through the use of a mark-up language, such as XML. With the exception of the automated generation of privacy constraints, PEM-PC should provide the user with the ability to define or change an existing constraint through the appropriate interface. An example of a privacy constraint, in XML, is:

```
<xml>
<privacy_constraint name="data_usage">
<time_of_storage> 500 </time_of_storage>
<access> limited </access>
<device> hard_disk </device>
</privacy_constraint>
</xml>
```

### 4.2.5 Message Encryption

While the effectiveness of current cryptosystems is probably adequate in many cases, their application in the context of pervasive computing is still questionable. The main reason for that is the processing, computational and storage limitations of the devices operated in such environments, which limit their capabilities of using strong cryptosystems. Therefore, alternative methods for data confidentiality should be explored.

For the purpose of PEM-PC, we explore the use of Lightweight Cryptography [23], which pose meaningful requirements and can be deployed without the support of a substantial - power demanding - technological infrastructure.

## 5 Conclusions and Further Research

In this work we described a generic PET, called Privacy Enhancing Model for Pervasive Computing, focused on future pervasive computing environments in order to fulfill privacy requirements in an holistic way by incorporating social as well as technical issues. Additionally, we set the basic framework, so as PEM-PC be effective and efficient, while – in the same time - respecting the specific privacy needs that may be raised by users in such an environment. The PEM-PC is to be modular, in order to be able to be enhanced easily by adding new and more robust privacy approaches and techniques.

We plan to enhance the negotiation entities component and to define a standard framework for negotiation process, so as to produce more concrete and accurate knowledge (conclusions), regarding the privacy preferences of the users and the privacy constraints defined by the pervasive environment. We also plan to identify and develop efficient solutions regarding the identity management of the users. Furthermore, we plan to enrich our ontology further, in order to support information related to location of the users, so as to maintain location privacy.

## References

- [1] Weiser M., “The computer for the 21<sup>st</sup> Century”, *Scientific American*, Vol. 265, no. 3, pp. 94-104, September 1991.
- [2] National Institute of Standards and Technologies, *Proc. of the IT Conference on Pervasive Computing*, 2001 ([www.nist.gov/pc2001/about\\_pervasive.html](http://www.nist.gov/pc2001/about_pervasive.html), accessed April 2, 2005).
- [3] Lahlou S., Langheinrich M., Rucker C., “Privacy and trust issues with invisible computers”, in *Com. of the ACM*, Vol. 48, no. 3, pp. 59-60, March 2005.
- [4] Russell D., Streitz N., Winograd T., “Building disappearing computers”, in *Com. of the ACM*, Vol. 48, no. 3, pp. 42-48, March 2005.
- [5] The Platform for Privacy Preferences 1.0 (P3P1.0) Specification World Wide Web Consortium, September 2001, [www.w3.org/TR/2001/WD-P3P-20010928](http://www.w3.org/TR/2001/WD-P3P-20010928).
- [6] A P3P Preference Exchange Language 1.0 (Appel 1.0), Working draft, World Wide Web Consortium, April 2002, [www.w3.org/TR/P3P-preferences](http://www.w3.org/TR/P3P-preferences).
- [7] Myles G., Friday A., Davies N., “Preserving Privacy in Environments with Location-Based Applications”, in *IEEE Pervasive Computing*, Vol. 2, No. 1, pp. 56-64, 2003.
- [8] Fisher-Hübner S., *IT Security and Privacy*, Sweden 2001.
- [9] Muhtadi A., Campbell R., Kapadia A., Mickunas M., Yi S., “Routing through the Mist: Privacy Preserving Communication in Ubiquitous Computing Environments”, in *Proc. of the International Conference on Distributed Computing Systems*, Austria, 2002.
- [10] Langheinrich M., “A Privacy Awareness System for Ubiquitous Computing Environments”, in *Proc. of the 4<sup>th</sup> International Conference on Ubiquitous Computing*, Sweden 2002.
- [11] Nguyen D., Mynatt E., “Privacy Mirrors: Understanding and Shaping Sociotechnical Ubiquitous Computing Systems”, Georgia Institute of Technology, *Technical Report GIT-GVU-02-16*, 2002.
- [12] Lederer S., Dey A., Mankoff J., “Everyday Privacy in Ubiquitous Computing Environments”, in *Proc. of the 4th International Conference on Ubiquitous Computing*, 2002.

- [13] X. Jiang, J. Landay, "Modelling Privacy Control in Context-aware Systems", *IEEE Pervasive Computing*, pp. 59-63, 2002.
- [14] Blarkom G., Borking J., Olk J., "Handbook of Privacy and Privacy-Enhancing Technologies. The case of Intelligent Software Agents", *PISA Project Deliverable*, The Hague, 2003.
- [15] Lederer S., Mankoff J., Dey A., "A conceptual model and metaphor of everyday privacy in Ubiquitous Computing", Intel Research, USA, July 2002.
- [16] Lessig L., "The Architecture of Privacy", In *Proc. of the Taiwan NET '98 Conference*, Taipei, March 1998.
- [17] Adams A., "The Implications of Users' Privacy Perception on Communication and Information Privacy Policies", in *Proc. of Telecommunications Policy Research Conference*, USA, 1999.
- [18] Balfanz D., Golle P., Staddon J. "Proactive Data Sharing to Enhance Privacy in Ubi-comp Environments", in *Proc. of UbiComp 2004 Privacy Workshop*, England 2004.
- [19] Jensen C., Tullio J., Potts C., Mynatt E., "STRAP: A Structured Analysis Framework for Privacy", *Graphics, Visualization and Usability Center (GVU) Technical Report*, Georgia Institute of Technology, January 2005.
- [20] Hong J., Ng J., Lederer S., Landay J., "Privacy Risk Models for Designing Privacy-Sensitive Ubiquitous Computing Systems" in *Proc. of Designing Interactive Systems*, USA, 2004.
- [21] Langheinrich M., "When trust does not compute-The role of trust in Ubiquitous Computing", in *Proc. of the UBICOM on Privacy (UbiComp 2003)*, USA, October 2003.
- [22] Dornan A., "Trusted computing: A matter of trust", in *Network Magazine*, Vol. 19, No. 7, pp. 26-32, July 2004.
- [23] Seys S., "Lightweight Cryptography Enabling Secure Wireless Networks", *Workshop on Security Issues in Mobile and Wireless Heterogeneous Networks*, Belgium, 2004.

# Bringing the User Back into Control: A New Paradigm for Usability in Highly Dynamic Systems

Sebastian Höhn

Albert-Ludwigs University Freiburg, Friedrichstr. 50, 79100 Freiburg, Germany  
`sebastian.hoehn@iig.uni-freiburg.de`

**Abstract.** The contribution of this paper is twofold. On the one hand, we report on the results of our investigation of different categories of usability issues. On the other hand, we introduce the ideas of context descriptive security models as a means of mastering the usability challenges of highly dynamic systems. Modern computer systems are involved in many situations of our daily lives. This means that newly deployed systems must be carefully designed in order to be correctly used by laypersons. The scenario we introduce shows that it is no longer feasible to argue that users must be educated in order to correctly operate these systems. As soon as such a system is deployed, for example, in a supermarket, the education-barrier will not be accepted: neither by the customer nor by the provider.

## 1 Introduction

Personal computers and Internet technology have taken hold of many areas of our daily lives. While personal computers and mainframes are disappearing from sight, calm and embedded systems have already begun their triumphal procession. Current mobile phones, for example, are equipped with wireless network technologies and come with more computing power than you need to phone your spouse. Since the omnipresence of unused computing power and the increasing potential to interconnect different devices enable a vast number of new services, these newly arising systems are highly dynamic.

## 2 The Brave New Supermarket

To exemplify the different aspects of a highly dynamic smart environment, we present the scenario of a supermarket: when customers enter the store a welcome message showing the latest bargains is sent to their mobile devices. The personal shopping lists appearing on the devices' screens are automatically transferred when they leave their homes or offices to go shopping. After scrolling through the list, they may take a look at the special offers that are individually prepared for regular customers. Some new type of convenience food arrests their attention, but they are in doubt about which wine goes with it: one click on the mobile

device and some recommendations show up on the display, together with the current prices and the location of the shelves where the bottles stand. When they have picked up all the goods and proceeded to the checkout, it is not necessary to unload the cart. The backstore system has registered all the goods and already calculated the price. It is now a simple matter of touching the “pay point” with the near field communication enabled cell phone and enjoying the convenience of mobile payment.

Although this scenario sounds a lot like science fiction, all the technologies necessary to build a prototype already exist. Many of the technologies are currently tested in Metro’s Future-Store initiative [1]. So why do we still hesitate to make this brave new world reality in our everyday lives?

One of the most important issues with the deployment of *highly dynamic systems* is the adequate management of security and privacy requirements. Unlike current computer systems, a smart environment is an open system and, to this end, can no longer be completely specified. We can neither anticipate all the possible devices that wish to communicate, nor can we anticipate every user’s intentions and goals. The system must cope with change during every minute of its operation. The requirements placed by the user’s protection goals can no longer be analyzed in advance. They must be handled by the user for each individual transaction. The problem is even worse if we bear in mind that the environment must, at the same time, act on behalf of its users and its providers. So it must consider two different – perhaps even contradicting – security and privacy policies. Within this area of conflict, the user must be brought into a position where he is able to express his goals, define his security and privacy policies and observe the operation of the system while it processes his inquiries. Hence, observability and controllability are important properties for gaining the user’s trust [2]. This is extremely important, otherwise the user must blindly trust the candor of the systems and their providers. The lack of observability and controllability is experienced as a lack of influence by the end users. This fact is reflected by the activities of organizations such as CASPIAN [3], with their claim that the customer must be in control of his data. But how can we achieve observable and controllable behavior for inexpert users?

This paper is set out as follows: first we provide an overview of the different usability and security issues we identified in our studies. From this experience, we derive the requirements for a solution. Finally, we present the context descriptive security models as a possible solution for the usability issues in highly dynamic systems. We conclude the paper with a summary and an outlook on future work.

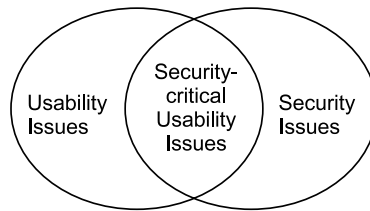
### 3 Lessons Learnt from Usability Studies

The actual problem is that the vast majority of users cannot cope with the complexity of security relevant applications. Roughly speaking, there are two reasons for this. On the one hand, usability of security relevant applications is often neglected and the user interfaces have severe usability issues. On the other hand, we have not found correct metaphors for the mechanisms themselves in

order to enable *end users* to easily use security relevant applications. In the following chapter, we will investigate the different aspects of security and usability and how they are related with the user's failure to adequately use security mechanisms.

### 3.1 Categories of Usability Issues

For this research we have to address two different aspects of computer systems: usability and security. It is often claimed that these two aspects require contradictory conditions, so that system designers are unable to deal with usability and security at the same time. We argue that a reasonable level of system security cannot be achieved without taking into account the users and their interactions with the system.



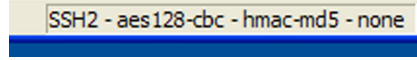
**Fig. 1.** Not all security issues are security-critical ones

Applications may suffer from a wide variety of usability problems that are certainly annoying and frustrating for the end user. Not all usability problems inevitably cause vulnerabilities within the system. However, not all vulnerabilities are necessarily caused by usability problems of the application. Multifaceted reasons may lead to vulnerabilities within computer systems. Buffer overflows or inconsistencies between protocols are some of the numerous causes for successful attacks that cannot be traced back to usability issues. So we notice that the areas of usability and security have a significantly interesting (and hence non-empty) intersection (cf. Figure 1). Security critical malpractice only arises from this intersection and thus endangers the compliance of the system's behavior with the user's protection goals. Security critical mistakes may compromise the confidentiality, integrity and availability of the whole system.

We provide a classification of usability problems that proved to be rather useful in the investigation of security-related usability issues [4]:

1. Usability problems that do not endanger the security of the system (i.e. roughly speaking lack of ergonomics)
2. Usability problems that in fact put the system's security at risk and arise despite the user's adequate security competence
3. Usability problems that arise directly from the user's insufficient security knowledge
4. Security problems that do not arise from user interaction

Several user studies we conducted at our institute revealed that for security critical applications approximately 75% of the security issues fall into the categories 2 and 3 of the classification given above [5]. This supports our assumption that usability problems within security-related applications very often lead to security problems. Hence security research must be geared towards its audience and the human factor must be adequately considered.



**Fig. 2.** How secure is the current connection?

Most computer users are more or less unskilled and are not aware of the manifold security mechanisms. System designers must be very careful when they implement methods for observing the current state of the system. It is absolutely necessary to find appropriate metaphors for given system states. If we consider Figure 2 and bear in mind that it is part of an interface that was designed to give the user information on the security level of the current connection, it is obvious that consolidated knowledge of different security mechanisms and their performance is necessary. The issues with observability are therefore twofold: on the one hand, applications provide information that is more or less useless to the end users due to their lack of technical knowledge. On the other hand the missing context information does not allow for a thorough analysis of the current transaction's security level. Our studies have shown that the majority of the issues with security critical applications are problems caused by a lack of observability [5].

### 3.2 Requirements of the Solution

Whitten and Tygar introduced the “barn door property” [6], which is based on the proverb stating that it is useless to close the door after the horse has bolted: secret information, once lost, may be the starting point of an attack that can scarcely be prevented. Consider the disclosure of a private key within a public-key infrastructure. It takes an attacker just a few seconds to copy the key and, owing to the user's unawareness, impersonating this user on the Internet.

To this end, we see the avoidance of errors as one of the most important requirements. With regard to security, fault-tolerance is rather difficult to achieve, because it is unclear how we can recover, once the barn door is open and cannot be closed again. Obviously, a disclosed private key must be revoked and replaced by a new one. As this is costly, avoidance is the only viable solution in most cases.

Observability of transactions and decisions is an inevitable prerequisite to achieve avoidance of errors. The user must be brought to a position where he can easily observe the behavior of the system at any time. Only if he is able to evaluate the system's current state can he adequately and effectively control the system's behavior. We would go even further and claim that he must be able

to observe the effects of a decision even before he finally makes the decision. A system that allows for the evaluation of the outcome of a transaction with different security settings and gives reliable information on the effects of the outcome is well suited to a security layperson's needs. The users are able to play around with the system configuration; can observe the outcome of different configurations and then – depending on the results – decide for the best one.

If the users do not have enough knowledge of the underlying security mechanisms they can still rate a specific system state by observing the *outcome's* compliance with their security requirements. If they are satisfied with the effect of an interaction, they may accept the configuration. If they are not, they can reconfigure the system until it satisfies all requirements. It is important to realize that all the aspects of usable security that are claimed important in the engineering process, for example in [5,7,8], are of great importance for this solution to work. If the users cannot efficiently control the system, they will soon be frustrated and are, in effect, unable to find an optimal configuration. They give up searching for an appropriate one and accept a mediocre security level just because they cannot find the an adequate one.

## 4 Context Descriptive Security Models

The exploration of the users' capabilities and behavior leads to the conclusion that the integration of different techniques and mechanisms into a context descriptive security model will provide us with a solution to the problem. The different scenarios of smart environments have one major property in common: they focus on the interaction of *end users* with the system. These are virtually unskilled, but must understand security and privacy models provided by the smart environment. Therefore, appropriate metaphors and interfaces for defining these models must be found.

Resuming the scenario of the supermarket, what security implications will the users have in this context? We consider two examples to illustrate our approach. First of all the users are interested in the information about themselves that is collected when they visit the store. Secondly they are interested in correctly identifying and authenticating devices. This is especially important if we consider mobile payment as an option at the checkout.

Context descriptive security models, as we define them, take into account the environment in which the users move and interact. The users are able to observe the different transactions and will no longer be confronted with the abstract configuration of security mechanisms. For example, they will realize the importance of authentication mechanisms, if they observe possible eavesdroppers within the current context.

### 4.1 Towards an Integrated Security Architecture

The context descriptive security model of the supermarket encompasses communication devices, such as the wireless LAN, Bluetooth or NFC hotspots. Furthermore, it contains the different devices and sensors that track the user's movement



and actions. This means all the cameras, RFID readers, motion detectors and other devices must be contained in the model. Location based sensors, such as cameras or motion detectors must be represented within the areas they cover. Additionally the data detected by the different sensors must be represented by the model. It does not suffice to tell the users that cameras will take pictures of the faces. It is, additionally, important to tell them that these photographs will be used to identify them, if they return to the store. All this information allows the users to gain deep insight into the supermarket's information collection techniques.

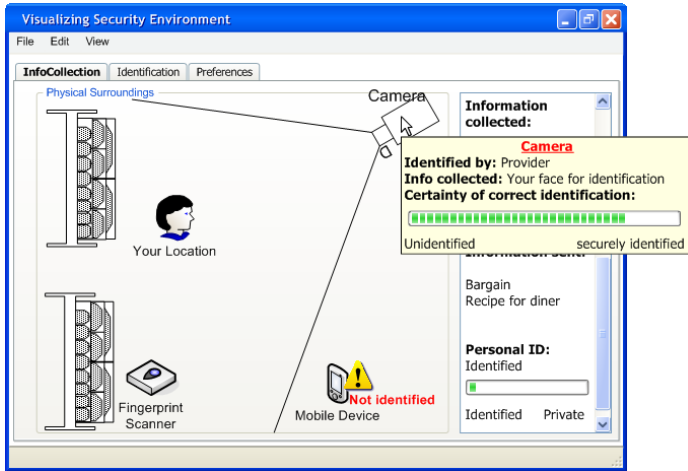
It will become impossible to prevent smart environments from collecting information on the customers. Hence, controlling the processing of data with security and privacy policies will play an important role.

Information can be collected secretly and even the amount of information that is legally collected is that countless that it cannot be controlled individually by the provider. Thus, the user must be enabled to control the processing of his information himself. The user can formulate his security and privacy policy and he must be convinced by appropriate techniques that the system complies with his policies. If the user, for example, claims that his location may be forwarded to authenticated staff of the supermarket only he must be able to observe the compliance of the system to these obligations. The formal foundation of these obligations is thoroughly investigated [9], but it is still unclear how they can be applied in a practical application.

Obviously, situations exist where it is undesired to reveal the complete information collection strategy. For example, surveillance cameras are intended to secretly collect information in order to prevent theft. If these devices are not part of the model that is presented to the user, they may not be linked with the information officially collected.

## 4.2 Presenting the Model to the User

Now that we have introduced the underlying security model, we will discuss the presentation of the model. From the usability point of view, this is the essential part of the prototype. We investigated the presentation of security and its perception by end users with mock-up interfaces of simple scenarios (cf. Figure 3). Individual devices are represented by symbols in a map of the covered area. Mobile devices are represented by their actual location, while omnipresent services – such as a camera that covers an area – are represented as areas or by icons that are contained in a dedicated window. This map shows not only devices and services, it furthermore integrates the surroundings of the physical world in which the users move. If, for example, the map of the floor shows their location and the locations of different services, they will more easily comprehend the interactions between the single devices as they pass along the way through the building. This visualization will show the users quite plainly that all transactions involve a variety of remote services and devices. This fact is often neglected by inexperienced users, who do not realize the distributed nature that even, to their notion, simple transactions (e.g. querying “Google”) may possess.



**Fig. 3.** Mock-up interface for visualization of the security model

Social navigation as it was proposed by [10] can be greatly extended by this security model. Social navigation uses the history of others to visually guide users to more frequently used solutions. In the real world this is like a worn path across a field of grass that shows us where others have walked and directs us towards points of interest. The underlying assumption of DiGioia and Dourish is that the solutions of others guide the users towards a solution for their own issues. From a security point of view, the information gathered from the history of others is useless at first sight. However, if we are able to indicate the individual risks by rating the results within the individual histories, this will become a precious source of information for all users.

Furthermore, this presentation allows for a trial and error approach of clueless people: it is possible to simulate certain actions by the application of formal methods and observe their outcome before they are actually performed. This enables the evaluation of the impact that the planned transaction has for privacy and security claims. Consequently, the users can reconsider their configurations and undo certain settings, reevaluate the transactions, and thus gradually increase the security and privacy level of the devices.

Authentication can be based on the location of specific devices. The customer talking to the salesclerk may identify the clerk's device by its location. So it is possible to identify different devices and allow for different types of interactions. The sending of product information from the clerk to the customer or the transmission of information from the customer to the salesclerk in order to receive special support (for example, to issue a guarantee) requires different security levels.

Different security mechanisms are not directly presented to the user. The application of the mechanisms is invisible and indirectly visualized by security levels. These levels (for example, "Certainty of Identification", "Confidentiality"

and “Anonymity”) are calculated according to the underlying attacker models and the current situation of the user. This type of visualization was proposed by Dourish and Redmiles [11]. We investigated the derivation of security levels from different measures based on the works by Sheyner et al [12], Dacier et al [13] and Ammann et al [14].

## 5 Conclusion and Future Work

In this article, we have given a detailed description of our current approach regarding the human computer interaction in smart environments. Our investigation of current usability issues and the different categories of users prompted us to propose the context descriptive security model as a solution to the different issues.

Our currently ongoing research is structured according to the challenges presented in the chapters above. We concurrently investigate the underlying security policies. Policies for smart environments must rely on a plethora of mechanisms to ensure their enforcement. We also investigate the formal foundations of secure log and audit in highly dynamic systems [15,16].

The most important goal of our current research is to enable the users to control the behavior of highly dynamic systems. To this end, they must be able to formulate their security needs, but it is equally important to have reliable and non-intrusive feedback mechanisms if the user’s current configuration cannot be guaranteed. This type of feedback is part of our ongoing research in the area of human computer interaction.

Even if users are unable to observe and control each individual transaction within a highly dynamic scenario, the proposed security model enables the user to observe the system’s behavior. This is a basis for the user to trust the provider: it is not necessary to really observe each individual transaction, the possibility to do so increases the risk for a disrespectful provider of being discovered.

## References

1. METRO AG: Metro’s future store initiative. <http://www.future-store.org> (2006)
2. Alan J. Dix, Janet E. Finlay, Gregory D. Abowd, Russel Beale: Human Computer Interaction. 2nd edn. Prentice Hall (1998)
3. CASPIAN: Consumers Against Supermarket Privacy Invasion and Numbering. <http://www.nocards.org> (2006)
4. Kaiser, J., Reichenbach, M.: Evaluating security tools towards usable security. In: Proceedings of the 17th IFIP World Computer Congress (WCC2002). (2002)
5. Daniela Gerd Tom Markotten: Benutzbare Sicherheit in informationstechnischen Systemen. PhD thesis, Albert-Ludwigs-University Freiburg (2003)
6. Whitten, A., Tygar, J.D.: Why Johnny can’t encrypt: A usability evaluation of PGP 5.0. In: 8th USENIX Security Symposium. (1999)
7. Zurko, M.E., Simon, R.T.: User-centered security. In: NSPW ’96: Proceedings of the 1996 workshop on New security paradigms, New York, NY, USA, ACM Press (1996) 27–33

8. Flechais, I., Sasse, M.A.: Developing secure and usable software. In: Workshop on Human-Computer Interaction and Security Systems, ACM (2003)
9. Hilty, M., Basin, D., Pretschner, A.: On obligations. In: 10th European Symposium On Research In Computer Security. LNCS, Springer (2005)
10. DiGioia, P., Dourish, P.: Social navigation as a model for usable security. In: SOUPS '05: Proceedings of the 2005 symposium on Usable privacy and security, New York, NY, USA, ACM Press (2005) 101–108
11. Dourish, P., Redmiles, D.: An approach to usable security based on event monitoring and visualization. In: NSPW '02: Proceedings of the 2002 Workshop on New Security Paradigms, New York, NY, USA, ACM Press (2002) 75–81
12. Sheyner, O., Haines, J., Jha, S., Lippmann, R., Wing, J.M.: Automated generation and analysis of attack graphs. In: SP '02: Proceedings of the 2002 IEEE Symposium on Security and Privacy, Washington, DC, USA, IEEE Computer Society (2002) 273
13. Dacier, M., Deswarte, Y., Kaâniche, M.: Models and tools for quantitative assessment of operational security. In: Information systems security: facing the information society of the 21st century. Chapman & Hall, Ltd., London, UK, UK (1996) 177–186
14. Ammann, P., Wijesekera, D., Kaushik, S.: Scalable, graph-based network vulnerability analysis. In: CCS '02: Proceedings of the 9th ACM conference on Computer and communications security, New York, NY, USA, ACM Press (2002) 217–224
15. Rafael Accorsi: On the Relationship of Privacy and Secure Remote Logging in Dynamic Systems. In: Proceedings of the International Information Security Conference Security and Privacy in Dynamic Environments. (2006)
16. Rafael Accorsi, Adolf Hohl: Delegating Secure Logging in Pervasive Computing Systems. In: Proceedings of the International Conference on Pervasive Computing. (2006)

# Extending SQL to Allow the Active Usage of Purposes

Wynand van Staden and Martin S. Olivier

Information and Computer Security Architecture Research Group  
University of Pretoria  
Pretoria, South Africa  
`wvs@adam.rau.ac.za`  
`http://www.mo.co.za`

**Abstract.** The protection of private information revolves around the protection of data by making use of purposes. These purposes indicate why data is stored, and what the data will be used for (referred to as specification/verification phases).

In this article, the *active specification* of purposes during access requests is considered. In particular it is argued that the subject that wishes to get access to data should explicitly specify their reason for wanting the data; as opposed to verification taking place by implicit examination of the subject's profile. To facilitate this *active specification* extensions to the SQL data manipulation language is considered.

## 1 Introduction

Current Privacy Enhancing Technologies (PETs) protect access to information by binding purposes to objects (data) – the specification phase, and only relinquishing data if a subject's purposes (reasons) for accessing data coincides with the purposes for storing data – the verification phase.

Recent work done on the verification phase [5] proposes the association of purposes with a subject's profile (or with a role in Role Based Access Control (RBAC)). Access is granted if a subject's profile contains the “correct” purposes for the data they are requesting access to. This method is implemented in a Mandatory Access Control (MAC) model, with a central authority deciding which purposes data are stored for, and which purposes are associated with subjects – thus no ownership of data.

In this paper it is argued that subjects should not be granted access to data because they “happened” to have the right purposes in their profile. If this were the case, purposes would be little more than additional “access modes” that are used to control access to data. As the enterprise storing the data is to be held accountable for its use of the data, so too must the employee of the enterprise be held accountable. Thus the mechanism that allows the employee access to the data should also provide him with a way of stating his purpose with the data; allowing proper access control, and facilitating the creation of verbose audit trails.

This paper considers extensions to the most widely used mechanism by which access is gained to data: SQL. If such extensions are to be useful it must be compatible with current SQL standards, and should not compromise the protection of data. The paper focuses primarily on the *application* of data, and thus considers the extensions in two phases.

The first phase considers extensions to the **grant** statement, which enables the implementation of an access control model which is a hybrid between Discretionary Access Control (DAC) and Mandatory Access Control, based on work done by the authors in [16]. This hybrid model relies on the “no ownership” of data of MAC to protect privacy, and the delegation of access control of DAC by subjects to provide flexibility. It thus avoids the central reason that DAC is not considered plausible in a privacy environment [8], ownership of data. In the second phase, an extension to the primary data access clause, *select*, is considered.

*Revoke*, *insert*, *delete*, and *update* fall outside of the scope of this paper, and will be reported on elsewhere. However, a few thoughts on revoke are presented. Revoking normal access modes in the extended model should be no different from the normal model. Revoking purposes becomes more complex, and it must be ensured that revoking a purpose results in a more restrictive privilege. We are currently investigating rewrite rules which are used to rewrite purposes in a binding to accomplish this.

The proposed extensions are flexible enough for implementation as they do not change the utilisation of the relevant SQL statements significantly. The extensions are clauses that can be specified when necessary or omitted – the semantics of the extensions are defined in both cases.

The primary intention is not to change the SQL specification, but to consider a possible syntax for such extensions, their semantics, and impact on protecting access to data. While it is not possible to provide a comprehensive solution to privacy issues, the extensions proposed in this paper at least take a step in the direction of integrating PETs with database technologies.

It is possible to construct a “preprocessor” which manages these extensions; reading purposes from the database and doing “pre-verification”, after which the extensions are removed from the original SQL statement, and the statement is passed to the DBMS for normal access control<sup>1</sup>.

The rest of the paper is organised as follows: Section 2 provides background information. Section 3 proposes the syntax and semantics of the extended SQL. Section 4 considers the case where the clauses are omitted. Finally, section 5 concludes the paper.

## 2 Background

The OECD’s use limitation and purposes specification document [13] is one of the earliest to put forward principles for the protection of an individual’s privacy. A collaborative document by Canada and the Netherlands take this further and

---

<sup>1</sup> The authors are in the process of constructing such a prototype.

proposes components which must be provided by a computerised system in order to protect privacy [10].

Many current systems, and policy models integrate these principles such as the Platform for Privacy Preferences (P3P) [6], the Extensible Access Control Markup Language (XACML) [12], the Enterprise Platform for Privacy Preferences (E-P3P) [2], and the Enterprise Privacy Authorisation Language (EPAL) [3]. A real world implementation to protect the privacy of individuals at the point of failure – the database – has emerged as the Hippocratic database [1].

It is common for purposes to be placed in a hierarchical structure such as a lattice [7], or a tree [5]. Previous work done by the authors [16] places purposes in a lattice, with the specific understanding that every purpose in the lattice is atomic in nature, and that purposes lower in the hierarchy subsume those that are higher. Thus, if a node of this lattice is bound to the data, a subject requesting access must provide exactly that purpose (not just a “subset” of it), or something that subsumes it (*ie* more specific). A purpose  $P_1$  that subsumes another purpose  $P_0$  is written as  $P_1 \geq P_0$ . Additionally our work also introduces *compound purposes*.

A compound purpose is a purpose which consists of other purposes, and is expressed by forming conjunctions or disjunctions with purposes contained in the purpose lattice, using well defined operators. These operators – **and**, **or**, and **andnot** are defined over purposes and *reasons*, with different and-, and or-operators for purposes and reasons, respectively.

Compound purposes allow one to express dependencies between certain purposes, for example one might state that a credit-card number is stored for “billing **and** refund” purposes – stating that a subject can only access the data if he provides both purposes as a reason for accessing the data. The **or** operator is less restrictive, as “billing **or** refund” only requires the subject to specify one of the two purposes.

Our work [16] makes a distinction between a purposes and reasons. A purpose is bound to data and is considered stable. Reasons are used during access requests, are transitory, and are defined in terms of (compound) purposes. They can thus be placed in a lattice structure, but it is unnecessary. The distinction between purposes and reasons is done to facilitate access control as verification handles purposes and reasons differently and will be reported on elsewhere.

Much work has been done on flow analysis, and illegal flow of information [7], unfortunately, due to lack of space, the subject cannot be considered in this paper (with this particular model and understanding of compound purposes and how the hybrid access control model is subject to such attacks); and is left for future work.

Work done on the verification phase [5] using Role Based Access Control (RBAC), requires that granting of permissions is controlled by a central authority. LeFevre et al [11] uses query rewriting to limit disclosure – also requiring a central authority to control authorisation. In most work, a subject has a set of associated purposes in his profile. These purposes are extracted by the Access Control Subsystem (ACS) of the DBMS, and matched against the purposes

bound to an object, access is granted *if the subject's profile contains any purposes that are bound to the object*. Byun et al. [5] allows the specification of *prohibited* purposes to make access control more restrictive.

The concept of extending SQL is not new and has, for example, been proposed Rosenthal et al [15] to extend the grant statement to limit the privileges that a subject receives. Allowing more flexibility in DAC in the relational environment, and ensuring that privileges are limited correctly.

### 3 Extending SQL for Specifying Reasons

The paper now proceeds to expound the proposed extensions to SQL, which are optional subclauses of the **grant**, and **select** statements. In all cases the syntax of the extensions allow the subclauses to be omitted providing a level of ease of use. The semantics of the extensions are defined to ensure predictable behaviour if they are omitted.

#### 3.1 The grant Statement

The classic grant/revoke model first proposed by Griffiths et al [9,4] creates a subject profile which stores the permissions a subject has on an object. The extended **grant** statement presented here is intended to add only two pieces of information. The reasons that the grantee may present for accessing the particular object forming part of the grant statement, and the reasons for which the grantee may grant onward (definition 1).

##### Definition 1 (Extended GRANT)

*GRANT* <privileges> [*for*<sub>1</sub>] on <object> to <subject>  
[with grant option [*for*<sub>2</sub>]][*for*<sub>3</sub>]

With  $for_i = \text{for } <reason_1, reason_2, \dots, reason_n>.$

The first of these for-subclauses (*for*<sub>1</sub>) is all the reasons for which the grantee may access the specified object. The reasons that can be specified here by the grantor depends the reasons afforded to him.

The second for-subclause (*for*<sub>2</sub>) is associated with the *grant-option* clause. This indicates the reasons for which the grantee is allowed to grant onwards. The third for-subclause (*for*<sub>3</sub>) is the reasons for which the grantor is granting onward. Note that the grantor's reasons for granting onward should be more specific (dominate) those he received.

Also note that *for*<sub>1</sub> is independent of *for*<sub>{2|3}</sub>, but that the grantee's *for*<sub>1</sub> must at least be as restrictive as the grantor's *for*<sub>1</sub> (see 3.2).

Every for-clause is a comma separated list of reasons for which an action may take place (accessing an object, or granting onward).

#### 3.2 Granting Onward

The grant option in the DAC model allows a subject to grant other subjects rights on objects that he himself has access to. These rights must be as restrictive as the rights of the grantor.



The extended grant statement ensures that the grantee does not receive more access than the grantor – it ensures that the grantee’s reasons for accessing and for granting onward is sufficiently restrictive.

**Granting reasons for granting onward.** A grantee’s grant-onward reasons are restricted by ensuring that they are more specific than the grantor’s grant-onward reasons. Suppose, for example,  $RS_2$  is the reasons that  $\sigma_1$  received from  $\sigma$  for passing on as part of a grant option. If  $\sigma_1$  issues a grant statement, passing the grant option to  $\sigma_2$  as: `GRANT  $\pi$  FOR  $RS'_1$  ON  $\omega$  TO  $\sigma_2$  WITH GRANT OPTION FOR  $RS'_2$  FOR  $RS'_3$` ; Then it is required that  $RS'_2 \geq RS_2$ . It is obvious that  $RS'_3 \geq RS_2$  must also hold.

**Granting reasons for accessing objects.** More sensitive objects will be protected by making use of reasons closer to the least upper-bound of the purpose lattice, as these purposes are more restrictive (specific). Thus, in order to adhere to the grant statement’s “narrowing” property, granting reasons onward requires that those reasons that are granted are dominated by the reasons the grantor has.

**Theorem 1.** (*Grant Onward*)

*If, for every reason  $R_i$  that a grantor passes onward he has a reason  $R_j$  such that  $R_j \geq R_i$ , the grantee’s access will not be less restrictive than the grantor’s.*

*Proof.* (*Grant Onward*)

Suppose that  $\sigma$  is granting reasons onward on  $\omega$ , and that  $\omega_P$  is the (compound) **purpose** bound to  $\omega$ .

Furthermore, suppose that  $RS_1$  is the reasons that  $\sigma$  received, and that  $RS'_1$  is the reasons he wants to grant onward.

If  $\sigma$  is only able to access portions of  $\omega$  for which  $RS_1 \geq \omega_P$ , then clearly if  $RS_1 \geq RS'_1$ ,  $\sigma_1$  will only get access to portions of  $\omega$  for which  $\sigma$  has access, or less.

Thus, if  $\forall R_i \exists R_j, R_i \in RS'_1, R_j \in RS_1$  such that  $R_j \geq R_i$ , then  $RS_1 \geq RS'_1$ , and the grantee will never be able to access information that the grantor does not have access to.

### 3.3 Accessing Objects

Access to objects can occur by either using an SQL Data Manipulation construct or by executing a procedure. Protecting access to procedures are not considered in this paper and is left for future work.

Requesting access to data through the normal constructs result in an access verification for every single construct that is found as part of the statement (query). This makes it possible to nest queries without having to redesign and implement the ACS. It is possible to merge certain types of queries [14], but this is largely done for faster query execution, and is thus not considered here. It is assumed queries can be considered in isolated form.

Every access request requires that the subject provide his reason for accessing data. In the same fashion as the grant-statement, the data access statements are augmented with an optional for-clause, however, there is a important difference between the grant statement and these statements.

Data access queries may include references to various (base or virtual) tables, and columns from these tables. Each one of these objects could have a purpose associated with it. Thus the subject must specify a reason for accessing each object, which is accomplished by having the for-clause provide a list of “key=value” pairs. Each pair consists of an object (key) and the reason for accessing that particular object. The data access statement is thus defined as per definition 2.

**Definition 2 (Extended SELECT)**

*SELECT*  $\theta$  [*for*  $\langle o_1 = r_1, o_i = r_2, \dots \rangle$ ]

Where  $\theta$  represents the body of the statement,  $o_i$  is an object that is known and protected by the DBMS, and  $r_i$  is a reason.

For example, suppose a subject wishes to get the names of all the customers in the database for “Public Relations” purposes. The query is presented as (for example purposes it is assumed that purposes are presented as strings): “**select name from cust for <name=“PR”, cust=“PR”>**”.

It is possible that a subject could request access to several objects. Specifying a reason for each explicitly can become tedious, thus it is possible to omit an object from the list. The ACS will automatically associate the greatest lower-bound of the purpose lattice as the reason for accessing that object, thereby assuming the subject is accessing the object for the most general reason.

If several objects are accessed for the same reason, the special keyword “default” is introduced. A subject can use this “object” to specify his default reason for accessing objects. For example in the query “**select name from cust for <default=“PR”>**” the ACS interprets the access request to mean that **name** and **cust** is accessed for the reason “PR”. If the “default” object is not set explicitly, the greatest lower bound of the purpose lattice will be associated with it (see 4).

**Reason Inference.** Subjects often access a table, and columns from that table for the same reason. The subject can omit the reason for access to the table, and the ACS can infer this reason by examining the reason specified for the columns.

If multiple columns from the same table are accessed, a compound reason for the table (consisting of the reasons specified for the columns of the table) is formed by using the **and** operator (defined in [16]).

In the same way that SQL requires each object in a SQL query to be uniquely identifiable (using fully qualified names in case of ambiguity), the ACS must be able to associate a reason with every object. Fully qualified names should be used in cases of ambiguity.

### 3.4 Controlling Access

A simple algorithm is presented in listing 1.1, which can aid in the verification process based on the extensions that have been given. Before the algorithm is provided, it is necessary to define several auxiliary functions.

**Listing 1.1.** Algorithm for Determining Access

```

1 PDACS(s, Olist)
2   default = def(Olist)
3   result = true
4   for o in Olist
5     if null(reason(o)) then
6       r = default
7       for o2 in getColumnsOf(o, Olist)
8         if null(reason(o2)) then
9           r = and_r(r, default)
10        else
11          r = and_r(r, reason(o2))
12        endif
13      endfor
14    else
15      r = reason(o)
16    endif
17    result = result & verify(s, o, r)
18  endfor
19  return result
20 end

```

1. *reason(o)*: returns the reason that is associated with an object as specified in the query, if no reason is associated, the a *null* value is returned. *null(o)* returns true if *o* is a null value.
2. *and<sub>r</sub>(r<sub>1</sub>, r<sub>2</sub>)*: forms an **and** conjunction of the reasons *r<sub>1</sub>* and *r<sub>2</sub>*.
3. *getColumnsOf(o, Olist)*: returns a list of all the objects in the list *Olist* which are columns of the first parameter *o* – if *o* is not a table, an empty list is returned.
4. *verify(s, o, r)*: returns true if *s* can access *o* for reason *r*.
5. *def(list)*: will return the “default” value if it is defined in *list*, else it returns the greatest lower-bound of the purpose lattice.

The Purpose Driven Access Control Subsystem (*PDACS*) function takes a list of objects (specified in the query), and the ID of a subject. It finds the default reason, iterates through the list of objects, and attempts to find a reason for every object in the list. If no reason is found for an object, it is assumed that the object is a table. In this case *PDACS* will try to infer the reason for accessing the table by finding the columns of that table that are listed in the query, their associated reasons, and constructing a compound reason for the table.

During each iteration verification for a specific object takes place, and is “anded” with past results. Any “false” returned by *verify* will thus result in access being denied completely.

## 4 Omitting the *for* Clauses

Having the *for*-subclauses as optional, allows users to continue using the **grant** and **select** statements as per usual. In those cases where any of the *for*-clauses are omitted the ACS modifies the statement to include the *for* clauses with *default* values.

Default values for **grant** can be inferred from the reasons associated with the grantor. Where a grantor has no specific reasons, the greatest lower bound of the purpose lattice is substituted.

Suppose for example subject  $\sigma$  is granted access to object  $\omega$  with grant-statement (in which  $RS_i$  represents a list of reasons): **GRANT SELECT FOR  $RS_{1:\sigma}$  ON  $\omega$  TO  $\sigma$  WITH GRANT OPTION FOR  $RS_{2:\sigma}$  FOR  $RS_3$ ;**

If  $\sigma$  wishes to grant access on  $\omega$  to subject  $\sigma_2$ , he can issue the grant statement: **GRANT SELECT ON  $\omega$  TO  $\sigma_2$  WITH GRANT OPTION;** The ACS will insert the missing *for*-clauses in the grantor's statement, based on those given to the grantor. Changing the statement to: **GRANT SELECT FOR  $RS_{1:\sigma}$  ON  $\omega$  TO  $\sigma_2$  WITH GRANT OPTION FOR  $RS_{2:\sigma}$  FOR  $RS_{2:\sigma}$ ;**

In the case where the *for* clause of the **select** is omitted, the ACS associates the most general purposes in the purpose lattice with the "default" object. For example, "**select name from cust;**" is changed to "**select name from cust for <default="noreason">**" (provided that the most general purpose is "noreason").

## 5 Conclusion

This article argued that subjects should actively participate in specifying what they want to do with data (using purposes) in order to facilitate better auditing.

To accommodate active specification of reasons, a simple extension to the SQL **grant**, and **select** was proposed. The extension to **select** is used to specify the reason for which data is accessed, and the extended **grant** statement allows subjects to delegate access control, whilst ensuring that it is impossible to pass on less restrictive reasons.

The impact of using these extensions are low, as the clauses can be safely omitted with well defined results, allowing subjects to continue using the SQL statements as per usual.

Future work on the topic includes extensions to **revoke**, and other data access statements, as well as flow analysis for determining illegal flow of information.

## References

1. AGRAWAL, R., KIERNAN, J., SRIKANT, R., AND XU, Y. Hippocratic databases. In *Proceedings of the 28th VLDB Conference* (Hong Kong, China, 2002).
2. ASHLEY, P., HADA, S., AND KARJOTH, G. E-p3p privacy policies and privacy authorisation. In *WPES'02* (Washington, November 2002).

3. ASHLEY, P., HADA, S., KARJOTH, G., POWERS, C., AND SCHUNTER, M. Enterprise privacy authorisation language (EPAL 1.1). Tech. rep., International Business Machines Corporation, 2003.
4. BERTINO, E. Data security. *Data and Knowledge Engineering* 25, 2 (1998), 199–216.
5. BYUN, J.-W., BERTINO, E., AND LI, N. Purpose based access control of complex data for privacy protection. In *SACMAT'05* (Stockholm, Sweden, June 2005), ACM.
6. CRANOR, L., LANGHEINRICH, M., MARCHIORI, M., PRESLER-MARSHALL, M., AND REAGLE, J. The platform for privacy preferences (P3P1.0) specification. Tech. rep., W3C, Available at <http://www.w3.org/TR/P3P/>, 2002.
7. FISCHER-HÜBNER, S. *IT-Security and Privacy: Design and Use of Privacy-Enhancing Security Mechanisms*. Springer-Verlag, 2001.
8. FISCHER-HÜBNER, S., AND OTT, A. From a formal privacy model to its implementation. In *21st National Information Systems Security Conference* (Arlington, VA, USA, October 1998).
9. GRIFFITHS, P. P., AND WADE, B. W. An authorization mechanism for a relational database system. *ACM Transactions on Database Systems (TODS)* 1, 3 (1976), 242–255.
10. HES, R., AND BORKING, J., Eds. *Privacy Enhancing Technologies: The Road to Anonymity*, revised ed. Dutch DPA, 1998.
11. LEFEVRE, K., AGRAWAL, R., ERCEGOVAC, V., RAMAKRISHNAN, R., XU, Y., AND DEWITT, D. Limiting disclosure in hippocratic databases. In *30th International Conference on Very Large Data Bases* (Toronto, Canada, 2004).
12. OASIS ACCESS CONTROL TC. OASIS extensible access control markup language (xacml) version 2.0. Tech. rep., OASIS, February 2005.
13. OECD guidelines on the protection of privacy and transborder flows of personal data. Tech. rep., Organisation for Economic Co-operation and Development, 1980.
14. PIRAHESH, H., HELLERSTEIN, J. M., AND HASAN, W. Extensible/rule based query rewrite optimization in starburst. In *SIGMOD Conference on the Management of Data* (San Diego, California, 1992), ACM.
15. ROSENTHAL, A., AND SCIORE, E. Extending SQL's grant operation to limit privileges. In *Data and Application Security, Development and Directions, IFIP TC11/WG11.3 Fourteenth Annual Working Conference on Database Security* (August 2000), B. M. Thuraisingham, R. P. van de Riet, K. R. Dittrich, and Z. Tari, Eds., Kluwer, pp. 209–220.
16. VAN STADEN, W. J., AND OLIVIER, M. S. Purpose organisation. In *Proceedings of the fifth annual Information Security South Africa (ISSA) Conference* (Sandton, Johannesburg, South Africa, June 2005).

# FGAC-QD: Fine-Grained Access Control Model Based on Query Decomposition Strategy

Guoqiang Zhan, Zude Li, Xiaojun Ye, and Jianmin Wang

School of Software, Tsinghua University, Beijing, 100084, China  
{zhan-gq03, li-zd04}@mails.tsinghua.edu.cn,  
{yexj, jimwang}@tsinghua.edu.cn

**Abstract.** Applications require fine-grained access control (FGAC) supported by DBMSs themselves. Though much literature has referred to the FGAC, its key problems still remain open. Thus, we develop a FGAC-QD model based on query decomposition strategy with incorporating two notions of *authorization rule* and *predicate transitive rule*. In our model, users' queries are decomposed into a set of one-variable queries (OVQ). For each OVQ, its validity is checked against the corresponding authorization rule; if all the OVQs are valid, the query is inferred to be valid and will be executed without any modification; otherwise the query has illegal access, and will be partially evaluated or rejected directly, according to the feature of applications. Finally, the results of experiments demonstrate the feasibility of FGAC-QD.

## 1 Introduction

In recent years, researchers have paid much attention to fine-grained access control (FGAC<sup>1</sup>) in databases. A motivation for FGAC is that it cannot be bypassed and therefore a high assurance of security and privacy can be achieved [1, 10], by enforcing access control in databases. Also, FGAC models are extended to support the privacy preserving as described in [2, 4].

So far, FGAC models are implemented by two approaches. One approach, called view-based approach, uses views or parameterized views based on the technique of query modification [11]. In other words, users' queries are restricted to access these views or modified by them transparently, that is, users are limited to access only those parts of the tables that are defined by views. The other approach uses labelling, where each data element (e.g., a cell or a record) is labelled with information identified its secure classification [6].

View-based approaches suffer from incorrect or unexpected partial result (such as Oracle VPD [1, 10]) or the change of query structure (such as Motro Approach [4, 8]) transparently based on the technique of query modification. However, labelling approach is not flexible to define FGAC policies, changes the relational data model and leads to high cost [6].

Non-Truman model tries to give a solution to limitations for view-based approach in [10] by checking whether users' queries can be answered using authorized views. While, the technique of answering queries using views cannot

---

<sup>1</sup> FGAC is specially referred to DBMS-level fine-grained access control.

validate complex queries efficiently, which is a NP problem [5], and the inference rules are incomplete in Non-Truman Model [10]. So, we propose a new FGAC model based on query decomposition (called FGAC-QD) to give a new way towards view-based FGAC model. In this model, we introduce the concept of authorization rules to define FGAC policies and support content-based authorization management, sequentially, develop an approach to checking whether users' queries are allowed by FGAC policies based on query decomposition strategy [13]. Based on the technique of query decomposition, we can notify users whatever they violate FGAC policies exactly, and can modify queries conditionally according to applications if necessary.

The main contributions include: 1) Develop the concept of authorization rule to define FGAC policies and support content-based dynamic authority management. 2) Develop FGAC-QD model based on query decomposition by introducing the notion of predicate transitive rule.

The rest of the paper is organized as follows. In Section 2, we briefly summarize related work on FGAC models. Sequentially, we describe how FGAC policies are defined in FGAC-QD model. In Section 4, we develop a simple FGAC-QD model to process simple conjunctive queries. In Section 5, in order to support complex queries, we extend the simple FGAC-QD model based on techniques of predicate inference and query decomposition. In Section 6, we introduce the implementing architecture of FGAC-QD model and demonstrate the feasibility of our model by experiments.

## 2 Related Work on FGAC

In commercial DBMSs, FGAC is implemented by the technique of query modification which is introduced into INGRES firstly [11]. Oracle VPD [1] is such a technique. This strategy suffers from some problems argued by Shariq Rizvi [10]: user queries will be modified transparently by the query engine, as will lead to semantic inconsistency by returning partial result without warnings; besides by incorporating the predicates derived from the FGAC policy, user queries may become more complex, and degrades the performance of query optimizer and executor, especially if nested queries are incorporated into the user query.

In [8], Motro proposed an access control model by introducing the notion of access permission views. In this model, FGAC policies are defined by a set of views  $V = \{v_1, \dots, v_m\}$  which are decomposed into each relation and stored as meta-relations. When a query is submitted to DBMS, the query is performed both on the meta-relations and the actual relations. A mask is derived with user queries evaluating on the meta-relations, and an answer is generated on actual relations. Before the answer is returned to users, the mask is applied to the answer, yielding the partial result that may be delivered to users. The approach of Motro can actually provide FGAC, but it also poses some subtle problems [4], such as the change of query structure.

Motivated by drawbacks of the technique of query modification, Shariq Rizvi et al. proposed a Non-Truman Model in which each user is associated with a set of parameterized authorized views [10]. As a result, when a user issues a query,

the system validates whether the query can be answered against the views associated with the user based on the technique of answering queries using views [5]. If the answer is positive, then the query is processed without any modification, otherwise, the query is rejected directly. Unfortunately, the inference rules for both the conditional validity and unconditional validity are incomplete [10]; besides, the technique of answering queries using views suffers from NP problem [5].

### 3 Definition of FGAC Policies

In this paper, we develop a FGAC-QD model in which the concept of *authorization rule* is introduced to provide a flexible way to define FGAC policies which can be granted to users as privileges defined in SQL do.

#### 3.1 Definition of Authorization Rule

**Definition 1:** Authorization Rule (AR) is a statement used to describe FGAC policies, and it is presented as a quaternion below:

AR:  $\langle \textit{Event}, \textit{Object}, \textit{Constraint}, \textit{Action} \rangle$  in which

**Event** means queries' types: SELECT, UPDATE, DELETE, INSERT.

**Object** refers to the schema object (table or view) on which AR is defined on.

**Constraint** is to decide whether the authorization rule can be used to validate current user requests by evaluating them against the current database instance, so that it can support the content-based dynamic authorization management, and users' privileges in DBMS evolve along with the database instance.

**Action** describes the actual access policy on *Object*, and is defined as  $\langle \textit{AAS}, \textit{TSF} \rangle$  where:

- **TSF** is a Boolean expression which is used to filter out tuples from *Object*. TSF may contain parameters which improve the scalability of authorization rules as parameterized views do in [10]. When authorization rules are activated, the parameters will be fixed by the user's session or application's context.
- **AAS** is a set of attributes which can be accessed in *Object*.

By extending **Action**, an AR is defined as:  $\langle \textit{Event}, \textit{Object}, \textit{Constraint}, \textit{AAS}, \textit{TSF} \rangle$ . The syntax for authorization rules in our prototype system is described as:

**CREATE AUTHORIZATION RULE**  $\langle \textit{rule\_name} \rangle$

**AS ON**  $\langle \textit{event} \rangle$  **TO**  $\langle \textit{object} \rangle$  **WHERE**  $\langle \textit{Constraint} \rangle$  **DO**  $\langle \textit{Action} \rangle$

**EXAMPLE** :Define a content-based FGAC policies. Given a sale record table: Records (User\_id, Order\_id, Book\_id, Fee, Amount, Order\_time, Status). Then an authorization rule **ShoppingGuide** which allows the user, whose total shopping amount exceeds 1000\$, to find the hot books by rank of sales, is defined as below.

**CREATE AUTHORIZATION RULE** ShoppingGuide **AS ON** select



**TO** Record **WHERE** exists (select 1 from Record where User\_id=\$User\_id having count (Fee) >1000\$) **DO** select Book\_id, count (Book\_id) from Record group by Book\_id sort by Book\_id;

### 3.2 Physical Storage Structure of Authorization Rule

In order to facilitate the validation for user's queries against authorization rules, we will negate TSF in AR, and then normalize it into DNF, finally modified AR store is into system catalog as meta-data. So, given an authorization rule  $\langle E, O, C, AAS, TSF \rangle$ , the storage structure of authorization rule is  $\langle E, O, C, AAS, TSF, \overline{TSF} \rangle$ , in which TSF is negated, and  $\neg TSF$  is denoted as  $(\overline{TSF} = TSF_1^* \vee \dots \vee TSF_n^*)$  in which each  $TSF_i^*$  is a simple conjunctive expression.

## 4 Simple FGAC-QD Model

We propose a simple FGAC-QD model which assumes that user's requests are simple conjunctive queries without nested queries, and these queries only involve one object. In this section, we define the concepts of *one-variable query* and *query validity* firstly, and sequentially give out the algorithm of query validity validation.

**Definition 2 One-Variable Query:** (OVQ) refers to a simple conjunctive query only involving one relation without any implicit predicates in its selection expression. OVQ is formally presented as  $\langle Object, TSF, AAS, Event \rangle$ , in which *Object*, *AAS* and *Event* are same to those of Authorization Rule, while *TSF* must be a conjunctive Boolean expression without implicit predicates.

Based on the technique of answering query using views [5], given a query Q and a set of views  $V : \{v_1, \dots, v_n\}$ , if there exists a derived view Q' defined only on V, which can give the same returned result as Q at any database instances. Then Q' is equivalent to Q, Q is contained in view set V, and the relationship between Q and V is called *containment*, denoted as  $Q \subseteq V$ . Actually, the action in an authorization rule is a query, thus based on containment, query validity is defined as:

**Definition 3 Query Validity:** Given a query Q and a set of available authorization rules  $ARS : \{AR_1, \dots, AR_n\}$ , if Q is contained in ARS (that is,  $Q \subseteq ARS$ ), then Q is valid.

In this preliminary model, it is assumed that each user's query is OVQ, and only one authorization rule can be authorized to per user on per object per event. Thus, we just need to check the containment between one OVQ and an authorization rule. Given an  $OVQ_0 : \langle O_{OVQ_0}, TSF_{OVQ_0}, AAS_{OVQ_0}, E_{OVQ_0} \rangle$  and an  $AR_0 : \langle E_{AR_0}, O_{AR_0}, C_{AR_0}, ASS_{AR_0}, TSF_{AR_0}, \overline{TSF_{AR_0}} \rangle$ , if  $OVQ_0 \subseteq AR_0$  is true, then  $AAS_{OVQ_0} \subseteq ASS_{AR_0}(I)$  and  $TSF_{OVQ_0} \wedge TSF_{AR_0} \equiv True(II)$

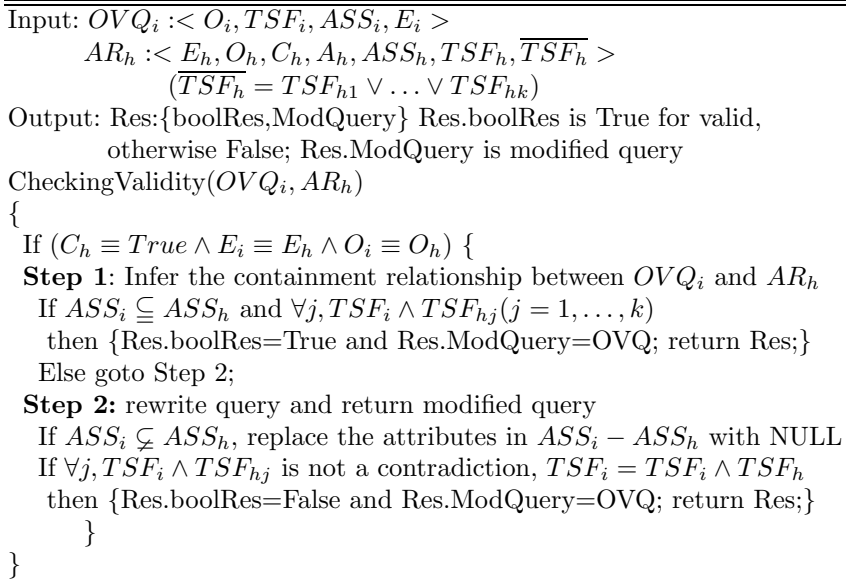
are held. We can easily validate (I), but it is some difficult [8] to (II) which is equivalent to  $TSF_{OVQ_0} \wedge \overline{TSF_{AR_0}} \equiv False$ .

According to logical algebraic theories, we can conclude the following proposition.

**Theorem 1.**  $\forall B_i(x) (i = 0, 1, \dots, n)$  is a boolean expression,  $B_0(x) \wedge (B_1(x) \vee \dots \vee B_n(x)) \equiv false$  is true, iff  $\forall B_0(x) \wedge B_j(x) \equiv false (j = 1, \dots, n)$  is true.

Assuming  $\overline{TSF_{AR_0}} = B_{AR_{01}}^* \vee \dots \vee B_{AR_{0k}}^*$ . So, the above (II) is extended to be  $TSF_{OVQ_0} \wedge (B_{AR_{01}}^* \vee \dots \vee B_{AR_{0k}}^*) \equiv False$ . So, to validate validity of the user's query, we just have to check if each  $TSF_{OVQ_0} \wedge B_{AR_{0i}}^* (i = 1, \dots, k) \equiv False$  is held. Also,  $TSF_{OVQ_0}$  and  $B_{AR_{0i}}^*$  are both conjunctive expression, so  $TSF_{OVQ_0} \wedge B_{AR_{0i}}^* (i = 1, \dots, k)$  is also a conjunctive expression. So, if any  $TSF_{OVQ_0} \wedge B_{AR_{0i}}^*$  is not a contradiction, then the user's query is not contained in its authorization rule, and is not valid.

Thus, validation algorithm is described in Fig 1. We check whether users' queries violate the authorized authorization rules and enforce FGAC policies by the technique of query modification if necessary. If applications accept partial results, the modified version of user's request is executed; otherwise user's request is rejected directly.



**Fig. 1.** Checking Validity Algorithm for OVQ

## 5 An Extending FGAC-QD Model

Usually, user queries are arbitrary and will be decomposed into a sequence of OVQs. Thus, for each OVQ, the corresponding authorization rules are activated

to check its validity. If all the OVQs are valid, the query is inferred to be valid; otherwise the system notifies user with the violations and rewrites the query.

Given a user's query  $Q : \{OS, TSF(x), AAS, Event\}$  in which  $OS$  is a set of objects, firstly, we transform  $TSF(x)$  into DNF:  $TSF(x) = TSF_1(x) \vee \dots \vee TSF_m(x)$ . Then  $Q$  is split into conjunctive queries (CQ) by eliminating logical operator OR, denoted as a set  $CQS = \bigcup CQ_i$  ( $CQ_i : \{OS, TSF_i(x), AAS, Event\}$ ,  $i = 1, \dots, m$ ).

## 5.1 Predicate Inference

Before decomposing a conjunctive query (i.e.  $CQ_i : \{OS_i, TSF_i(x), AAS_i, E_i\}$ ), we have to generate all the implicit predicates from  $TSF_i$  to ensure the completeness of accessing information on each objects. For example, let  $TSF$  be  $\{T1.col1 > 10 \wedge T2.col1 = T1.col1\}$ , if we don't infer implicit predicates from  $TSF$  before decomposing the query, then  $OVQ$  for  $T2$  will lose the access information  $T2.col1 > 10$ . And thus it may lead to incorrect result of validation against granted authorization rules. So we will introduce the algorithm of predicate inference based on the concept of predicate transitive rules.

In this paper, we only take operators  $\{>, \geq, <, \leq, =\}$  into consideration. These operators are classified into two categories: *BigSet* :  $\{>, \geq, =\}$  and *SmallSet* :  $\{<, \leq, =\}$ , and in each class priority for each operator is descending. Thus, given two operators  $\theta_1, \theta_2$ , if they belong to the same class, and  $\theta_1$  is prior to  $\theta_2$ , then their relationship is denoted as  $\theta_1 \geq \theta_2$ .

**Definition 4 Predicate Transitive Rule:** Given two predicates:  $\langle X, \theta_1, Y_1 \rangle$  and  $\langle Y_1, \theta_2, Z \rangle$ , if  $\theta_1 \geq \theta_2$  is true, then  $P^* : \langle X, \theta_1, Z \rangle$  is concluded, thus  $\{\langle X, \theta_1, Y_1 \rangle, \langle Y_1, \theta_2, Z \rangle\} \rightarrow \langle X, \theta_1, Z \rangle$  is a predicate transitive rule.

Given a  $CQ_i : \langle OS_i, TSF_i, AAS_i, Event_i \rangle$ , when we use predicate transitive rule to generate predicates from  $TSF_i$ , we must pay attention to the complexity and features of nested queries in  $TSF_i$ , especially correlated nested queries. We also have to know to which nested queries the predicates belong, so we assign each nested query with a unique number. Assuming there are  $(n-1)$  nested queries in the  $CQS_i$ , so each nested query is assigned with a number  $L$  ( $L = 1, \dots, (n-1)$ ), and  $L$  for the top query is assigned with zero. Thus, each object (including table, view or column) in  $OS_i, TSF_i$ , and  $AAS_i$  must identify nested queries which it belongs to. Finally, assuming predicates are presented as the following form:  $\langle X, \theta, Y \rangle$ , and they are extended to be  $\langle (X, L_1), \theta, (Y, L_2), L_3 \rangle$  ( $L_1, L_2, L_3$  are possible not to equal with each other). All predicates in  $B(x)$  are clustered into two sets:  $B_{simple}(x)$  and  $B_{join}(x)$ .  $B_{join}(x)$  is a set of join clauses (JC, i.e.  $T1.col1 \theta T2.col2$ ), while  $B_{simple}(x)$  is a set of simple predicates (SP) except join clauses. Now, based on predicate transitive rule and extension of predicate presentation, we can infer implicit predicates by algorithm in Fig 2. All the new implicit simple predicates are appended into  $B_{simple}(x)$ , and a join predicate can be added into  $B_{simple}(x)$  if two attributes of it come from the same objects belonging to the same nested queries, that is,  $L$  is equal.

## 5.2 Query Decomposition

By predicate inference, all the implicit predicates are generated, and appended into  $B_{simple}(x)$ . So, according to the definition of OVQ, the main tasks of query decomposition are to tie relevant predicates from  $B_{simple}(x)$  to specific relations in CQ and form ASS for them.

Given a conjunctive query  $CQ_i :< OS_i, TSF_i, AAS_i, Event_i >$  with  $(n - 1)$  nested queries in  $TSF_i$ , that is  $L$  varies from 0 to  $(n - 1)$ . The decomposition of  $CQ_i$  follows the algorithm described in Fig.3, and all OVQs are stored in a set, denoted as SOVQ.

---



---

```

PredicateInference( $B_{simple}(x), B_{join}(x)$ )
{ For each  $JC_i :< (X_i, L_{i1}), \theta_i, (Y_i, L_{i2}), L_{i3} >$  from  $B_{join}(x)$ 
Do { FLAG= False
Step 1: Generate Simple Implicit Predicate
 $\forall SP_j :< (X_j, L_{j1}), \theta_j, (Y_j, L_{j2}) > \in B_{simple}(x)$ ;
if  $(Y_i, L_{i2}) = (X_j, L_{j2}) \wedge (\theta_1 \geq \theta_2)$  then generate a predicate
 $SP^* :< (X_i, L_{i1}), \theta_i, (Y_j, L_{j2}), L_{i1} >$ ,  $B_{simple}(x) = B_{simple}(x) \cup \{SP^*\}$ 
Step 2: Generate Join Predicate
 $\forall JC_h :< (X_i, L_{i1}), \theta_h, (Y_j, L_{j2}), L_{h3} > \in B_{join}(x)$ 
if  $(Y_i, L_{i2}) = (X_h, L_{h1}) \wedge (\theta_1 \geq \theta_h)$  then
{
FLAG = TRUE; Generate implicit join predicate  $JC^*$ :
 $JC^* :< (X_h, L_{h1}), \theta^*, (Y_h, L_{h2}), L_{i1} >$  ( $\theta^* = \max(\theta_h, \theta_i)$ )
If  $(X_i, Y_h \in O^*) \wedge (L_{i1} == L_{h2})$  then  $B_{simple}(x) \cup JC^*$ ;
 $JC_i = JC^*$ 
}
} While(FLAG)
}
```

---



---

**Fig. 2.** Algorithm of Implicit Predicate Inference

---



---

```

Decomposition( $CQ_i, B_{simple}(x), SOVQ$ ) {
For each  $< Q_k, L_k >$  in  $OS_i$ ,  $OVQ_{O_k}$  is constructed as:
Step 1: Form  $AAS_{O_k}$ 
 $\forall < col_j, L_j > \in AAS_i$ , if  $(col_j \in O_k \wedge L_j = L_k)$  then
 $AAS_{O_k} = AAS_{O_k} \cup \{col_j\}$ 
Step 2: Form  $TSF_{O_k}$ 
 $\forall P_i :< (X_i, L_{i1}), \theta_i, (Y_i, L_{i1}), L_{i1} > \in B_{simple}(x)$ 
if  $((L_{i1} = L_k) \wedge (X_i \in O_k))$  then  $P_i^* :< X_i, \theta_i, Y_i >$  is formed.
 $P_i^*$  is added into  $TSF_{O_k}$  as conjuncts:  $TSF_{O_k} = \bigwedge P_i^*$ 
Step 3: Finally, form  $OVQ_{O_k}$ 
 $E_{O_k}$  is assigned as the type of  $CQ_i$ ;
 $OVQ_{O_k} = < O_k, TSF_{O_k}, AAS_{O_k}, E_{O_k} >$ ;  $SOVQ = SOVQ \cup \{OVQ_{O_k}\}$ 
}
```

---



---

**Fig. 3.** Algorithm for Query Decomposition

### 5.3 Checking Validity

A conjunctive query CQ has been decomposed into a sequence of OVQs collected in the set SOVQ, for each  $OVQ_i$  from SOVQ, the eligible authorization rule is activated to check its validity according to the validation algorithm (in Fig. 1). Given an  $OVQ_i$ , when  $OVQ_i$  violate authorization rules  $AR_i$ , we have to modify CQ but not  $OVQ_i$  directly: append the *TSF* in  $AR_i$  to the CQ as a conjunct, as is different from that in simple FGAC-QD model.

It seems that our model is similar with Non-Truman Model, actually, there are great differences between the two. In FGAC-QD model, FGAC policies are defined by authorization rules, which are defined on one relation only. In order to avoid the complexity of answering queries by views, our model introduces the techniques of predicate inference and query decomposition, and based on these techniques, we can determine whatever users violate the FGAC policies exactly.

To summarize our FGAC-QD model, a user's query is transformed into a set of conjunctive queries, then these conjunctive queries are extended by predicate inference algorithm, and finally are decomposed into a set of one-variable queries which will be checked against the corresponding authorized authorization rule one by one. If all the OVQs for all conjunctive queries are valid, the user's query is valid and executed without any modification. Otherwise, according the validation algorithm, exact notification of violating information against authorization rules is sent to the user, and the system returns a modified query by adding TSF in the authorization rules which user violates as conjuncts into the user's query. If a partial result is acceptable, this modified query is executed, and a partial result is returned, otherwise the user's query is rejected directly.

## 6 Implementation of FGAC-QD Model

We use PostgreSQL to implement and validate our model. PostgreSQL Rule System (PRS) [12] has provided a flexible mechanism to support our authorization rule. Based on PRS, we extend existing rule syntax. However, relations that are used due to rules get checked against the privileges of the rule owner, not the user invoking the rule. This means that a user only needs the required privileges for the tables/views that he/she refers explicitly in his queries. Obviously, this feature is not allowed in FGAC-QD model, only if a user has authorized with the authorization rule, it can be activated to check or modify users' queries. So, we have to extending existing PRS model to support the management of authorization rules, such as granting, revoking and so on.

According to above three algorithms, we have developed validation module based on Theorem 1, including inference module, decomposition module and query modification module. The implementing architecture of FGAC-QD model consists of FGAC policies management (which is to define security policies by authorization rules and manage the authorization distribution), FGAC policies enforcement, and violation notification. These modules form a new facility interposing between the parser and rewriter in PostgreSQL.

Based on the above implementation, we do some experiments to evaluate the performance of FGAC-QD model in OLTP applications according to TPC-C benchmark. In our experiment, we define access policies for each target relation by authorization rule: each user can access all the data in each relation, so each query has to suffer from FGAC-QD model and can be validated as a valid query. As a result, all the transactions can be processed by TPC-C model normally. The experiment environment is set up as below: Redhat 9.0 (OS), 180G hard disk, P4 2.6G (CPU), 1G memory. System parameters for tested DBMS are: 11 Warehouses, shared\_buffer=35000, wal\_buffers = 128, checkpoint\_segments=128, checkpoint\_timeout=1800. The results of TPC-C tests are present in the following Fig.4.

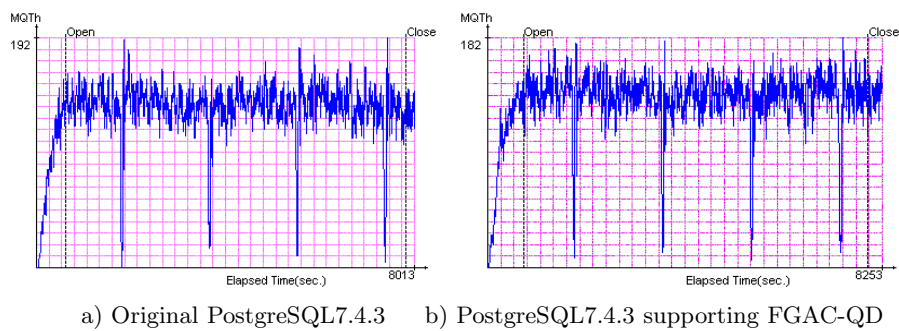


Fig. 4. TPC-C Test Result

Their transaction throughput is 136.35 tmpC for a) and 135.08 for tmpC for b) respectively, and corresponding responding times are presented in Table 1. The column of Ratio% shows the increase percentage of responding time b) against a).

Table 1. Comparison of Responding times

	90% (seconds)			average (seconds)			maximum (seconds)	
	(a)	(b)	(Ratio%)	(a)	(b)	(Ratio%)	(a)	(b)
NewOrder	1.16	1.56	34.48	0.99	1.22	23.23	58.9	60.63
Payment	1.01	1.31	19.80	0.68	0.78	17.71	60.4	58.26
StockLevel	0.6	0.84	40.00	0.47	0.4	-17.50	47.26	50.55
Delivery	1.29	1.82	41.09	0.77	1.05	36.37	49.13	55.62
OrderStatus	1.08	1.37	26.85	0.66	0.93	40.91	26.53	55.67

The result of experiments demonstrate that there is little decrease in throughput, and there is great performance degradation in responding time, about reduced by 20% ~ 30%. The result is expectable, because security facility always cost many computing resources. Considering the balance between security and performance, the performance is somewhat acceptable.

## 7 Conclusion

We have developed the notion of authorization rule to declared fine-grained security policies and develop a simple FGAC-QD model to process simple conjunctive queries. Based on the technique of query decomposition and the notion of predicate transitive rules, we extend the simple FGAC-QD model to process complex queries. Because of the technique of query decomposition, our model avoids the pitfalls of the query modification and the complexity of answering queries using views, and can efficiently to process most complex queries. Queries are modified if there is violation against the authorization rule. And according to applications, if a partial result is acceptable for an application, then the partial result for the query is returned; otherwise the query is rejected directly. By supporting the feature of parameter and constraint in authorization rule, our model provides a more flexible content-based dynamic authorization mechanism.

## References

- [1] The Virtual Private Database in Oracle9ir2: An Oracle Technical White Paper. <http://otn.oracle.com/deploy/security/oracle9ir2/pdf/vpd9ir2twp.pdf>.
- [2] R. Agrawal, P. Birdz, T. Grandisony, J. Kiernany, S. Loganz, and W. Rjaibi. Extending Relational Database Systems to Automatically Enforce Privacy Policies, In Proc. of ICDE, 2005:1013-1022.
- [3] G.J. Ahn and R.Sandhu. Role-based authorization constraints specification. ACM Transactions on Information and System Security,3(4),2000: 207 - 226.
- [4] E. Bertino, J.W. Byun, and N.H. Li. Privacy-Preserving Database Systems, LNCS 3655, 2005:178-206.
- [5] A.Halevy. Answering queries using views: A survey. The VLDB Journal, 10(4), 2001. pages: 270-294.
- [6] S. Jajodia and R. Sandhu. Toward A Multilevel Secure Relational Data Model .In Proceedings of SIGMOD Conference 1991: 50-59.
- [7] T.F. Keefe, B.M. Thuraisingham, W. T. Tsai: Secure Query-Processing Strategies. IEEE Computer 22(3), 1989: 63-70.
- [8] A Motro, An access authorization model for relational databases based on algebraic manipulation of view definitions. In Proc. Of ICDE 1989: 339-347.
- [9] R. Pottinger, A. Levy. A Scalable Algorithm for Answering Queries Using Views. In Proc. of VLDB 2000: 484 - 495.
- [10] S. Rizvi, A. Mendelzon, S. Sudarshan, and P. Roy: Extending Query Rewriting Techniques for Fine-Grained Access Control. In Proc. of SIGMOD 2004: 551-562.
- [11] M. Stonebraker and E. Wong, Access control in a relational database management system by query modification, In Proc. of ACM Conference 1974: 180-186.
- [12] M.Stonebraker, and al."On rules, procedures, caching and views in database systems", In Proc. of SIGMOD 1990: 281-290.
- [13] E. Wong and K. Youssefi. Decomposition-A Strategy for Query Processing. ACM Transactions on Database Systems, 1(3),1976: 223-241.

# A Framework for Modeling Restricted Delegation in Service Oriented Architecture

Muhammad Alam, Michael Hafner, Ruth Breu, and Stefan Unterthiner

Quality Engineering, University of Innsbruck  
Innsbruck, Tirol, Austria

{muhammad.alam, m.hafner, ruth.breu, csad2761}@uibk.ac.at

**Abstract.** We present a novel approach for modeling restricted delegation of rights in a distributed environment based on web services. Our approach is based on SECTET-PL [5], a predicative language for modeling access rights based on the concept of Role Based Access Control (RBAC). SECTET-PL is part of the SECTET framework for model-driven security for B2B workflows. Our Rights Delegation Model combines the concept of roles from RBAC with the predicative specification of SECTET-PL. The Rights Delegation Models are translated into XACML Delegation Policies, which are interpreted by a security gateway.

## 1 Introduction

Conventionally, trust is enforced by a central authority that knows all actors and possibly all relationships between them. But the realization of the concept of trust through a central authority is not always a viable option. In distributed scenarios, where actors do not know all of their partners, where they cannot keep track of every relationship between all of them and do not want a central authority to enforce access rights to the resources in their domains, authorization remains a local responsibility and thus distributed by nature. Delegation of rights is a concept that supports the notion of trust in distributed environments and thereby fosters cooperation of partners across domain boundaries.

Inter-operability – always a major issue in the context of distributed scenarios – is meant to be guaranteed by the use of standards. Current approaches provide solutions for the specification of delegation policies based on proprietary standards, protocols and technologies. They are only suitable for a group of peers and of restricted usability. In order to provide maximum flexibility, a framework for the realization of the concept of trust in a distributed environment would have to rely on *open* standards. Some efforts of OASIS [18] focus on the standardization of a framework for the specification of restricted delegation policies but the approaches and the resulting solutions (e.g. eXtensible Access Control Mark up Language (XACML) [26]) remain very close to the technical level and are therefore inaccessible for the domain expert or business analyst.

In order to provide a satisfying alignment between the security requirements in a particular delegation scenario and a corresponding implementation, all stakeholders involved in the realization of the distributed system – from the domain



experts to the software engineers – must have a common understanding of the security requirements, each one at the appropriate level of abstraction. A *model-driven* framework [17] is most appropriate to abstract the complexities of the technical platform but remaining expressive enough to model rights and their delegation at the domain level.

The main innovative features we propose in this paper is a model-driven framework for restricted delegation of rights in a Service Oriented Architecture (SOA) – an integral aspect of our model-driven Trust Management (TM) framework. Restricted delegation of rights means that rights of the delegator or the delegatee may not depend only on their roles but also on other kind of information like credentials of the delegator, data of the business logic or parameters of the delegated web service. For more information about our approach, we refer e.g. to [20,8] (design method), [6] (reference architecture) and [9,4,7] (model-driven security).

Compared to other approaches that support a policy language for TM, our primary goals are different for TM in that we intend to apply a model-driven framework to advanced aspects of TM like restricted delegation of rights, information release, access control, trust negotiation strategies etc. In order to enhance inter-operability in open environments like the internet, our model-driven TM framework is built on top of Shibboleth protocols [23] which is one of the most significant TM frameworks that bolsters open standards.

In order to limit the propagation scope of delegation, the authors [11] have proposed a (**logic-based**) Role-based Constrained Delegation Model (RCDM) which uses a *spatial constraint*. The approach has formally analyzed the correctness of spatial constraints and the depth of the delegation tree. The spatial constraints are specified through a policy language called REAL which supports the delegation at two different levels: the delegation of authority and capability based delegation. Compared to our approach, their goal is to formally analyze the whole process of delegation of rights in distributed systems. In [15], the authors have described a Role-based Trust management framework (RT for short) for representing policies and credentials in distributed authorization. The credentials specified in RT are then translated to DataLog rules. In [24], the authors propose a delegation chaining protocol with a trust value. The approach however, leaves the responsibility of the specification of restricted delegation and accompanying issues on the group of peers involved in a delegation scenario. The use of logic programming for the specification of delegation statements provides a powerful mechanism to prove the correctness of the overall system but the specifications are difficult to understand and obscure transcription. Moreover, no logic based mechanism is yet incorporated into a standard authorization framework.

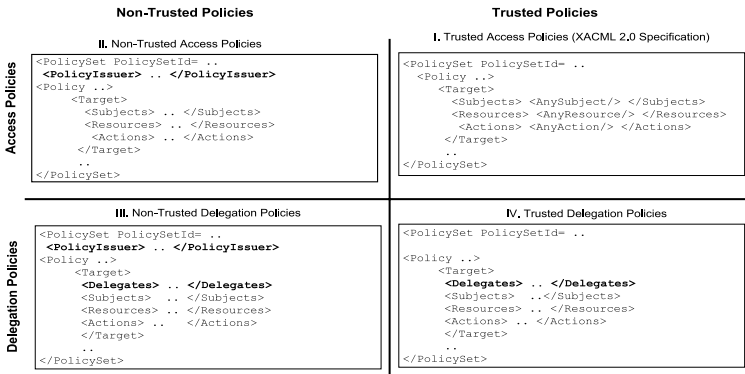
KeyNote [16] is a very well-known trust management language. Its main feature is the support for distributed authorization through delegation policies. These policies (called assertions) contain the description of the authorizer, constraints and licensees (to whom rights are delegated). However, the KeyNote language is more implementation specific and the specification of delegation policies in **proprietary languages** compared to open standards like XACML for rights delegation [26] restricts its usability in open environments.

In [12], the authors present a classical model for **role-based** delegation using sub-role hierarchies. This approach divides a role in to different sub-roles and a senior role can only inherit the allowed sub-roles of the junior role for delegation. Compared to their approach, our work is more focused on combining the RBAC with a modeling framework for restricted delegation based on UML techniques.

There are some **open standards-based** approaches e.g. [14,21] which propose the use of X.509v3 certificates to include the delegation information. However, the delegation information itself is specified in proprietary formats and therefore of restricted usability in open environments like the internet. Moreover, these approaches only provide technical solutions for the specification restricted delegation policies. In [13], the authors extend the Security Assertion Mark up Language (SAML) [22] syntax to achieve different delegation goals. The framework of this approach itself is designed to only transport any kind of delegation information. However, the approach does not discuss restricted delegation. In order to describe more complex, constrain based scenarios in a distributed environment, a policy based payload (like XACML) is needed. Recapitulating this work with XACML policies can offer a flexible delegation framework for distributed and dynamic environments.

The rest of the paper is structured into two parts: in the technology part, section 2 provides an overview of XACML profile for rights delegation and presents an example scenario which will be used in the rest of the paper to illustrate our approach. In the modeling part, section 3 presents a UML profile for restricted delegation of rights. Section 4 deals with the specification of delegation policies using our policy language SECTET-PL and in section 5, a conclusion is drawn.

## 2 Background



**Fig. 1.** XACML Profile for Rights Delegation

XACML [25] is an OASIS standard that allows for the specification of access policies to (web) services. The language provides a standard set of XML elements for the formulation of access control policies. It also specifies a request/response

protocol for related queries and defines an abstract data flow model between functional components. In the following we briefly elaborate the XACML profile (work in progress) for rights delegation [26].

This profile extends the core XACML specification for access policies (Fig 1–part I) by one more type of policy called *delegation policy*. The purpose of this policy is to authorize other policies issued by non-trusted sources e.g. users, systems etc. In addition to `<Subject>`, `<Resource>` and `<Action>` elements (basic XACML concepts), this profile defines an element called `<Delegate>` in the `<Target>` element. The presence of this element qualifies the policy either as a delegation policy (Fig 1–part III&IV) or otherwise as an access policy (Fig 1–part I&II).

The `<Delegate>` element defines a subject (e.g. a role, UserID or a complex object) which is allowed to delegate rights for the situation specified by `<Subject>`, `<Resource>` and `<Action>` elements. This profile also defines a policy-level element called `<PolicyIssuer>`. This element specifies a subject which can issue an access or a delegation policy. The absence of the `<PolicyIssuer>` element qualifies an access or a delegation policy as a trusted policy (Fig 1–part I&IV) and therefore, there is no need for further validation. The presence of the `<PolicyIssuer>` element specifies that the corresponding access or delegation policy is issued by a subject. In this case, the policy needs further validation against a trusted delegation policy (Fig 1–part II&III). The delegation policy files can restrict delegation by means of `<condition>` and `<IndirectDelegatesCondition>` elements of the XACML.

In order to describe our rights delegation framework for distributed systems, we take an example co-operation between our Research Group within the University of Innsbruck (UIBK) and the University of Vienna (UV) for a research project. In our case, research members of the UIBK have given restricted access to the resources located at UIBK [5] which they can further delegate to other research members of the UIBK and to the research members of the UV (focus of this paper). For security, privacy and management reasons, we assume that every peer maintains the attributes of the users associated to his domain. In order to check the access rights of the Service Requester (SR), the Service Provider (SP) clarifies the requester's attributes with a third party – the Attribute Authority (AA) – usually the requester's home domain. The groups use a Collaborative Research Program Management Tool (CRPMT) which offers a web service interface.

In the above scenario, suppose a research member A (where A belongs to UIBK and represents a complex object) has been authorized by means of an XACML trusted delegation policy to execute and delegate a web service `addResource()` hosted at UIBK, her home-domain (cf. Fig 2a). The research member A delegates the access rights on the web service `addResource()` to an external research member B (from UV) and B further delegates to an external researcher C by means of non-trusted XACML delegation policy files (cf. Fig 2a,2b). The external research member C uses her delegated right and accesses the web service `addResource()`. After authentication (we assume that a proper security mechanism to validate

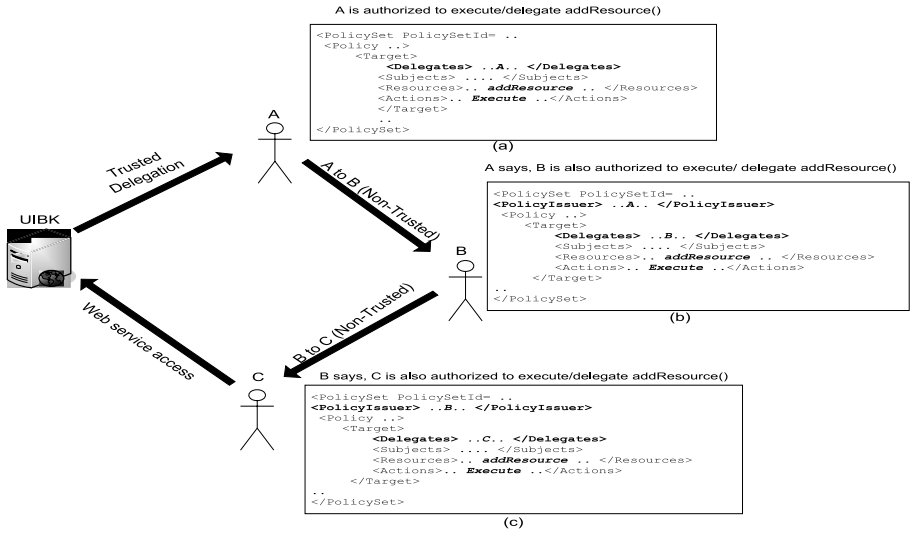


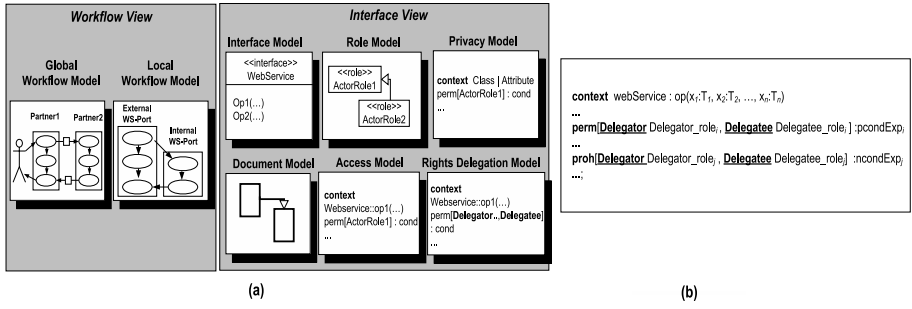
Fig. 2. A Delegation Scenario using XACML

the security requirements like integrity of the delegation policies based on Public Key Infrastructure is in place), the research member C presents a pile of XACML delegation policies as a chain of trust to the authorization component of UIBK to prove his/her eligibility for the web service `addResource()`. These delegation policies include the policy issued by A to B i.e. A→B and B→C. Based on the result of the evaluation of these policies (**chain check**), the research member C will be allowed/disallowed to use the web service `addResource()` [26].

To conclude this section, using existing approaches, the specification of restricted delegation is either limited to a single domain or subject to severe interoperability issues. XACML on the other hand, provides a generic solution, as it can express fairly complex delegation policies and can be deployed across domain boundaries.

### 3 A UML Profile for Restricted Delegation of Rights

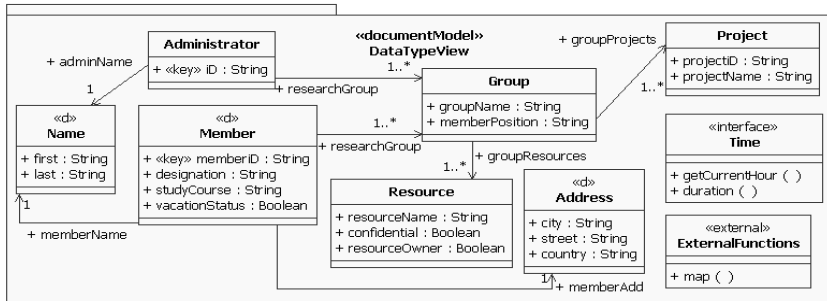
We conceive the example given in section 2 as a workflow within a network of partners cooperating in a controlled way by calling services and exchanging documents. Our approach is based on two orthogonal views: the *Interface View* and the *Workflow View* (cf. Fig 3a). The sub-models in the *Workflow View* depict the message exchange protocol between the cooperating partners with a special focus on security requirements like confidentiality, integrity and non-repudiation [20,7,9]. In the *Interface View*, each partner is conceived as a node offering services with a given data type, access control [5], privacy [3] and rights delegation requirements. In this paper, we concentrate only on the sub-model *Rights Delegation Model* of the Interface View.



**Fig. 3.** (a) Model Views (b) General Form of Rights Delegation Model

**Table 1.** Model Elements of the Interface View Mapped to UML Stereotypes

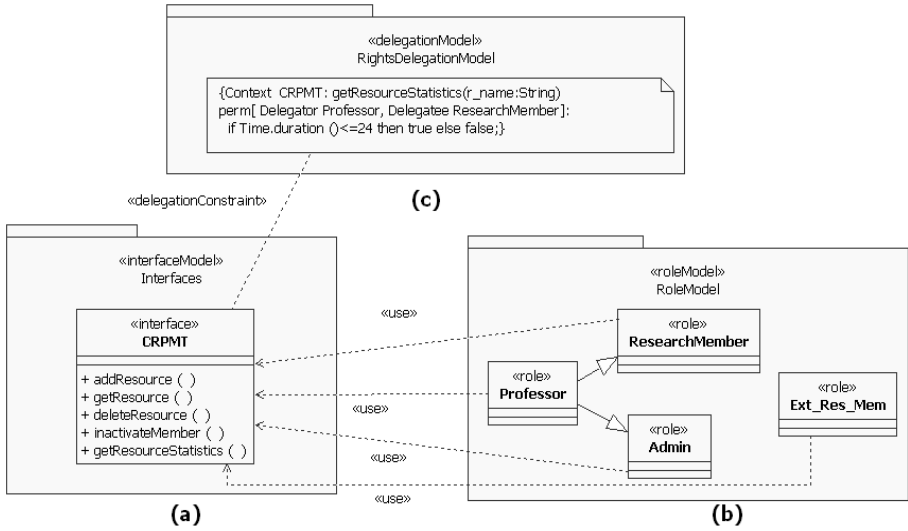
Interface View Model Type	UML Model Element	UML Stereotype
Interface View	Package	<<interfaceView>>
InterfaceModel	Package	<<interfaceModel>>
Interface	Interface	<<interface>>
RoleModel	Package	<<roleModel>>
Role	Class	<<role>>
DocumentModel	Package	<<documentModel>>
UniqueAttribute	Attribute	<<key>>
DistributedEntity	Class	<<d>>
ExternalFunctions	Interface	<<external>>
RightsDelegationModel	Package	<<delegationModel>>
DelegationRules	Constraint	<<delegationConstraint>>
AccessModel	Package	<<accessModel>>
AccessRules	Constraint	<<accessConstraint>>
PrivacyModel	Package	<<privacyModel>>
PrivacyRules	Constraint	<<privacyConstraint>>



**Fig. 4.** Sample Document Model

Table 1 shows the stereotypes used to map the model elements of the Interface View to their representation in UML (cf. Fig 4,5a,5b,5c).

In the Interface View, the sub-model *Document Model* (cf. Fig 4) describes a data type view in the form of UML class diagram and the sub-model *Interface Model* (cf. Fig 5a) defines an abstract set of UML operations, the component offers to its clients. The sub-model *Role Model* (cf. Fig 5b) describes roles having access to the (web) services and is expressed as a UML class diagram. Due to



**Fig. 5.** Sample (a) Interface Model (b) Role Model (c) Rights Delegation Model

limited space and to avoid repetition, please refer to [1] for a detailed information about these supporting models for the Rights Delegation Model.

The sub-model Rights Delegation Model (cf. Fig. 5c) extends each operation definition in the Interface Model with a rights delegation section. The rights delegation section specifies rules and conditions under which a given role (delegator) in the Role Model is permitted (or prohibited) to delegate rights about a particular web service of the Interface Model to another role (delegatee). These rules are described in the predicative language SECTET-PL described in the next section.

## 4 Predicative Specification of Delegation of Rights in SECTET-PL

SECTET-PL [5,10] is a predicative language in OCL-style [19] allowing the specification of fine-grained data dependent access permissions based on roles. Originally developed with the goal of integrating aspects of authorization in use case based development [2], we use SECTET-PL in our Rights Delegation Model to associate each operation *op* of the Interface Model with a set of rules according to the general structure given in Figure 3b.

The positive rule *pcondExp<sub>i</sub>* describes the condition under which a role *Delegator\_role<sub>i</sub>* is permitted to delegate rights to a role *Delegatee\_role<sub>i</sub>*, the negative rule *ncondExp<sub>j</sub>* describes the condition under which a role *Delegator\_role<sub>j</sub>* is prohibited to delegate rights to a role *Delegatee\_role<sub>j</sub>* for a specific operation *op* of the Interface Model. These specifications including the delegator and the delegatee roles, the *op* and the specified rules are then transformed to trusted XACML delegation policy files (transformations are not discussed in this paper).

**DR1:** A ResearchMember role can delegate access rights to add a resource to an administrator provided that both are working in the same research group and the corresponding resource is not confidential.

```
Context CRPMT: addResource (g_name:String,res:Resource)
perm[ Delegator ResearchMember, Delegatee Admin]:
  let rm= delegator.map(Member) , ad = delegatee.map(Administrator) in
  ad.researchGroup.selectOne(groupName=g_name)= rm.researchGroup.selectOne(groupName=g_name)
  and res. confidential = false;
```

**DR2:** Only Research members with designation professor can add a resource with status confidential. The professor can delegate his/her access rights to all Post-doc research members, when the corresponding professor is on vacations.

```
Context CRPMT:addResource (g_name:String,res:Resource)
perm[ Delegator Professor, Delegatee ResearchMember]:
  let prof = delegator.map(Member), rm= delegatee.map(Member) in
  res.confidential = true implies
    (prof.vacationStatus = true and
     rm. designation = "Post-doc" );
```

**DR3:** Only resource owner of a resource can delete a resource. Owners of a resource can delegate their access rights to all research members provided both are working in the same research group.

```
Context CRPMT: deleteResource (g_name:String,res_name:String)
perm[ Delegator ResearchMember, Delegatee ResearchMember]:
  let owner = delegator.map(Member), rm= delegatee.map(Member) in
  owner.researchGroup.selectOne(groupName=g_name)= rm.researchGroup.selectOne(groupName=g_name)
  and
  owner.researchGroup.
  select(groupName=g_name).groupResources.
  select(resourceName = res_name and resourceOwner=true).notEmpty();
```

**DR4:** Only research members with designation professor can inactivate a research member. The professor cannot delegate his/her right to an external researcher if he/she is not a post-doc.

```
Context CRPMT: inactivateMember (mem_ID:String)
proh[ Delegator Professor, Delegatee Ext_Res_Mem]:
  let ext_research_mem = delegatee.map(Member) in
  ext_research_mem.designation <> "Post-doc" ;
```

**Fig. 6.** Sample Delegation Rules

The conditions are permission predicates over the formal parameters of the web service ( $x_1 : T_1, x_2 : T_2, \dots, x_n : T_n$ ). The Document Model is supported by a library of *external functions* e.g. `map(T)` defined in the interface `ExternalFunctions` stereotyped with `<<external>>`. This stereotype indicates that the corresponding interface is not transformed to XML schema but refers to the security infrastructure in order to verify a certain relationship between the caller of the web service and a particular entity of the Document Model. The *identification variables* associated with these external functions differentiates between different types of the caller e.g. the identification variable `delegator` classifies the individual that issues a delegation policy and the identification variable `delegatee` classifies the individual (who) accesses the web service. The example Delegation Rules (DRs) presented in Figure 6 refer to the Document Model, Interface Model and Role Model in Figures 4, 5a, 5b respectively.

Used in some permission or prohibition expressions (cf. Fig 6 – DR1), the special constructs `delegator.map(T)` and `delegatee.map(T)` authenticate the caller of the web service (where the way how authentication is done can be freely chosen), check if the participants are in the specified roles and map the participants to an internal representation (of type T) in the Document Model. In case the participants belong to the some other domain, the `map(T)` function requests the attribute values that are not present locally from the corresponding domain through an attribute requesting service. The `delegator.map(T)` maps the issuer of the delegation policy and the `delegatee.map(T)` maps the caller of the web service to a class in the Document Model. The only difference is the context in which these special variables `delegator` and `delegatee` are used.

RBAC has an inherent limitation in assigning the roles to the subjects according to some limited attributes or attribute values (single dimension) [27]. As more attributes or attribute values are involved, the number of roles grows accordingly. Our approach addresses this problem by checking multiple attribute or attribute values through dynamic constraints using SECTET-PL. There is no need to add additional roles for different attributes or attribute values. (e.g. **PostDocResearchMember** role for all Post-doc research members (cf. DR2)). In general the evaluation strategy of a set of permissions and prohibitions referring to some role **role** and operation **op** is such that all permission conditions  $pcond_1, \dots, pcond_n$  and all prohibition conditions  $ncond_1, \dots, ncond_m$  are connected by a logical "or" leading to the following access condition:

$$pcond_1 \text{ or } \dots \text{ or } pcond_n \text{ and not}(ncond_1 \text{ or } \dots \text{ or } ncond_m)$$

Using SECTET-PL, different roles can correspond to the same entity in the Document Model e.g. **Professor**, **ResearchMember** role to the **Member** class and based on the various attributes of the participants (e.g. their personal data), delegation can be further restricted (cf. DR3). Similarly rights can be delegated among the same roles having different attributes (cf. DR3).

The DR4 describes our approach for restricted delegation in distributed environments. The `delegatee.map(T)` maps an external researcher to a distributed entity **Member** (stereotyped <<d>>) in the Document Model. The external researcher is assigned the role **Ext\_Res\_Mem** after authentication. Attributes of the entity that are not present locally will be requested from the corresponding domain through an attribute requesting service [1].

## 5 Conclusion

In this paper, we presented the restricted delegation aspect of our model-driven TM framework for B2B workflows. The primary goal of our model-driven TM framework is to bridge the gap between the underlying implementation and the domain expert. We provide the specification of dynamic delegation policies via high level language SECTET-PL. One of the most important advantages of SECTET-PL is that it is tightly integrated with the UML models which is a de-facto standard for modeling and it combines the use of a predicative language at a high level of abstraction with an underlying platform independent access policy standard XACML. Currently, we are extending our framework to include advanced parameters of the restricted delegation such as depth of the delegation, the inclusion of a trust value and our tool support [5,10] for restricted delegation.

## References

1. M. Alam, M. Hafner, and R. Breu. Modeling Authorization in a SOA based Distributed Workflow. IASTED Software Engineering 2006, ISBN: 0-88986-572-8.
2. R. Breu and G. Popp. Actor-centric modelling of access rights. FASE 2004. Springer LNCS Vol. 2984, p. 165-179, 2004.



3. M. Alam et al. Model-Driven Privacy Management. Submitted.
4. M. Alam et al. Model Driven Security for Web Services (MDS4WS). INMIC 2004,Digi Obj Id 10.1109/INMIC.2004.1492930.
5. M. Alam et al. Modeling Permissions in a (U/X)ML World. To Appear In ARES 2006.
6. M. Hafner et al. A Security Architecture For Inter-organizational Workflows- Putting WS Security Standards Together. ICEIS 2005,ISBN: 972-8865-19-8.
7. M. Hafner et al. Modeling Inter-organizational Workflow Security in a Peer-to-Peer Environment. IEEE ICWS 2005,ISBN: 0-7695-2409-5.
8. R. Breu et al. Model Based Developement of Access Policies. Submitted.
9. R. Breu et al. Model Driven Security for Inter-Organizational Workflows in e-Government. TCGOV 2005,Proceedings. ISBN 3-540-25016-6.
10. SECTETPL : A Predicative Language for the Specification of Access Rights. <http://qe-informatik.uibk.ac.at/~muhammad/TechnicalReportSECTETPL.pdf>.
11. G. Yin et al. Trust Management with Safe Privilege Propagation. APPT 2005, LNCS 3756,pp.174-183,2005.
12. H. Lee et al. A New Role-Based Delegation Model Using Sub-role Hierarchies. ISCIS 2003,LNCS 2869,pp.811-818,2003.
13. J Wang et al. Extending the SAML to Support Delegation for Web Services and Grid Services. IEEE ICWS 2005,ISBN: 0-7695-2409-5.
14. K. Stoupa et al. XML-Based Revocation and Delegation in a Distributed Environment. EDBT 2004 workshops LNCS 3268, pp. 299-308.
15. N. Li and J. Mitchell. RT: A role-based trust-management framework, 2003. cite-seer.ist.psu.edu/li03rt.html.
16. M. Blaze et al. The KeyNote Trust-Management System. RFC 2704, Sept. 1999.
17. Model Driven Architecture. <http://www.omg.org/mda>.
18. OASIS Organization for the Advancement of Structured Information Standards. [www.oasis-open.org](http://www.oasis-open.org).
19. UML 2.0 OCL Specification. <http://www.omg.org/docs/ptc/03-10-14.pdf>.
20. R. Breu et al. Web service engineering - advancing a new software engineering discipline. ICWE 2005, LNCS 3579 Springer 2005.
21. S. Kim et al. Wokflow-based Authorization. Journal of Grid Computing 2004 Kluwer Publishers, Netharland.
22. SAML 2.0 Specification. [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=security](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security).
23. Shiboleth protocols and profiles: August 2005. <http://shibboleth.internet2.edu/shib-intro.html>.
24. W. Jiang et al. Using Trust for Restricted Delegation in Grid Envoirnmnts. ISPEC 2005, LNCS 3439,pp.293-301,2005.
25. XACML 2.0 Specification Set. [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=xacml](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml)
26. XACML v3.0 administration policy Working Draft 05 December 2005. [http://www.oasis-open.org/committees/documents.php?wg\\_abbrev=xacml](http://www.oasis-open.org/committees/documents.php?wg_abbrev=xacml).
27. E. Yuan and J. Tong. Attributed Based Access Control (ABAC) for Web Services. IEEE ICWS 2005,ISBN 0-7695-2409-5.

# Reputation-Based Trust Systems for P2P Applications: Design Issues and Comparison Framework

Eleni Koutrouli\* and Aphrodite Tsalgatidou

Department of Informatics & Telecommunications,  
National & Kapodistrian University of Athens, Greece  
{ekou, atsalga}@di.uoa.gr

**Abstract.** In Peer-to-Peer (P2P) computing area trust issues have gained focus as a result of the decentralized nature of P2P systems where autonomous peers interact with each other without relying on any central authority. There is, thus, the need of a trust system to ensure a level of robustness against malicious nodes. Various reputation-based trust models have been proposed for P2P systems which use similar concepts but focus on different aspects and address different set of design issues. As a result, there is a clear need to investigate the design aspects of reputation-based trust systems that could be deployed in P2P applications. In this paper we present the basic elements and design issues of such systems and compare representative approaches, aiming at supporting the design of reputation systems suitable for particular P2P applications.

## 1 Introduction

Peer-to-Peer (P2P) systems are decentralized applications where heterogeneous peers, which are autonomous and have intermittent presence in the network and a high level of anonymity interoperate for purposes such as file sharing, distributed computing and eCommerce transactions without the need of a centralized server. The decentralized nature of P2P systems poses the need for enhanced trust between peers that will enable the reliable communication and exchange of services between them.

Peers in P2P systems need to make trust decisions for choosing peers they will transact with or resources they have asked for among the offered ones. There is, thus, the need of at least a minimal trust system to ensure a satisfying level of robustness against various kinds of attacks that have been monitored in P2P systems [8]. Such a trust system should be decentralized so that each peer can make autonomous trust decisions based on other peers' reputation. By "peer reputation" we refer to a measure that indicates the trustworthiness of a peer in a particular context. This measure is estimated based on both direct experiences and other peers' transaction information.

Reputation-based trust models for P2P systems have recently gained a lot of attention by the research community in the areas of trust and P2P systems. Several trust models are found in the literature that use similar concepts (like reputation,

---

\* Eleni Koutrouli is also with the Bank of Greece, Panepistimiou 21, Athens, Greece, email: ekoutrouli@bankofgreece.gr

trustworthiness, recommendation, etc.) but focus on different aspects (like social or probabilistic modeling of behavior, trust data management, etc.). In most cases, these models, although having been simulated and tested, have not been deployed in real P2P applications and, thus, they usually do not fully address all the design aspects that should be taken into account for the design of an effective reputation system that could be deployed in a real P2P application. They also differentiate regarding their approach to the various design issues (such as the kind of input information, the methods of reputation estimation, the reputation representation, etc.).

As a result, no general model for P2P reputation systems exists and the choice of a reputation model for a particular P2P application is challenging. Furthermore, the designer of a decentralized reputation-based trust system needs to take the necessary design issues into consideration and make critical decisions about them.

In this paper we present the concepts which are central to any *reputation-based trust model* for P2P applications and a conceptual representation of such a model. We also present the components and design considerations of a *reputation-based trust system* that could be deployed in a P2P application to provide reputation-based trust functionality. Furthermore, we provide a comparison framework for P2P reputation systems and compare existing approaches. Our objectives are to

- identify the elements of reputation-based trust models for P2P systems and present the major considerations for reputation-based trust systems design
- enable the right choice of either a reputation system or specific elements of a reputation system for particular P2P applications through our comparison.

The rest of the paper is organized as follows: in section 2 we discuss the concept of trust and its applicability in P2P systems and present a conceptual representation of a P2P reputation-based trust model. In section 3 we discuss the components and design considerations of a reputation system, as well as our classification framework, based on these considerations. In section 4, we present representative P2P reputation systems for various application areas and use our framework to compare them according to the presented issues. Discussion and conclusions follow in section 5.

## 2 Trust, Reputation and P2P Systems

Trust in computer science is a concept that has been borrowed from the human society, where people constantly apply it in their interactions. In the World Wide Web, where interactions in widely-distributed, open, decentralized systems that span multiple administrative domains are enabled, the need for establishing trust between interacting entities is posed. As a result, recent research focuses on trust management as a framework for decentralized security decisions in such systems.

Trust is a complex, multifaceted, and context-dependent notion, which is representatively defined in Sloman [15] as “the quantified belief by a trustor with respect to the competence, honesty, security, and dependability of a trustee within a specific context”. In P2P systems, peers which do not know each other need to exchange services and resources without central control. There is, thus, the need for a decentralized trust system that will support peers to identify reliable peers [1][13][14], reliable resources [14], or malicious peers [11]. Various trust systems for P2P applications have been proposed that can be classified in the following categories:

- **Policy-based trust systems**, where peers use credential verification to enable access control to restricted resources [12]
- **Reputation-based trust systems**, which use information considering previous interactions with an entity to establish a reputation measure that will support a trust decision [1][5][11][14][16][18][19]. In our paper we examine this category of trust systems, due to their wide applicability in P2P systems.

## 2.1 Conceptual Representation of P2P Reputation-Based Trust Systems

The basic elements of a reputation-based trust model are the following:

1. **Trustee:** The entity that is given a reputation value for a service it provides, e.g. the output of a transaction, or an attribute it possesses, e.g. the genuity of a file.
2. **Trustor:** The peer that needs to evaluate the trustee's reputation in order to make a trust decision about it, such as to decide whether to perform a transaction with it.
3. **Third Party or Witness:** A peer that provides a recommendation for the trustee based on its own experiences with the latter.
4. **Context:** The reputation of a peer depends on the specific context in which it applies, like a specific service the trustee provides, attributes of such a service, etc.
5. **Recommendation:** Refers to the feedback provided by peers about another peer's trustworthiness. In the following we will alternatively use the terms *recommendation*, *trust information*, or *feedback* to refer to this kind of information.
6. **Trustworthiness or reputation:** An indicator of the quality of the trustee's services or attributes, based on recommendations, as well as the specific context and time.

Figure 1 provides a graphical representation of the elements of a reputation-based trust model, based on the conceptual models presented in [4].

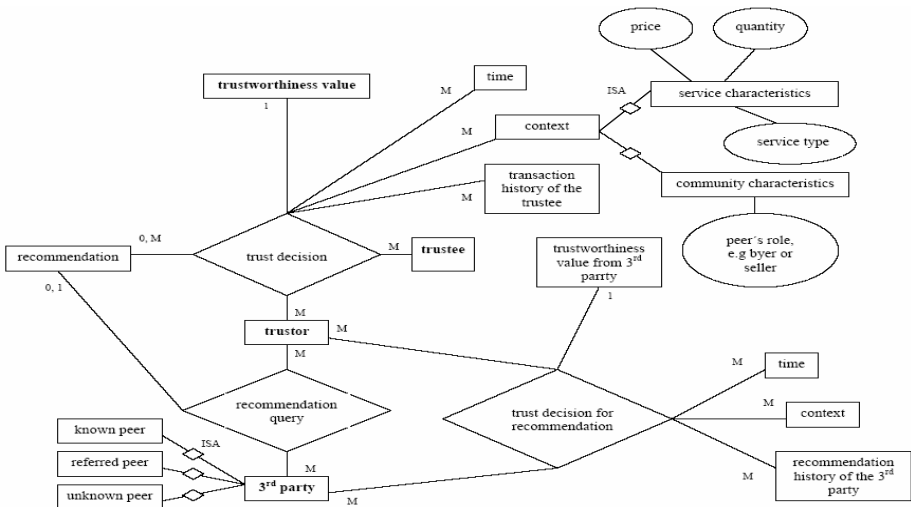


Fig. 1. Conceptual model of a reputation-based trust system

### 3 Design of P2P Reputation Systems and Comparison Framework

In general, a reputation-based trust system assists peers in choosing a reliable peer to transact with. To provide this function, a P2P reputation system:

- collects information on the transactional behavior of each peer. Transacting entities produce ratings about each other's performance, which are often locally aggregated to form an entity's opinion about others. Individual ratings or opinions constitute recommendations, which are distributed in the P2P network. Each peer can store such information and can provide it on request or by propagating it in the network.
- aggregates the trust information that concerns the transactional behavior of the trustee and produces a trustworthiness (or reputation) value for it. As it is often impossible or too costly to obtain ratings or opinions resulting from all interactions with a given peer, a reputation score is based on a subset of ratings.
- ranks peers according to their trustworthiness or compares a peer's trustworthiness with a threshold in order to allow the trustor to choose a peer to transact with and the system to take action against malicious peers while rewarding contributors.

The functionality of reputation system can, thus, be broken down into the components illustrated in Figure 2. When designing each component, various design issues should be taken into consideration, which we present in the following. These design issues constitute our *comparison framework*, which has been used for evaluating a number of P2P reputation systems that are presented in Section 4. The result of this comparison is illustrated in Table 1.

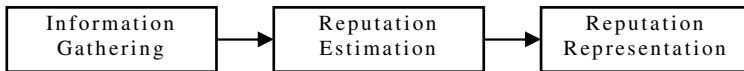


Fig. 2. Components of a P2P reputation-based trust system

#### 3.1 Design Considerations for Information Gathering

1. **Trust information storage, dissemination and search mechanisms:** An important issue in a decentralized trust system is data management, which refers to which trust information is used, where it is stored and how it is propagated and acquired. Some P2P reputation systems [1] use the underlying P2P structure to store and retrieve trust information. In others [14][19] each peer keeps information regarding both its own transactions and a set of neighbors, which it can ask for recommendations. Broadcasting, flooding and probabilistic flooding methods can be used to send queries for recommendations or disseminate own experiences.
2. **Local control over trust information stored locally on a peer:** Whether peers have or do not have local control over the trust information that is stored locally on them, has impact on the reliability of the reputation system, as a malicious peer with local control could change the information stored locally on it.

Table 1. Comparison of reputation-based trust systems for P2P systems

Design Issues		P2P Ecommerce applications					File Sharing		Cooperation	
	Regret	Social mechanism for reputation	Managing the dynamic nature of trust	Peer Trust	Fuzzy Trust	Managing Trust	Maximum Likelihood Estimation	Reputation-based trust management	NICE	
Information Gathering	Storage of trust information	each peer stores information about the social relationships in its environment and also about its transactions	each peer stores an interest vector, an expertise vector and information about its neighbours' expertise and reputation	each peer stores a small portion of the trust data (transaction history and feedbacks) using the P-Grid Structure	each peer maintains transaction records and remote peers' evaluated trust scores, DHT-based overlay	P-Grid Structure	structured P2P network based	each peer stores a trust vector for each peer it has dealt with at the past	each peer stores positive reputation information concerning itself or negative reputation information concerning other peers	
	Feedback dissemination and search mechanisms	fuzzy logic based mechanisms for witnesses identifying	recommendations are answers to queries for services	N/A	DHT-based overlay network	P-Grid based	structured P2P network based	first generation P2P systems based, e.g. Gnutella	probabilistic flooding-based search	
	Local control	yes	yes	yes	yes	no	no	yes	yes	
	Recommender's credibility	yes, based on <ul style="list-style-type: none"><li>social relationships of witness and trustee</li><li>witness's trustworthiness regarding its services</li></ul>	yes (a witness's recommendation is taken into account if the witness's trust rating is above a threshold)	yes (as a function of the witness's trustworthiness or using a personalized similarity measure)	yes (trust score of a peer is taken into account as a weight for global reputation estimation)	indirectly (it is very likely that a peer providing a lot of complaints about others lies)	yes (possibility of a peer to lie when providing a recommendation)	yes (different trust and credibility scores)	no can be incorporated	
	Positive/negative behavior	positive and negative	positive and negative	positive and negative	positive and negative	positive and negative	negative	positive and negative	positive and negative	
Method of Feedback Aggregation	Context dependence	yes	no	yes (transaction context and community context)	yes (local transaction parameters)	no (could be integrated)	N/A	no	no	
	Initialization of trust information:	default value based on peer's role	initial trust rating = 0	default value = 1	N/A	at the beginning every peer is trusted	N/A	N/A	depends on the chosen algorithm	
	Method of feedback aggregation	statistic	Statistic	statistic	statistic, fuzzy logic for the estimation of weights	probabilistic	probabilistic	statistic	statistic	
	Transitivity level of trust induction	one level of induction	recommendation chain	one level of induction	one level of induction	one level of induction	one level of induction	one level of induction	recommendation chain	
	Recency dependence	yes, a time function is used as a weight	yes (each time trust rating is updated based on the previous rating)	yes (weights give more importance to recent experiences)	yes (can be incorporated in the context)	yes (transaction date is taken into account. Recent transactions lead to higher weights)	no	N/A	yes no can be incorporated	
Output	Scope of trust: global vs. localized	localized	localized	global	localized	global	global	localized	localized	
	Threshold/Rank	threshold	threshold	threshold	N/A	rank	threshold	rank	threshold	
	Range of trust values	[-1,1]	[-1,+1]	[0,1]	N/A	no specific range, the higher price means less trust	[0,1]	trust ratings and credibility ratings: (0,1)	depends on the chosen scheme	
	Distrust Representation	N/A	yes, if a peer's trust rating is below a threshold it will not be trusted	No	N/A	yes	N/A	yes (distrust ratings)	no can be incorporated	

3. **Credibility of the recommender:** As peers may provide inaccurate recommendations, the recommender's credibility should be taken into account. Some systems suggest maintaining separate ratings on a peer's likelihood to defect on a transaction and its likelihood to recommend malicious peers. Some others use the trustworthiness value of a peer with respect of the services it provides as a filter for the recommendations it makes, assuming that a peer, which provides trusted services will also provide honest recommendations.
4. **Type of behavior taken into account:** Reputation evaluation can be based only on positive behavior (e.g. contribution rate in a file sharing system), or only on negative behavior (e.g. cheating in a transaction), or both on positive and negative behavior of the trustee. In the last case negative behavior should normally be taken into account with a higher weight, as it is has a greater impact on the trustworthiness of a peer than its positive behavior.
5. **Context dependency:** Reputation estimation and trust decisions depend on the particular context of a transaction. This context can constitute of transaction related factors, such as quantity and price of a transaction in the case of e-commerce applications or can refer to the type of the transaction, such as service provision or recommendation provision. Some trust systems (e.g. [13][18]) take context into account explicitly or implicitly while some others (e.g. [14]) ignore context, assuming that it is the same for all transactions.

### 3.2 Design Considerations for Reputation Estimation

1. **Initialization of trust information:** Assigning an initial value to a peer, when no information about its transactions exists, is challenging, as it is important to distinguish between a new peer and a peer with poor long term performance and also prevent peers with poor trustworthiness to enter into the system with a new identity in order to gain higher trustworthiness. The choice of an initial trust value depends on the strategy followed: it can represent complete distrust, complete trust, neutral trust, or default values depended on the role of the peer in a community.
2. **Scope of trust information (global vs. localized information):** Some reputation systems ([1][18]) assume that every peer has the same access to existing trust information and, thus, when different peers evaluate the trustworthiness of another peer their evaluation will be the same. In these systems, trust has a global scope and can be said to be objective. In other systems ([13][19]), trust evaluation is a localized process, based on direct information and on information coming from a set of trustor's neighbors. In localized trust systems trustworthiness is subjective.
3. **Trustworthiness estimation method:** Feedback regarding past interactions with the trustee is aggregated to produce the trustee's trustworthiness value. Various aggregation methods have been proposed such as simple statistic functions (e.g. average), probabilistic methods, fuzzy logic, etc.
4. **Transitivity extent:** Trust transitivity is implicitly taken into account in reputation-based trust systems, as they assume that if A trusts B and B trusts C and B recommends C to A then A can estimate a trustworthiness metric for C based on B's recommendation and A's trust in B. Transitivity is assumed either through a recommendation chain (multiple levels of trust indirection) or only through one level of trust indirection.

5. **Recency dependency:** While estimating reputation more recent transaction behavior should have a greater impact on a peer's score than older transactions, e.g. weights or aging factors can be used to give more importance to recent experience.

### 3.3 Design Considerations for Trustworthiness Representation

1. **Range of trustworthiness values:** Trustworthiness values can be discrete or continuous and can have a varying range reflecting different trust semantics. Examples of such domains are the interval  $[0,1]$  (e.g. when a value represents a probability) and the set  $\{0,1\}$  where 0 represents distrust and 1. Another example is the use of a specific interval which is divided in smaller intervals to represent different levels of trust [7].
2. **Rank or threshold based:** Trust decisions in reputation systems are usually taken either after comparing a peer's trustworthiness with a threshold (threshold-based systems) or after comparing the trustworthiness of different peers (rank-based systems). When a threshold is used, its selection depends on the semantics of trustworthiness values and the requirements of the specific implementation.
3. **Distrust representation:** Representation of distrust can isolate malicious peers. In some trust systems distrust is explicitly represented, either as a specific range of reputation values or by keeping different ratings of trust and distrust.

## 4 Comparison of P2P Reputation Systems

We have used the comparison framework described in Section 3 to examine and compare a number of reputation systems, with respect to the way they address the identified issues and, thus, identify existing approaches, deficiencies and possible design choices. Most existing P2P reputation systems have been developed for general-purpose P2P applications, such as file sharing. However, reputation systems have been proposed for other classes of P2P systems too, such as cooperative and P2P e-commerce applications. We have selected systems which we believe that are representative works on reputation in the aforementioned P2P application areas, although further approaches exist (e.g.[10]), which are not presented here due to space limits. Finally, we present our observations in Table 1, which illustrates both our framework and the choices of the various reputation-based trust systems regarding the identified design considerations.

In the following, we briefly describe the selected reputation-based trust systems:

- **Regret [13]:** Regret is a reputation system, designed for multiagent marketplaces, which is based on the social relations between peers. It concerns three different dimensions of reputation: *individual dimension* considers only the direct interactions between peers, *social dimension* considers information about the trustee coming from other peers and from the social relations between peers, and *ontological (or context dependent) dimension* refers to combining reputations on different aspects. Various kinds of reputation along with their reliability measures are estimated and then combined to form the final reputation of a peer.



- **A Social Mechanism for Reputation [19]:** In this system peers can have two kinds of reputations: for providing services and for providing recommendations. Peer A assigns a rating to B based on its direct experience with B as well as recommendations, and A's ratings to recommenders. A peer receiving a query decides whether it has the expertise to answer or not and forwards the query to a set of neighbouring peers. A response can contain an answer, or a recommendation, or both, or neither and can be used to evaluate the expertise of the responding peer and of its recommenders.
- **Managing the Dynamic Nature of Trust [7]:** In this system when a peer wants to make a trust decision about another peer within the current time slot, it uses its local rating if it has interacted with the trustee in the same time slot. Otherwise, it asks for reputation information and estimates the trustee's trustworthiness as an average of the received reputation values, weighted with the witnesses' trustworthiness. The trustworthiness of a peer for a future time slot can also be estimated probabilistically. After an interaction the trustor modifies both the trustworthiness values of the trustee and those of the witnesses.
- **PeerTrust [18]:** PeerTrust is designed for P2P eCommerce communities. It takes into account the feedback in terms of the amount of satisfaction a peer obtains through transactions, the number of the trustee's transactions, the credibility of the feedback, the transaction context factor, addressing transaction characteristics, and the community context factor, referring to community characteristics (such as the availability of pre-trusted peers). Each peer stores a small portion of the trust data using the P-Grid structure [2]. A peer collects the necessary trust data and evaluates the trustworthiness of another peer on the fly when needed.
- **FuzzyTrust [16]:** A fuzzy logic reputation system for P2P eCommerce applications. Peers perform fuzzy inference on local parameters to generate local scores for the peers with whom they have transacted. These local scores are collected from qualified peers, which meet an aggregation threshold and aggregated into global reputation values. The FuzzyTrust system uses a DHT-based P2P overlay network for the global reputation aggregation.
- **Managing Trust [1]:** This system is based on binary trust. If a peer cheats in a transaction, it becomes untrustworthy and a complaint is formed against it and stored and replicated in a P-Grid data structure [2]. When a peer wants to calculate the trustworthiness of another peer, it searches for complaints that this peer has both received and filed. Trustworthiness of a peer is then evaluated based on the global knowledge on these complaints.
- **Maximum Likelihood Estimation based trust system (MLE) [5]:** The proposed system uses a structured P2P overlay for the storage and retrieval of trust information, which consists of reports on a peer's performance. These reports are either 0 if the peer acted dishonestly or 1 if the peer acted honestly. The reputation of a peer is its probability to perform honestly in its transactions with others. This is estimated based on a probabilistic method, taking into account the probability of a peer to lie when it reports on another's peer's performance.
- **A Reputation-based Trust Management System for P2P systems [14]:** In this system, which is designed for P2P file sharing, every peer can estimate trust and distrust ratings for other peers, based on binary trust vectors. A peer receiving responses for a resource query organizes them into groups according to their file

hash value. A trust score for each file version is calculated as the average of the trust ratings of the offering peers. If there is not enough local information about a peer, trust queries are issued about it. Credibility ratings of the responses are used as weights for the trust rating estimation of the recommended peer. The file version with the highest trustworthiness is downloaded from one of the peers who offer it.

- **NICE [11]:** NICE is designed for Internet cooperative applications. After a transaction, each transacting peer signs its opinion regarding the quality of the transaction. If this signed opinion (cookie) is positive it is stored in the other transacting peer, otherwise, it is stored in the peer signing it. When a peer A wants to access B's resources, it has to prove its trustworthiness to B and, thus, sends B cookies signed by B. If A does not have such cookies, it may collect chains of cookies from B to A and present them to A.

Table 1 contains a comparison of the aforementioned reputation systems against the design considerations issues identified in section 3. When no information is given in the description of a system regarding one of these issues, the respective cell of the table has been filled with "N/A", whereas in some cells additional explanative information is provided. This table can also be viewed as a multifaceted classification of reputation systems as well as a supporting tool for the designing procedure.

## 5 Discussion

Reputation plays a vital role in the process of establishing trust between communicating peers in a P2P system. Motivated by the lack of a complete design framework for reputation systems for P2P applications, we have presented the elements of reputation-based trust systems and the basic issues that need to be taken into account in the design of reputation systems that can be used in P2P applications. We have also examined some representative approaches for three P2P application areas and have compared them against the way they deal with these issues.

The designer of a reputation-based trust system for P2P systems should consider the presented design issues and make careful decisions about them in order to develop an effective solution for reputation functionality in such systems. Our presented comparison aims at supporting the right choices regarding these issues when designing a reputation system for a particular P2P application.

However, there are further issues regarding the design of a reputation-based trust system for P2P applications that need to be addressed, such as handling of anonymity, supporting fault tolerance and scalability and various types of misbehavior and attacks that can affect a reputation system's reliability. We are aiming at examining these issues in future work in order to provide a more comprehensive framework for the design of effective reputation systems for P2P applications.

## Acknowledgements

This work has been partially funded by the European Commission under contract 04559 SODIUM and by ELKE under contract 70/4/5829.

## References

- [1] Aberer, K., Despotovic, Z., Managing trust in a peer-2-peer information system, 10th Intl Conference on Information and Knowledge Management (CIKM), Atlanta, 2001
- [2] Aberer, K., P-Grid: A self-organizing access structure for P2P information systems, 6th Intl Conference on Cooperative Information Systems (CoopIS), 2001
- [3] Chang, E., Dillon, T., Hussain, F. K., Trust and Reputation Relationships in Service-Oriented Environments, 3rd Intl Conference on Information Technology and Applications
- [4] Chang, E., Hussain, F. K., Trust and Reputation Relationships in Service-Oriented Environments, Keynote, ICITA 2005
- [5] Despotovic, Z., Aberer, K., Maximum Likelihood Estimation of Peers' Performance in P2P Networks, 2nd Workshop on the Economics of Peer-to-Peer Systems, 2004
- [6] Despotovic, Z., Aberer, K., Possibilities for Managing Trust in P2P Networks, Technical Reports in Computer and Communication Sciences (EPFL Technical Report IC/2004/84), November 2004
- [7] Dillon, T.S., Chang, E., Hussain, F.K., Managing the dynamic nature of trust, IEEE Transaction of Intelligent Systems, vol. 19, no. 5, pp. 79-82, Sept/Oct 2004
- [8] Jøsang, A., Ismail, R., Boyd, C., A Survey of Trust and Reputation Systems for Online Service Provision (to appear), Decision Support Systems, 2005
- [9] Jurca, R., Faltings, B., An Incentive Compatible Reputation Mechanism, IEEE Conference on E-Commerce, Newport Beach, CA, USA, 2003
- [10] Kamvar, S., Schlosser, M., Garcia-Molina, H., The Eigentrust Algorithm For Reputation Management in P2P Networks, 12th Intl World Wide Web Conference, 2003
- [11] Lee, S., Sherwood, R., Bhattacharjee, B., Cooperative Peer Groups in NICE, 22<sup>nd</sup> Conference of the IEEE Computer and Communications Societies (INFOCOM), 2003
- [12] Li, N., Mitchell, J., RT: A Role-based Trust-management Framework, 3<sup>rd</sup> DARPA Information Survivability Conference and Exposition (DISCEX), Washington, 2003
- [13] Sabater, J., Sierra, C., Reputation and social network analysis in multi-agent systems, 1st Intl Joint Conference on Autonomous Agents and MultiAgent Systems, Bologna, 2002
- [14] Selcuk, E. Uzun, and M. R. Pariente, A Reputation-Based Trust Management System for P2P Networks, 4th Intl Workshop on Global and Peer-to-Peer Computing (GP2PC), 2004
- [15] Sloman, M., Trust-management in Internet and pervasive systems, IEEE Intelligent Systems, vol. 19, no. 5, pp. 77-79, Sept/Oct 2004
- [16] Song, S., Hwang, K., Zhou, R., Kwok, Y. K., Trusted P2P Transactions with Fuzzy Reputation Aggregation, IEEE Internet Computing Magazine, Special Issue on Security for P2P and Ad Hoc Networks, Nov/Dec 2005
- [17] Suryanarayana, G., Taylor, R., A Survey of Trust Management and Resource Discovery Technologies in Peer-to-Peer Applications, ISR Technical Report # UCI-ISR-04-6, 2004
- [18] Xiong, L., Liu, L., A Reputation-Based Trust Model for Peer-to-Peer eCommerce Communities, IEEE International Conference on E-Commerce (CEC), 2003
- [19] Yu, B., Singh, M. P., A social mechanism of reputation management in electronic communities, 4<sup>th</sup> Intl Workshop on Cooperative Information Agents, 2000

# Towards Trust in Digital Rights Management Systems

Jürgen Nützel and Anja Beyer

Technische Universität Ilmenau, Institut für Medien und Kommunikationswissenschaft  
D-98693 Ilmenau, Germany  
{Juergen.Nuetzel, Anja.Beyer}@tu-ilmenau.de

**Abstract.** Digital transactions are usually based on mutual trust. In case of DRM (Digital Rights Management) this initial trust is missing on both sides. Neither do the content providers trust their clients – therefore DRM was established. Nor do the clients trust the content providers and react with not using these systems. The release of an open DRM standard by the Open Mobile Alliance (OMA) was a first step to increase the trustworthiness of DRM. But from the content providers' perspective a secure implementation for PC Platforms was missing. Especially the mechanisms to obfuscate and install the device private key which is the security anchor were not established there. This paper shows a software solution for that. A more riskless way to solve this problem is the involvement of Trusted Computing which is also shown by the authors. Finally the authors claim the necessity not to leave the users' security behind.

## 1 Motivation and Introduction

Digital Rights Management Systems (DRMS) were developed to enforce the rights of the content providers but the ultimate success is still missing [1]. The reason for it is obvious. Singh et al [2] accurately expressed it: “DRM policies were driven by lack of trust and treated everyone like criminals”. According to [3] successful digital transactions always depend on security, trust and benefit. In order to increase the acceptance of the DRMS there definitely have to be improvements in these areas. As commercial operations always are multilateral [comp. 4], these changes have to be made on both consumers and providers side. Therefore we introduce in chapter 2 an open DRM standard which was released by the Open Mobile Alliance (OMA) [5]. In chapter 3 we present our proposal for a mechanism which uses obfuscation techniques to protect the device private key which is the security anchor of the OMA DRM agent. Another attempt to build a trustworthy and secure OMA DRM agent is Trusted Computing which is discussed in chapter 4. Afterwards we argue how these mechanisms have the ability to improve the trustworthiness of the providers. We also give recommendations where there have to be further improvements.

## 2 Digital Rights Management of OMA

The Open Mobile Alliance (OMA) [5] is an organization which develops open standards to increase the interoperability of mobile services. Nearly all mobile operators and

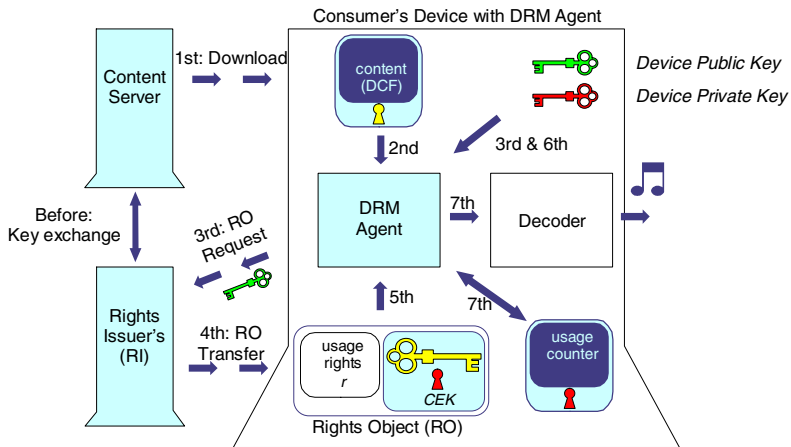
device manufacturers are members of OMA. One of OMA's standardization activities focuses Digital Rights Management (DRM). The authors are of the opinion that OMA will become a leading role in the DRM business because all the leading mobile operators and mobile hardware manufactures support this standard.

The main goal of any DRM [6] solution is the enforcement of permissions and constraints associated with the content. The main threat comes from unauthorized access to protected content beyond the grants of the associated rights objects.

OMA DRM V1.0 only provides three simple protection schemes: forward-lock, combined delivery and separate delivery. See [7] for more information. It becomes obvious that these simple protection schemes do not fulfill the requirements of a second generation DRMS. Therefore OMA developed a second release.

## 2.1 The DRM Reference Model and OMA DRM V2.0

The first release of the DRM specification lacks the complete security necessary for a robust, end-to-end DRMS that enables a secure distribution, the authentication of devices, revocation and other aspects like a domain concept [9]. V2.0 which is the focus of this paper addresses these missing aspects. In [10] a DRM reference model was introduced which well describes the fundamental structure and functions of most of the OMA DRM V2.0. The three major components are the content server, the rights issuer (RI) and the DRM agent. According to this reference model (see fig. 1) the download and usage of content files is proceeded as follows.



**Fig. 1.** DRM reference model with usage counter and device key pair for device identification

Before a download can start the content server has to prepare the encrypted content. In the same manner as every modern DRMS, the content is packaged in a secure container (DCF). The content is encrypted with a symmetric content encryption key (CEK). The content server adds additional data like unique content ID and the address of the RI to the content package and hands the applied CEKs (or a seed information to retrieve the keys from) over to the RI. The RI stores the CEKs

and provides them on request (3<sup>rd</sup> step in fig. 1) together with the appropriate usage rights in form of a rights object (RO). This is an XML document, expressing the permissions and constraints (using ODRL 2.1 [8]) associated with the content and also contains the CEK. Therefore the content cannot be used without the appropriate RO.

All DRM agents have a unique device key pair and a certificate. The certificate (not shown in fig. 1) includes information such as issuer, device type, software version, serial numbers, etc. This allows the RI to securely authenticate a user's device. The certificate with the device public key is transferred during the rights object acquisition protocol (ROAP, 3<sup>rd</sup> and 4<sup>th</sup> step) from the DRM agent to the RI.

In [10, p. 82] the DRM agent is described as "...the real nerve center of the DRM system." It enables the user to exercise his rights, to render the content and it organizes the communication with the content and the RI [10, p. 79ff].

Figure 1 also shows a typical sequence for a DRM system: In the first step the user receives a content package either by downloading it from a content server or from another user (superdistribution). In order to render it, the DRM agent needs an appropriate RO. If no local stored RO was found the DRM agent sends a RO request to the RI (3<sup>rd</sup> step). The request includes the identity of the device (by sending the device certificate with the device public key) and the content ID from the content package (DCF). Before the forth step might happen a financial transaction is initiated. Afterwards, the RI creates the RO containing usage rights and CEK.

Before delivering the RO in step 4 to the client device, sensitive parts like the CEK are additionally encrypted using the symmetric REK (right object encryption key). The RO is cryptographically bound to the target DRM agent. This is done by using the device public key which encrypts the REK. This ensures that only the target DRM agent with corresponding device private key can access the RO and thus the content.

The DRM agent is able (in step 5 und 6) to access the CEK using the device private key (and the REK). Depending on the usage counters and the usage rights ("play three times") the content will be decrypted and rendered in the decoder.

Like the CEK the device private key may not leave the trusted environment of the DRM agent due to being it's security anchor. If it becomes disclosed, the user will be able to decrypt every content package without a proper RO. These are the most important aspects of the OMA DRM security model. Further aspects refer to state of the RO (e.g. remaining number of play-back or usage time) and the time on the user's device which may not be modified by the user. An unauthorized modification of the play-back counters or the device time has to be prevented as well.

In chapter 3 we introduce our approach to obfuscate this key on open/insecure platforms like Linux and Windows PC's. In chapter 4 we show how Trusted Computing makes risky software obfuscation obsolete. This will enable a trustworthy OMA DRM even on complete open platforms like Linux.

### 3 Software Implementation of OMA DRM Agents

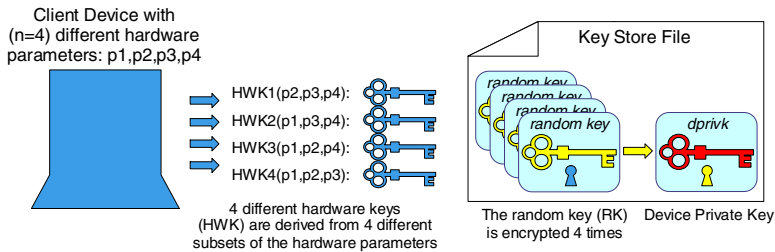
An attack on the CEK could be possible if the REK in the ROs could be deciphered if the device private key becomes obvious, therefore the device private key has to be kept secret in the trusted environment of the DRM agent.

A second possible threat is the loss of the device authentication. The complete security of the communication between the RI and the DRM agent relies on the device key pair. If a private key gets “lost” an attacker could implement (according to the open OMA protocols) a corrupt DRM agent which simply decrypts the DCF instead of rendering it. The RI can react (if it is detected) on this threat only by the revocation of this corrupted device (identified by the device certificate) or even by the revocation of all devices of this device type. After the revocation the revoked device is no longer able to receive valid ROs. The developer of the agent’s software gets seriously into trouble with the revocation of DRM agents [11].

### 3.1 Obfuscation of the Key Store

Mobile devices are the main focus of OMA DRM. Nevertheless, there is a great interest in the industry to port the DRM agent also to other platforms, especially to Windows XP. The problem is that the Windows XP operating system does not support trusted storage facilities to hide the device private key.

In chapter 4 we introduce an OMA DRM implementation based upon the efforts of the TCG (Trusted Computing Group) [12]. But currently Trusted Computing even with Windows Vista [13] comes very slowly onto Windows platforms. Therefore we have to think about less secure solutions based upon software obfuscation techniques. Such obfuscated software solutions have to solve two goals:



**Fig. 2.** Hiding and device binding of the device private key using hardware parameters

**Hiding and device binding:** The device private key must not become visible and must not be transferred to any other device (PC). The following procedure is our proposal for a solution to this problem. Our approach is to store the private key encrypted in a key store file (see figure 2). The encryption is done by a randomly generated symmetric key (RK). RK will be stored also in the key store. The RK will be encrypted by symmetric keys (the hardware keys, HWK), which will be derived directly from several hardware (or system) parameters like MAC address, hard disc and graphic card IDs and others. To avoid the loss of the key store after the replacement of the hard disk, the RK has to be encrypted several times with different subsets of the system parameters (three from four parameters in figure 2). The state of the rights objects (RO) is stored using the same method. In the final implementation we have 8 different parameters and we allow the loss of two of them. In this case RK has to be encrypted more than 50 times with different subsets of the 8 parameters.

**Obfuscation of the software:** A way to disclose the private key even if it is protected as shown in figure 2 is to reverse the software. In [13] and [14] the reader may learn more about such reversing techniques and about techniques to prevent reverse engineering of program code. If an attacker is able to trace the software using a debugger after a while she will find the location where the private key will be applied. This allows her to locate the key in the memory. To avoid this attack several obfuscation techniques should be applied. We mention only a few of them here [11].

Another (very difficult) method is to modify parts of the operating system. This was done by Sony (see chapter 5.1) with big drawbacks for the user's security.

### 3.2 Different Options to Install the Device Private Key

In case of Windows XP the user has to install additional software to receive rights objects from an OMA compatible RI. This software has to manage also the device private key. This includes also the installation of the key.

- **Embedded in installation package:** One option would be to embed (in an obfuscated way) the private key in an individualized installation package. This solution is insufficient because the individual installation package could be installed on several PCs.
- **Created in the DRM agent:** Another option is to create the device key pair within the DRM agent. This option has many drawbacks and security risks. The DRM agent has to send the device public key to a certification authority (CA) to receive a signed certificate (for the authentication against the RI). This communication has to be secured to make the CA believe the public key comes from a valid DRM agent. In chapter 4 we show that a Trusted Platform Module is able to solve this.
- **Hidden download:** The authors of this paper are of the opinion that the hidden download of a unique device private key is the most practicable software solution. The security anchors for this download are two shared master keys (*mk1*, *mk2*). The keys are embedded in the installation package (which is the same for all users) for the DRM agent. To obfuscate the two 128 bit AES keys they will be split into smaller parts and will be spread over many kilobytes of random data.

Fig. 3 shows the proposed communication between the client PC and the "Device CA". At the beginning the installation package with a specific software ID (*sw\_id*) will be downloaded and installed. After DRM agent's first start it contacts the Device CA which has to be trusted by the RI. The agent sends the software ID, a random session ID (*session1*), the local time and a first message authentication code (*mac1*). The server uses *mac1* to proof that the transferred parameters are unmodified and that the sender is a valid DRM agent. The correct local time has to be sent to prevent reply attacks. If the local time is within a defined tolerance the server answers with the server time (*time2*) and a second MAC (*mac2*), which enables the client to proof the authentication of the server. The server time could be used to adjust the local clock.



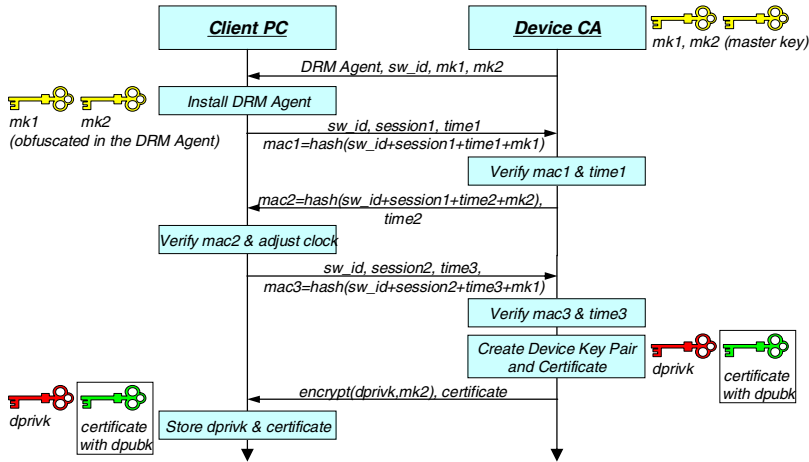


Fig. 3. The proprietary communication between client PC and the Device CA

After this mutual authentication the client requests the device certificate and the device private key. This second request uses a new session ID (*session2*). If the sent parameters could be verified the server creates a new key pair. Every request produces a new key pair. Finally the certificate and the encrypted device private key are transferred. The client PC receives the encrypted device private key. It decrypts the key and re-encrypts it using the RK (fig. 2). After this step the key store is initialized and the DRM agent is ready to request RO [11].

## 4 OMA DRM and Trusted Computing

The TCG (Trusted Computing Group) [12] is an organization which develops and promotes open, vendor-neutral industry standard specifications for trusted computing building blocks and software interfaces. The TPM (Trusted Platform Module), which comes into play in the new Windows Vista [15], is one of the developments of the TCG. The TPM is a co-processor on the PC's main board which provides cryptographic operations (like RSA, AES, SHA-1, random numbers generator) and stores (shielded) individual private keys generated by the TPM. It is able to provide a root of trust. Windows Vista uses the current TPM specification (V1.2) [16] [17].

### 4.1 Certificates in TCG

A core concept of TCG is attestation. "Attestation is the process of vouching for the accuracy of information. External entities can attest to shielded locations, protected capabilities, and Roots of Trust. A platform can attest to its description of platform characteristics that affect the integrity (trustworthiness) of a platform. All forms of attestation require reliable evidence of the attesting entity" [17].

Each TPM will be shipped with an embedded key called Endorsement Key (EK) by the manufacturer. The EK is a RSA key pair bound to the platform. The

manufacturers provide the key pair and a certificate with the Endorsement Credentials (public part of EK, TPM model, TPM manufacturers). The platform manufacturer provides a similar certificate which is linked to the EK. The Platform Credentials which are signed by the platform manufacturers are a pointer to Endorsement Credentials, the platform type and the platform manufacturer. If the TPM attests a specific system state it provides signed hashes over collected data or parameters (using the Platform Configuration Registers, PCR).

Due to privacy issues the public part of EK is never disclosed to a challenger. Otherwise the challenger would be able to identify the TPM directly, which is not needed for the platform attestation. Therefore the TCG provides the concept of multiple Attestation Identity Keys (AIK) and AIK certificates (AI also called pseudonymous identities). AIKs are enrolled using a trusted third party (a Privacy CA). It is recommended to use a different AIK for every challenger. The Attestation Identity Credentials provided and signed by the third party (TP) include pointers to Endorsement and Platform Credentials, the issuer's name (the TP), the identity public key and an identity label. The identity label is an arbitrary textual string which was chosen by the platform owner to define an identifier for the pseudonym.

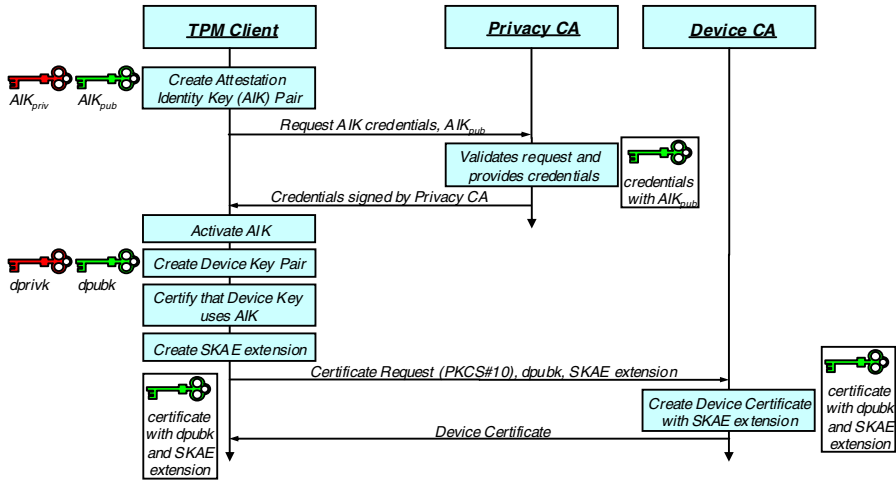
In OMA, the DRM agent would always send the same device public key to different RI. This would enable the content providers to merge their collected user profiles even if the users choose different logins at different content providers. Compared with the privacy efforts of the trusted computing group, the privacy issues of the device owner were not respected in OMA DRM V2.0.

## 4.2 TCG SKAE Certificates to Implement OMA DRM Device Certificates

The purpose of an AIK is to sign values in the PCR. AIK certificates are not intended to be used as general purpose X.509 certificates. Therefore it is not possible to use them as OMA DRM device certificate. Instead of using the AIK the Subject Key Attestation Evidence (SKAE) X.509 certificate extension (from the TCG) could be used to provide hardware bound OMA compatible device certificates.

The SKAE extension specification [18] defines a standard mechanism to represent a certified (AI) credential in X.509 certificates. This mechanism allows an OMA RI to ensure that the use of the device private key, represented by the corresponding device public key certificate, was performed with a secure TPM [18].

Fig. 4 shows the protocol to obtain an OMA DRM X.509 certificate with a SKAE extension (figure adapted from [18]). The extension embeds AI credentials which have been provided by a Privacy CA (a trusted third party). After the AIK creation the TPM requests an AIK certificate. This request includes public part of AIK ( $AIK_{pub}$ ), references to EK and platform credentials. The Privacy CA validates the request and the included credentials, and then issues an AIK credential and sends it back to the TPM. After that the TPM activates the AIK. In the next step the TPM creates a new key pair which will become an OMA device key pair ( $dpubk$  and  $dprivk$ ). TPM creates a SKAE extension which includes the AIK credentials. Finally the TPM requests the X.509 certificate from the Device CA. The request includes the extension, the device public key and a proof of possession for the device private key. The CA creates an OMA compatible certificate with the embedded SKAE extension.



**Fig. 4.** Creation of an OMA DRM device certificate with an embedded SKAE extension

The RI can use the SKAE extension to validate the AIK Credential according to certificate validation procedure defined in RFC3280. RI is also able to verify (using AIK<sub>pub</sub> embedded in the SKAE extension) TPM's certification of the device key pair.

## 5 Trust and What It Is All About

It is not a secret that the communication via Internet is often insecure. There have been various efforts to build secure applications with encryption and signature mechanisms (e.g. PGP, SET, HBCI, etc.). But especially in the context of trade there is still a huge uncertainty among clients although the e-commerce boomed in recent years. Due to [1] 38% of the Internet users avoid to buy digital goods online. The lack of trust in providers and systems may be the hardest barrier. This fact shows that there is still a big challenge in e-commerce.

"Trust" is not an objective term. It implies that a user is free in his decision whether to trust someone or not and in what circumstances [4]. In [3] the correlation between trust, security and benefit is described. They define trust as the willingness to supply a risky input that is based on the expectation that trust objects (persons, systems) do not misuse the resulting dependence in an opportunistic way respectively that they prove functioning. The higher the expected benefit from a transaction is the higher is the willingness to carry it out despite potential lack of security.

### 5.1 What Does Trust Mean in Case of DRM

Also the digital goods industry and with it the DRM content providers suffer from the uncertainty of the people. And this uncertainty is stoken by bad news headlines. The most sensational example recently was the Sony DRM XCP (Extended Copy Protection): it was installed on the computer with rootkit functions [19]. With this it

was possible to hide additional software and files from the users view and to open doors for attackers to hide malware [20]. Beyond, the users do not have a proof that the providers handle the personal data of the users confidential. Another problem is for example what happens when the user has paid for a download (he is forced to do that before downloading it) and the download breaks down for some reasons. They cannot prove it and have to trust the provider receiving the paid goods. Apparently, the expected benefit from a DRM transaction is not high enough to carry it out. There are two ways to canvass customers: either to increase the benefit for them (e.g. through additional services) or to increase the trustworthiness of the systems. Security may be an important aspect in this case. The main question arising is: “Can trust be created by the initiation of technical mechanisms?”

As a matter of fact one could say that there is no trade without mutual trust. This means that trust has to be on both sides, the merchant’s and the client’s side – this also applies in the DRM field. On the one hand the providers have to trust their customers that they pay for the requested digital goods and that they do not misuse the content (e.g. by illegal transmission of the virtual goods). On the other hand that the clients only use the files and the rights objects in that way they are allowed to. The providers solved the first problem by forcing the customer to pay for the content before he or she can download it. Potential clients have to trust the providers that they handle their personal data confidential (privacy) and that they do not install additional software on their system. Furthermore, they need to have a proof that they have paid for certain goods (non-repudiation) in case they have to provide evidence.

## 5.2 Putting It All Together

The main question we have to face up is if the OMA DRM standard together with a Trusted Platform Module (TPM) can raise the benefit from DRM systems.

The OMA DRM trust model is based on a Public Key Infrastructure (PKI). “A Rights Issuer trusts a DRM Agent to behave correctly if the DRM Agent’s certificate is verifiable by the Rights Issuer and not revoked. Similarly, a DRM Agent trusts a Rights Issuer to behave correctly if the RI’s certificate is verifiable by the DRM Agent and not revoked [9]”. This mutual authentication contributes a higher security level.

With a TPM the DRM provider possesses a solid security anchor and is not depended on obfuscation techniques. This technology helps him to balance his lack of trust to the customers. It is necessary that the DRMS with trusted computing mechanisms is evaluated according to the Common Criteria [21]. Such an accreditation could be the basis for the users trust in DRMS – as there is a proof that the software does not interfere his system otherwise as intended [17].

## 6 Conclusions and Further Work

The success of DRM systems mainly depends on mutual trust. The OMA DRM standard, created for mobile devices, implies a trust model based on PKI. Our solution provides an expanded approach which adopts this standard to untrustworthy platforms like Windows. The inclusion of the Trusted Computing technology allows a root of trust and is therefore qualified for enforcing the providers’ rights.

If these improvements are communicated in a proper way there could be a chance for improving the users trust in the whole DRM process but there is no guaranty for it due to trust being a “feeling” and very subjective. DRM providers should see trust also as an opportunity to an advantage in competition.

## References

1. Ausge. Ergeb. der Online-Umfrage IZV7, Inst. f. Wirtschaftspolitik u. Wirtschaftsforschung, Universität Karlsruhe, 2004, [www.iwww.uni-karlsruhe.de/izv/pdf/izv7\\_auswertung.pdf](http://www.iwww.uni-karlsruhe.de/izv/pdf/izv7_auswertung.pdf)
2. Singh, S.; Jackson, M.; Beekhuyzen, J.; Waycott, J.: Downloading vs Purchase: Music Industry vs Consumers, presentation to DRMTICS 2005, Sydney, 30 Oct-2 Nov 2005, available at: [www.titr.uow.edu.au/DRM2005/presentations/drm05-beekhuyzen.pps](http://www.titr.uow.edu.au/DRM2005/presentations/drm05-beekhuyzen.pps)
3. Petrovic O., Fallenböck M., Kittl C., Wolkinger T.: Vertrauen in digitale Transaktionen, *WIRTSCHAFTSINFORMATIK* 45 (2003) 1, p. 53-66
4. Pfitzmann, A.; Pfitzmann, B.; Schunter, M.; Waidner, M.: Trustworthy User Devices in Multilateral Security in communications, Volume 3, Technology, Infrastructure, Economy, Addison Wesley, München u.a., 1999
5. Website of the Open Mobile Alliance, [www.openmobilealliance.org](http://www.openmobilealliance.org)
6. Iannella, Renato: Digital Rights Management (DRM) Architectures. D-Lib Magazine, Volume 7 Number 6, June, 2001, [www.dlib.org/dlib/june01/iannella/06iannella.html](http://www.dlib.org/dlib/june01/iannella/06iannella.html)
7. OMA Digital Rights Management V1.0, DRM Specification, Approved Enabler, [www.openmobilealliance.org/release\\_program/drm\\_v1\\_0.html](http://www.openmobilealliance.org/release_program/drm_v1_0.html), Release Date: 2004/06/25
8. Website of the ODRL initiative, [www.odrl.org](http://www.odrl.org)
9. OMA Digital Rights Management V2.0, DRM Specification, Candidate Enabler, [www.openmobilealliance.org/release\\_program/drm\\_v2\\_0.html](http://www.openmobilealliance.org/release_program/drm_v2_0.html), Release Date: 2005/09/15
10. Rosenblatt, B.; Trippe, B.; Mooney, S.: Digital Rights Management, Business and Technology, M&T Books, New York, 2002
11. Nützel, J.; Beyer, A.: How to increase the security of Digital Rights Management Systems without Affecting Consumer's Security, paper accepted at ETRICS 2006
12. Website of the Trusted Computing Group (TCG), [www.trustedcomputinggroup.org](http://www.trustedcomputinggroup.org)
13. Eilam, Eldad: Reversing: Secrets of Reverse Engineering, Wiley Publishing, USA, 2005
14. Cerven, Pavol: Crackproof Your Software, No Starch Press, San Francisco, 2002
15. Website of Microsoft's Windows Vista, [www.microsoft.com/windowsvista/](http://www.microsoft.com/windowsvista/)
16. TPM v1.2 Specification Changes, October 2003, [www.trustedcomputinggroup.org](http://www.trustedcomputinggroup.org)
17. TCG Specification Architecture Overview, Specification Revision 1.2, 28 April 2004, [www.trustedcomputinggroup.org/groups/TCG\\_1\\_0\\_Architecture\\_Overview.pdf](http://www.trustedcomputinggroup.org/groups/TCG_1_0_Architecture_Overview.pdf)
18. TCG Infrastructure Workgroup, Subject Key Attestation Evidence Extension, V1.0, Rev. 7, [www.trustedcomputinggroup.org/specs/IWG/IWG\\_SKAE\\_Extension\\_1-00.pdf](http://www.trustedcomputinggroup.org/specs/IWG/IWG_SKAE_Extension_1-00.pdf), 16 June 05
19. Russinovich, M.: [www.sysinternals.com/blog/2005/10/sony-rootkits-and-digital-rights.html](http://www.sysinternals.com/blog/2005/10/sony-rootkits-and-digital-rights.html),
20. World of warcraft hackers using Sony BMG rootkit, <http://www.securityfocus.com/brief/34>
21. Website of Common Criteria, [www.commoncriteriaportal.org](http://www.commoncriteriaportal.org)

# Cluster-Based Analysis and Recommendation of Sellers in Online Auctions

Mikołaj Morzy and Juliusz Jezierski

Institute of Computing Science

Poznań University of Technology, Piotrowo 2, 60-965 Poznań, Poland

{Mikolaj.Morzy, Juliusz.Jezierski}@put.poznan.pl

**Abstract.** The expansion of the share of online auctions in electronic trade causes exponential growth of theft and deception associated with this retail channel. Trustworthy reputation systems are a crucial factor in fighting dishonest and malicious users. Unfortunately, popular online auction sites use only simple reputation systems that are easy to deceive, thus offering users little protection against organized fraud. In this paper we present a new reputation measure that is based on the notion of the density of sellers. Our measure uses the topology of connections between sellers and buyers to derive knowledge about trustworthy sellers. We mine the data on past transactions to discover clusters of interconnected sellers, and for each seller we measure the density of the seller's neighborhood. We use discovered clusters both for scoring the reputation of individual sellers, and to assist buyers in informed decision making by generating automatic recommendations. We perform experiments on data acquired from a leading Polish provider of online auctions to examine the properties of discovered clusters. The results of conducted experiments validate the assumptions behind the density reputation measure and provide an interesting insight into clusters of dense sellers.

## 1 Introduction

By the year 2006, 63% of online population will have engaged in e-commerce activities. It is estimated that already in 2006 e-commerce transactions will account for 18% of all global sales. Meanwhile, auctions are triumphantly coming back in the form of customer-to-customer (C2C for short) e-commerce model. Today, more than 250 online auction sites enable customers to trade all possible types of goods using a variety of bidding protocols. More than 1.3 millions of transactions are committed daily on online auction sites. eBay, the global leader in the online auction market, reports 95 millions of registered users and 5 millions of transactions each week. At any given point in time there are approximately 12 millions of items posted on eBay. Examination of the latest financial data published by eBay reveals an astonishing development: for the second quarter of 2005 eBay reported net revenues of \$1.09 billion (40% increase year on year), operating income of \$380 million (49% increase year on year), and net income of \$290 million (53% increase year on year).

Huge success of online auctions can be attributed to many reasons. Bidders are not constrained by time, bids are placed 24/7 and potential users are given enough time to search and bid for interesting items. The Internet removes geographical constraints on users as they do not have to physically attend an auction. Large number of sellers and buyers reduces selling costs and influences prices of offered goods. Last but not least, many users describe their bidding experiences as similar to gambling. Apart from offering new and unprecedented possibilities, online auctions provide opportunities for dishonest participants to commit fraud [13]. The lack of physical contact between involved parties decreases the degree of trust exposed by users. According to a recent Eurobarometer poll, 73% of customers who do not participate in e-commerce, refrain from doing so motivated by concerns about the security of payment, delivery issues, and warranty terms. This fear is caused mainly by the growing number of complaints regarding online auctions. American Federal Trade Commission reports that 48% of all complaints concerning e-commerce involved fraud committed in online auctions, with the total loss of \$437 million in one year. National Consumers League reveals that 63% of complaints about Internet fraud concerned online auctions, with an average loss of \$478 per person. Online fraud can occur during bidding process and after bidding ends. Popular fraudulent practices include bid shielding and bid shilling. Bid shielding consists in providing artificially high bid for an item, thus discouraging other bidders from competing for an item. At the last moment, the shielder withdraws the bid, so the winner of an auction becomes the second highest bid cooperating with the shielder. Bid shilling consists in using a false bidder identity to drive up the price of an item on behalf of the seller. After the bidding process is over, potential fraud includes refraining from paying (bidder) and sending no merchandise or sending merchandise of lower quality and inconsistent with the offer (seller). These types of fraud are dangerous from the economical point of view, because they undermine the trust that users develop toward the online auction site.

One of the mechanisms to build trust between anonymous participants of online auctions are reputation systems [11]. Reputation is perceived by auction participants as a fundamental issue in developing a successful customer-to-customer relationship [10]. Furthermore, reputation of sellers has an economically observable and statistically significant effect on price of items [5]. Unfortunately, simple reputation systems used by online auction sites do not protect participants from malicious users. Typically, the reputation of a participant is measured by the number of committed auctions, where each auction is judged by the second party as “positive”, “neutral”, or “negative”. Such simple schema is both unreliable and fraud-prone, because dishonest users can easily deceive the system into assigning unfairly high reputation ratings. A seller can create a set of virtual bidders who will “win” seller’s auctions and provide the seller with additional positive feedback points. This technique is known as “ballot stuffing” and it biases the entire system, because unearned reputation allows the seller to obtain more bids and higher prices from other users [7,12]. In order to better disguise this fraudulent practice, a seller could create a network of auctions

between virtual bidders, turning them into a clique and inflating their reputation. Another possibility is to use virtual bidders to provide artificially negative feedbacks to seller's competitors. This technique is referred to as "bad-mouthing". Bad-mouthing is more difficult to implement, because it requires to actually win a competitor's auction. Nevertheless, if the gain of driving a competitor out of the market exceeds the investment cost, bad-mouthing can be beneficial. One thing that should be stressed is the fact, that sellers and buyers are exposed to different types of risk. Sellers can postpone the shipment of an item until the payment is delivered, so the sellers are not threatened financially. On the other hand, buyers pay before receiving an item, unless using a trusted third-party, such as PayPal. The reputation of buyers has little importance for sellers, whereas the reputation of sellers is of crucial importance to buyers, who have to decide upon participating in an auction solely based on seller's reputation.

In this paper we introduce a new measure of reputation of sellers in online auctions. We draw inspiration from social network analysis. We mine the topology of links between auction participants to discover clusters of densely connected sellers. We evaluate the usefulness of discovered clusters in assessing the reputation of sellers and in providing automatic recommendations based on discovered clusters. Our original contribution includes the definition of the density reputation measure, the concept of using dense clusters for generating automatic recommendations, and the experimental evaluation of the proposed solution. The paper is organized as follows. In Sect. 2 we present the related work on the subject. Section 3 introduces the density reputation measure and presents the idea of clusters of densely connected sellers. The procedure for automatically generating valid recommendations based on discovered clusters is also explained. The properties of the new density measure are examined using thorough experiments, and the results of the experiments are presented in Sect. 4. We conclude the paper in Sect. 5 with a summary of the future work agenda.

## 2 Related Work

An anonymous, heterogeneous, and geographically distributed environment for commercial transactions requires an efficient mechanism for building trust between participants. Reputation systems [11] allow users to develop long-term business relationships and receive financial benefit for their past honest behavior. Most auction sites use the reputation system developed by eBay, where credibility is expressed as the number of positive feedbacks minus the number of negative feedbacks received by a user [5,10]. This simple mechanism suffers from several deficiencies, as pointed out in [6]. Feedbacks issued by users are subjective, lack transactional and social context, and contain highly asymmetric information. Neutral feedbacks are very rare, the spectrum for positive feedbacks is very broad, and negative feedbacks occur only when the quality of service becomes unacceptable, otherwise users refrain from posting a negative feedback in the fear of retaliation.



In recent years several new solutions have been proposed that aim at overcoming at least some of the deficiencies of feedback-based models. An interesting proposal was formulated in [1] where the authors develop a complaint-only trust model. Although originally developed for peer-to-peer environment, this highly decentralized model can be successfully used in online auctions. Another model originating from peer-to-peer environment is PeerTrust [14]. PeerTrust is a complex model consisting of many parameters, such as feedback in terms of satisfaction, number of transactions, credibility of feedback, transaction context, and community context. Method presented in [9] and further investigated in [8] does not use feedbacks to compute the reputation of participants. Instead, it uses a recursive definition of credibility and performs data mining to discover credibility estimation for each participant. A solution presented in [3] tries to prune false feedbacks and accepts only feedbacks that are consistent with the majority of feedbacks received by a given user. The need for a trusted third party is advocated in [2]. The authors propose to introduce a trusted judge that could authorize, identify, and manage the reputation of auction participants. An efficient method for assessing the level of trust between any two individuals based on a small amount of explicit trust/distrust statements per individual is presented in [4]. An interesting comparison of typical fraudulent behavior in online auctions with the abuse of customers by pay-per-call industry in the 1990s is presented in [13]. In the opinion of the author, the ability of online auction business to self-regulate is limited and not adequate to solve the problem, so legislation must be introduced to guarantee sufficient customer protection.

### 3 Density Reputation Measure

The main drawback of all feedback-based reputation systems is the fact that the reputation estimation for a given user is strongly influenced by the reputation of users directly involved in auctions with the user. This allows dishonest participants to artificially inflate their reputation estimates. Therefore, we propose a new reputation measure for sellers. Our density reputation measure computes the reputation of a given seller based on the reputation of other “similar” sellers.

Given a set of sellers  $S = \{s_1, s_2, \dots, s_m\}$ . Two sellers  $s_i$  and  $s_j$  are linked if there are at least *min\_buyers* who committed an auction with sellers  $s_i$  and  $s_j$ , and the closing price for each auction was at least *min\_value*. The number of such buyers is called the *strength* of the link and is denoted by *link* ( $s_i, s_j$ ). The *neighborhood*  $N(s_i)$  of a seller  $s_i$  consists of sellers  $\{s_j\}$ , such that the seller  $s_i$  is linked with  $s_j$ , given user-defined thresholds *min\_buyers* and *min\_value*,  $N(s_i) = \{s_j \in S : \text{link}(s_i, s_j) > 0\}$ . The cardinality of the neighborhood  $N(s_i)$  is called the *density* of the neighborhood,  $\text{density}(s_i) = |N(s_i)|$ . The rationale behind user-defined thresholds is the following: *min\_buyers* selects sellers with significant number of sales, and *min\_value* prunes low-value transactions. The density measure can be interpreted as follows: a buyer who buys from sellers  $s_i$  and  $s_j$  acknowledges the quality of both sellers. Unexperienced buyers are unlikely to link many sellers, these are rather experienced buyers who are used to link sellers. In this way the

density measure discards unreliable information from unexperienced buyers. The fact that two sellers are linked indicates that either they trade similar and popular goods (such as music or books), or that their offer is complementary (like bicycles and bicycle add-ons). Obviously, a link between two sellers may be coincidental and may not bear any useful information. Nevertheless, high density of a seller is a good indicator of seller's trustworthiness. Another important issue is the type of a cluster to which a seller is linked. Density reputation measure discovers natural groupings of sellers around product categories, thus allowing to automatically generate meaningful recommendations.

The density reputation measure does not consider the strength of the link between any two sellers, only the density of a given seller's neighborhood. In order to distinguish between strongly and weakly linked sellers we introduce another reputation measure, called score, defined as

$$score(s_i) = \sum_{s_j \in N(s_i)} density(s_j) * \log_{min\_buyers} link(s_i, s_j)$$

The score measure uses the density of each seller in the neighborhood of the current seller and multiplies it by the strength of the link between the two sellers. The logarithm is used to reduce the impact of very strong links between sellers.

The density reputation measure is very resistant to fraud and manipulation. Let us consider a malicious seller trying to enter the cluster of reliable sellers. Linking to a single trustworthy seller requires to create *min\_buyers* and investing at least *min\_buyers\*min\_value* in winning auctions of a trustworthy seller. Still, this links only to a single seller and places the cheater in the outskirts of the cluster. In order to receive higher density the cheater has to repeat this procedure several times. We attribute this feature of the density reputation measure to the fact that it uses other sellers to rate a current seller, rather than using information from buyers. We believe that it is much more difficult for cheaters to manipulate other sellers than to create virtual bidders and use them to inflate cheater's reputation.

The density reputation measure is used to provide users with automatic recommendations. When opening a page containing a given item, a user is presented with a set of top  $n$  dense sellers, who belong to the same cluster as the seller of the given item. Let  $R$  denote a set of target  $n$  sellers. Let  $d(s_i, s_j)$  denote the distance between the sellers  $s_i$  and  $s_j$  defined as the length of the shortest path connecting sellers  $s_i$  and  $s_j$  in the graph. The group density of the set  $R$  of sellers is defined as

$$density(R) = \frac{\sum_{s_r \in R} density(s_r)}{\sum_{(s_p, s_q) \in R \times R} d(s_p, s_q)}$$

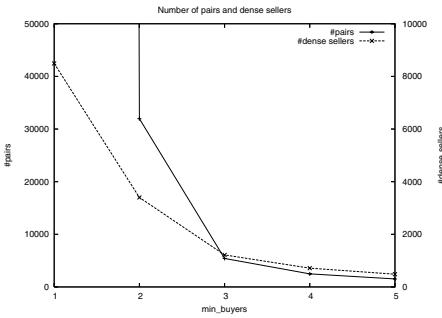
When displaying top  $n$  sellers as a recommendation for currently selected seller  $s_i$  we are trying to find the set  $R(s_i)$  of sellers who are characterized by high group density and who are close to a given seller  $s_i$ ,

$$R(s_i) = \arg \max_R \frac{\text{density}(R)}{\sum_{s_r \in R} d(s_i, s_r)}$$

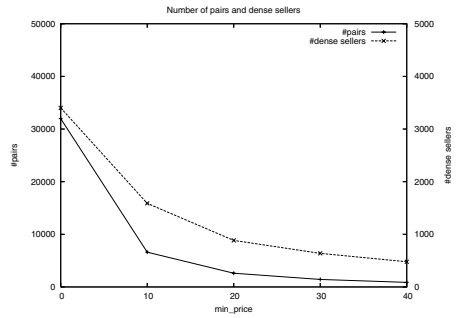
Therefore, using the recommendation system the user gains access to auctions of reliable sellers who trade goods that are similar to the searched item. Most notably, the recommendation depends on neither textual descriptions provided by sellers nor category assignments of items. This is an important feature of the recommendation system, because it allows to generate description-independent and taxonomy-independent suggestions.

## 4 Experimental Results

The data have been acquired from [www.allegro.pl](http://www.allegro.pl), Polish leader of online auctions. The dataset consists of 440 000 participants, 400 000 auctions, and 1 400 000 bids. The number of participants is greater than the number of auctions, because for each participant their highest bid is stored, whether it was the winning bid or not. Therefore, we have data on some participants who did not win any auction. Analyzed dataset is a small subset of the original data and it has been created using the following procedure: 10 000 sellers have been selected, and for this seed set all their auctions from a period of six months and participants of these auctions have been collected. Analogously, 10 000 buyers have been selected randomly and a similar procedure has been applied to this seed set. Altogether, complete information on 20 000 participants was available. Data were stored and preprocessed using Oracle 10g database.

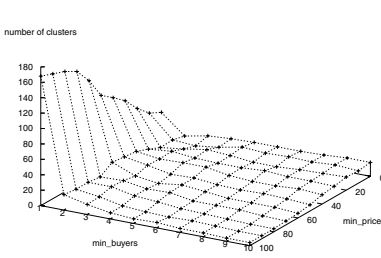


**Fig. 1.** Pairs and dense sellers (a)

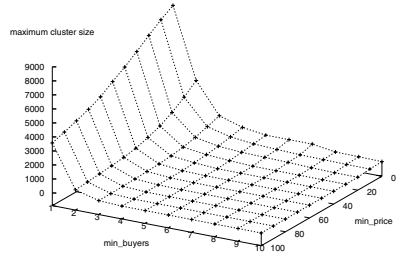


**Fig. 2.** Pairs and dense sellers (b)

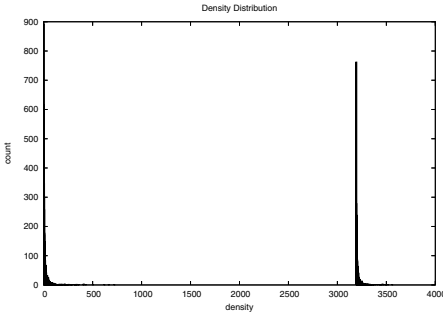
Figure 1 presents the number of linked pairs of sellers and the number of dense sellers when increasing the value of the *min\_buyers* threshold. As can be seen, even for small threshold value the number of pairs and the number of dense sellers becomes manageable. Figure 2 presents analogous results for varying the values of the *min\_price* threshold.



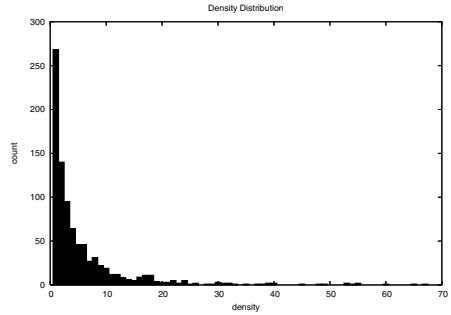
**Fig. 3.** Number of clusters



**Fig. 4.** Size of the biggest cluster



**Fig. 5.** Density distribution (a)



**Fig. 6.** Density distribution (b)

Figure 3 presents the changes in the number of discovered clusters when varying *min\_price* and *min\_buyers* thresholds. As can be seen, the *min\_price* threshold has stronger impact on the number of discovered clusters, except when *min\_buyers* = 1. The number of clusters is relatively small and for most combinations of both thresholds the space of sellers is dominated by a few clusters (usually with one main cluster being significantly bigger than the others). Figure 4 presents the size of the biggest discovered cluster for a given combination of *min\_price* and *min\_buyers* thresholds. When no thresholds are set, almost all sellers are assigned to a single cluster. Interestingly, this result suggests that Milgram’s concept of six degrees of separation applies also to the online auction environment (when discovering the borders of each cluster we never used more than five iterations to identify all members of the cluster). Another thing to notice is the fact, that the *min\_price* threshold has very little impact on the size of the biggest cluster when more than two links are used to connect sellers. For realistic settings of both thresholds the size of the biggest cluster becomes relatively small, which makes this approach suitable for automatic recommendation generation. We believe that only the most trustworthy and reliable sellers are left in the clusters, thus making respective clusters a valid source of meaningful recommendation.

Two examples of density distribution are presented in Fig. 5 (no limits on *min\_buyers* and *min\_price*) and Fig. 6 (*min\_buyers*=2 *min\_price*=\$20). When

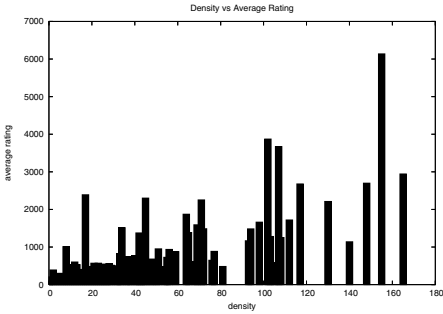


Fig. 7. Rating distribution (a)

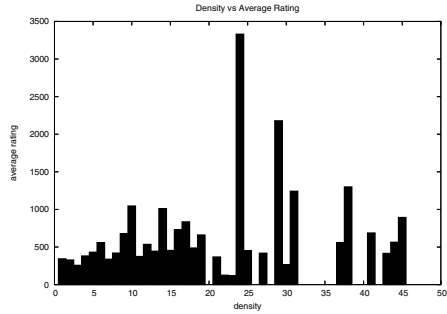


Fig. 8. Rating distribution (b)

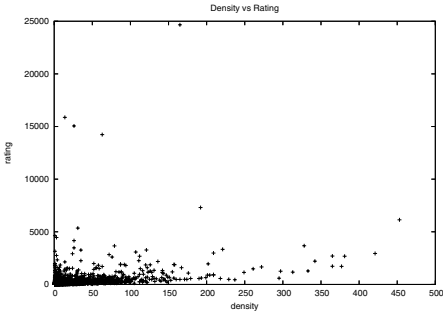


Fig. 9. Projection of density on rating

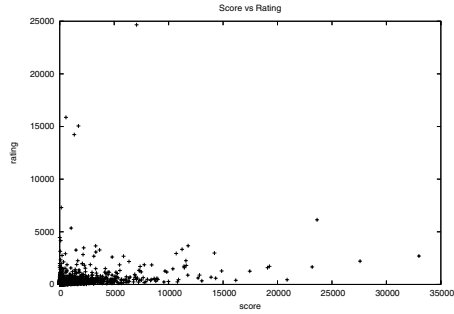
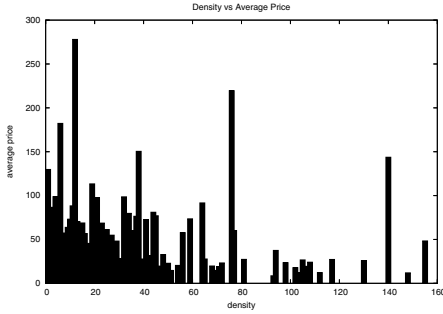


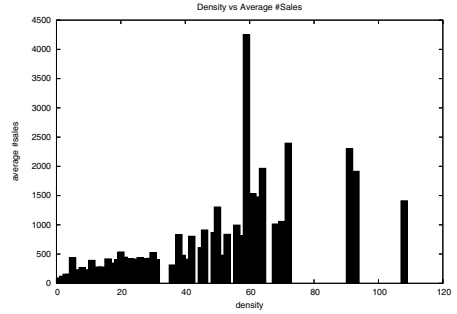
Fig. 10. Projection of score on rating

no thresholds are defined, two clusters of sellers are visible. The majority of sellers are characterized by the density from the range  $\langle 1, 500 \rangle$ , but there is also a small group of very densely connected sellers, and their density is  $\langle 3200, 3500 \rangle$ . Average density is 1217 and 8493 sellers (85% of the entire population) turned out to be dense. When thresholds are set, the average density drops to 5.9 and the number of dense sellers is 885 (8.8% of the entire population). One might argue that the *min\_price* threshold is set too prohibitively, but the average price of items in the mined dataset is close to \$30, so we rather believe, that the algorithm really discovers the set of most important and credible sellers.

An interesting question is how does the new density measure relate to traditional reputation rating computed as the aggregation of positive and negative feedbacks. The average rating distribution with respect to density is presented in Fig. 7 (*min\_buyers*=3, *min\_price*=0) and Fig. 8 (*min\_buyers*=2, *min\_price*=\$30). In general, higher density is a good indicator of high rating, but this relationship is not linear, specially when *min\_price* threshold is set to prune out low value transactions. Fig. 9 (*min\_buyers*=2, *min\_price*=0) shows the projection of average rating vs density. Many high rated sellers have low density, which is even more visible when *min\_price* is set. Sellers with high ratings are usually trading popular goods that are not expensive, so *min\_price* threshold is punishing them.



**Fig. 11.** Density and average price



**Fig. 12.** Density and average number of auctions

Similar analysis of average rating vs score is presented in Fig. 10 ( $\text{min\_buyers}=3$ ,  $\text{min\_price}=0$ ). These figures reveal a shift along the x-axis. This suggests that the sellers with low density and high rating have much higher average strength of the link than densely connected sellers.

The distribution of average price of offered items with respect to density is depicted in Fig. 11 ( $\text{min\_buyers}=3$ ,  $\text{min\_price}=0$ ) (on the figure prices are given in Polish zloty). Surprisingly, there is no clear evidence that higher density has any impact on the closing price reached by sellers. Finally, Fig. 12 ( $\text{min\_buyers}=4$ ,  $\text{min\_price}=0$ ) presents the distribution of average number of sales with respect to density. This time it is easily noticeable that highly dense sellers enjoy much larger volume of sales. This fact, more than the distribution of average price of items, convinces us, that density is a good predictor of future performance of a participant of an online auction.

## 5 Conclusions

In this paper we have presented a new density reputation measure for sellers in online auctions. Our measure considers the network of interconnections between participants and mines the topology of the network to derive useful knowledge about users. Discovered clusters of densely connected sellers can be used as a predictive of future performance of a user, thus providing additional insight into the data. In addition, discovered clusters can be used to generate description-independent and taxonomy-independent recommendations for participants of online auctions. We believe that the density of a seller can be successfully used as an indicator of seller's reliability. Main advantages of the proposed solution include resistance to manipulation, ability to discover complex fraudulent activities, and practical usability proved by experiments. The support exhibited by our commercial partners encourages us to follow the work in this area of research. Our future work agenda includes other models of user reputation, efficient use of negative and lacking feedbacks, and thorough investigation of the properties of clusters of sellers.

## References

1. K. Aberer and Z. Despotovic. Managing trust in a peer-2-peer information system. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, pages 310–317, New York, NY, USA, 2001. ACM Press.
2. S. Ba, A. B. Whinston, and H. Zhang. Building trust in online auction markets through an economic incentive mechanism. *Decision Support Systems*, 3(35):273–286, 2003.
3. M. Chen and J. P. Singh. Computing and using reputations for internet ratings. In *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 154–162, New York, NY, USA, 2001. ACM Press.
4. R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412, New York, NY, USA, 2004. ACM Press.
5. D. Houser and J. Wooders. Reputation in auctions: Theory, and evidence from ebay. Technical report, University of Arizona, 2001.
6. R. A. Malaga. Web-based reputation management systems: Problems and suggested solutions. *Electronic Commerce Research*, 4(1), 2001.
7. M. I. Melnik and J. Alm. Does a seller's ecommerce reputation matter? evidence from ebay auctions. *The Journal of Industrial Economics*, L(3), September 2002.
8. M. Morzy. New algorithms for mining the reputation of participants of online auctions. In *WINE 2005, 1st Workshop on Internet and Network Economics, 15-17 December 2005, Hong Kong*. Springer Verlag, 2005.
9. M. Morzy, M. Wojciechowski, and M. Zakrzewicz. Intelligent reputation assessment for participants of web-based customer-to-customer auctions. In *AWIC 2005, 3rd International Atlantic Web Intelligence Conference, June 6-9 2005, Lodz, Poland*, pages 320–326. Springer Verlag, 2005.
10. P. Resnick and R. Zeckhauser. *Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System*, volume 11 of *Advances in Applied Microeconomics*, pages 127–157. Elsevier Science, 2002.
11. P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. *Communications of the ACM*, 43(12), 2000.
12. P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: A controlled experiment. Technical report, School of Information, University of Michigan, 2003.
13. J. M. Snyder. Online auction fraud: Are the auction houses doing all they should or could to stop online fraud? *Federal Communications Law Journal*, 52(2), 2000.
14. L. Xiong and L. Liu. A reputation-based trust model for peer-to-peer ecommerce communities [extended abstract]. In *EC '03: Proceedings of the 4th ACM conference on Electronic commerce*, pages 228–229, New York, NY, USA, 2003. ACM Press.

# Trust Model Architecture: Defining Prejudice by Learning

M. Wojcik, J.H.P. Eloff, and H.S. Venter

Information and Computer Security Architectures Research Group (ICSA)  
Department of Computer Science, University of Pretoria  
{hibiki}@tuks.co.za,  
{eloff, hventer}@cs.up.ac.za

**Abstract.** Due to technological change, businesses have become information driven, wanting to use information in order to improve business function. This perspective change has flooded the economy with information and left businesses with the problem of finding information that is accurate, relevant and trustworthy. Further risk exists when a business is required to share information in order to gain new information. Trust models allow technology to assist by allowing agents to make trust decisions about other agents without direct human intervention. Information is only shared and trusted if the other agent is trusted. To prevent a trust model from having to analyse every interaction it comes across – thereby potentially flooding the network with communications and taking up processing power – prejudice filters filter out unwanted communications before such analysis is required. This paper, through literary study, explores how this is achieved and how various prejudice filters can be implemented in conjunction with one another.

## 1 Introduction

Technological development has influenced the principles required to run a successful economy [1]. However, the advent of new technologies and the subsequent implementations thereof have resulted in exposure to new risks. Two risk factors exist that continually drive research towards lessening the risks encountered: effective communication and security.

In order to accomplish an organisation's desired task, effective and timely communication is required. An organisation makes use of technology to communicate and share information. This information is an asset to the organisation and is used to assist decision-making processes. It is important that this information be reliable and accurate so that it can be trusted [2].

Trust models have been proposed in order to minimise the risk of sharing and successfully analysing information [3], [4]. Trust models rely on the abstract principle of trust in order to control what information is shared and with whom. Trust models evaluate the participants of a transaction and assign a numerical value, known as a trust value, to the interaction. This numerical value is used to determine the restrictions placed on the transaction and the nature of information shared. This process of analysis



occurs with all interactions an agent running a trust model encounters and can lead to an overwhelming communication load. In order to control the number of interactions a trust model encounters, prejudice filters have been proposed.

This paper introduces and defines the concepts of prejudice, trust and trust models in Section 2 by introducing a basic trust management architecture and expanding on work already done in these areas. The concept of prejudice filters and their interdependencies is explored in Section 3, with special focus on one relationship involving the learning filter. This is followed by a discussion of concepts in Section 4 and a conclusion in Section 5.

## 2 Background

Since trust model architecture is based on the concept of trust, a basic understanding of trust is required. This section introduces the concept of trust in the context of human relationships and then explores how this concept is put into practice by trust model architecture. The concept of prejudice is also explored, with special attention to how this concept can lighten communication load required to make trust-based decisions.

### 2.1 Trust Models and Trust

Trust models rely on the concept of agents [4]. Within the context of trust models, an agent refers to a non-human-coded entity used to form and participate in machine-based trust relationships. This agent would usually be situated on a computer and implement some form of logical rules to analyse the interactions with which it comes into contact in order to determine whether another agent is to be trusted or not. These logical rules may be static or adjustable by the agent in a dynamic manner, based on results of transactions the agent has participated in.

Trust is a subjective concept – the perception of which is unique to each individual. Trust is based on experience and cognitive templates. Cognitive templates are templates formed by experiences that are later used to analyse future experiences of a similar nature. Trust is dynamic in nature and influenced by environment, state and situation. According to Nooteboom [5], "[s]omeone has trust in *something*, in some *respect* and under some *conditions*".

Each of the four key concepts highlighted by Nooteboom exists within trust model architecture. *Someone* and *something* define two agents participating in an interaction. The former refers to the instigator of the interaction while the latter refers to the agent accepting the request. The *respect* is defined by the reason for instigating an interaction. Finally, the *conditions* refer to the situational factors that influence the success of an interaction.

### 2.2 Trust Model Architecture

Trust models assist agents that have not previously encountered one another by forming and participating in trust-based interactions. Various experts have already

proposed numerous trust models [6], [7], [8]. A survey of the literature conducted by the author has identified four components that have been used in trust model implementation: trust representation, initial trust, trust dynamics and trust evaluation.

Catholijn M. Jonker and Jan Treur [9] focus on how trust is represented by agents in order to simulate intelligence and make trust-based decisions. They propose a simple qualitative method of representing trust that defines four basic trust values. These values include unconditional distrust, conditional distrust, conditional trust and unconditional trust. Other issues of trust representation, as identified by Damiani, De Capitani di Vimercati and Samarati [10], include protocols that are required in order to communicate and discern trust related information. These protocols are required to identify and analyse trust related information in anonymous environments as well as to control what identity information is released under specific circumstances.

Jonker and Treur in further research state that trust models incorporate trust characteristics that can be divided into two states. These states refer to initial trust – the initial trust state of an agent – or trust dynamics – the mechanisms that allow for the change in and updating of trust [9]. The initial trust state of an agent determines the agent's predisposition wherein the agent can be predisposed towards trust, distrust or neutrality. Taking the dynamic nature of trust in consideration, Marx and Treur [8] concentrate on a continuous process of updating trust over time. Experiences are evaluated and used by a trust evolution function.

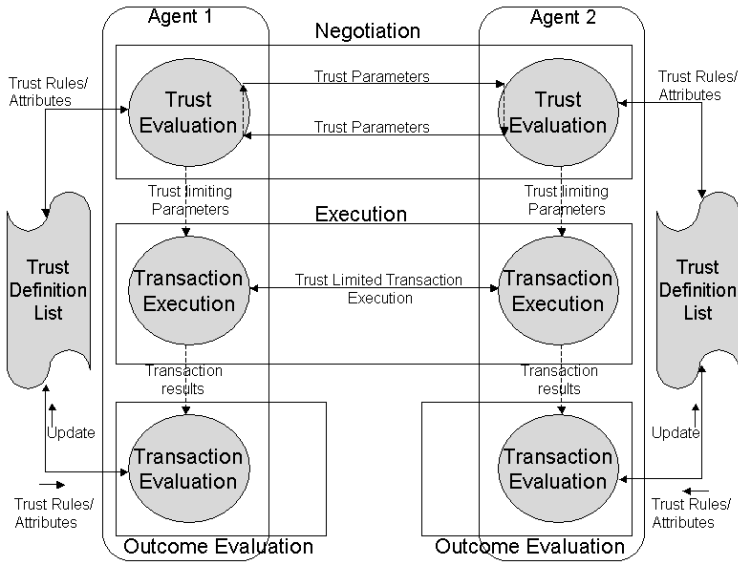
Changing trust values requires that some form of trust evaluation should take place. The reputation-based model of Li Xiong and Ling Liu [11], known as PeerTrust, emphasises the importance of this evaluation process by evaluating various parameters, such as nature of information shared and purpose of interaction, in order to update the trust value an agent retains.

Trust models are able to obtain trust values in several manners. Trust information and state can be pre-programmed into the agent as a list of parameters. These parameters can also be dynamically formulated, based on pre-defined and logically formed trust rules that an agent uses to evaluate trust.

### 2.3 Example of a Typical Trust Architecture

According to Ramchurn *et al.* [12] basic interactions among agents go through three main phases. These phases are negotiation, execution and outcome evaluation. Trust plays an essential part in all three of these phases. This is illustrated by Figure 1.

Two agents attempting to communicate with one another are first required to establish a communication link, usually initiated by one agent and accepted by another. This process initiates a negotiation process whereby two agents negotiate various parameters, such as the security level of information that is to be shared or the services for which permission will be granted, that will define boundaries of the interaction. A trust value for the interaction is defined through comprehensive analysis of logical rules. The simplest way of storing and implementing these rules is to have them present in a list that the agent accesses and processes. In Figure 1, storage of these rules occurs in the trust definition list.



**Fig. 1.** Operation of an agent using a trust model

The successful negotiation and establishment of a trust value triggers an analysis of the trust value. Provided the trust value is above a certain acceptable threshold, the transaction execution process is started. Trust models control the context of the interaction during the execution phase, limiting trust given and hence controlling which information or services are accessible and which are not.

Once transaction execution has terminated, the results of the interaction are sent to the transaction evaluation process. This process evaluates the results and updates the trust definition list in either a positive or negative manner. Negative updating of the logical rules occurs due to business transaction failure, while business transaction success will trigger a positive update.

The evaluation of trust among agents is a time-consuming process that requires comprehensive evaluation of the defined logical rules in order to attain an accurate trust value to be used during an interaction. Only once the trust value has been obtained, the agents will decide whether to participate in a transaction or not.

In a networking environment, the amount of possible agents that will request participation in such an interaction can be vast. To successfully assess another agent, agents pass several messages to obtain the required information that is to be analysed against the defined trust parameters. For instance, the formal model for trust in dynamic networks proposed by Carbone, Nielson and Sassone [7] passes delegation information between agents in order to create a global trust scheme. Delegation allows a particular agent to trust another agent, based on the fact that the other agent is trusted by agents that the agent in question trusts. This reliance on the passing of messages exposes the network to the possibility of network overload. Another potential problem arising during the process of establishing trust is the level of comprehensiveness

required by the analysis process. Having a large number of strict rules define a trust relationship limits the communications an agent will be able to participate in, while at the same time adding to the analysis load. Rules that are too generic open the system up to a higher level of risk by allowing an agent to participate in interactions with other agents that have not been fully analysed for trustworthiness.

Prejudice filters have been proposed to lessen the number of interactions that require comprehensive trust evaluation [13] so as to solve the problems mentioned above. Stereotyped grouping of interactions allows for characteristics to be assumed instead of evaluated in detail. It also allows trust evaluation to focus on characteristics that are not assumed, instead of evaluating the interaction against the entire list of logical rules that represent expectations.

### 3 Prejudice Filters

In order to understand the concept of prejudice filters, an understanding of prejudice is required. Prejudice is an extension of the concept of trust-building processes and is defined as a negative attitude towards an entity, based on stereotype. It is important to note that the negative nature of prejudice allows negative assumptions in order to evaluate trust. Prejudice influences trust by allowing certain negative assumptions to be made about certain groups. These negative assumptions are based on prior knowledge and experience with such groups. All entities of a certain stereotyped group are placed in the same category, allowing assumptions to be made and simplifying the processing required before trust can be established [14]. This way an agent only needs to analyse attributes it does not have assumptions about in order to adjust trust value. An agent is allowed to completely distrust an agent simply because it falls into a category which it perceives as negative.

Agents see prejudice filters as simplified trust rules that rely on the concept of prejudice in order to limit the number of interactions an agent needs to analyse in detail. Prejudice filters rely on broad definitions of attributes that lead to distrusted interactions, thus denying interactions that can be defined by these attributes. For example, if an agent has interacted with another agent from a specific organisation and the interaction failed in terms of expectations, future requests from agents belonging to the same organisation will be discriminated against. Figure 2 illustrates where prejudice filters extend the trust architecture as originally depicted in Figure 1.

Prejudice filters affect two phases of the three-phase interaction cycle: the negotiation and outcome evaluation phases. In the negotiation phase, the prejudice filters are consulted first to provide a quick, simplistic evaluation of trust in order to filter unwanted communications before they are required to go through detailed trust evaluation and definition. Once an interaction has passed the prejudice evaluation, it moves onto the trust evaluation in order to acquire a trust value. When the execution phase concludes, the outcome evaluation phase includes the prejudice parameters when it evaluates the interaction. Failed transactions update the prejudice filters in order to filter out other transactions of a similar nature at an earlier stage.

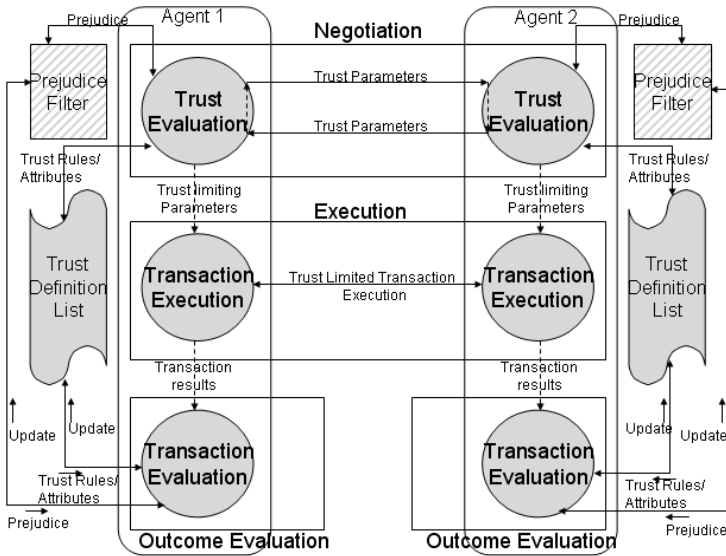


Fig. 2. Operation of an agent using a trust model with prejudice filters

### 3.1 Extending Existing Models to Include Prejudice Filters

Existing trust models rely on various means of establishing trust, which include recommendation, reputation, third party reference, observation, propagation, collaboration, negotiation and experience [1], [2], [4], [6], [7], [8], [9], [10], [11], [12]. Based on these, five means of implementing prejudice filters have been identified by the author in order to simplify the extension of existing models to include prejudice. These five are as follows [13]:

*Learning:* When using the learning filter, prejudice is not defined explicitly. An agent relies on ‘first impressions’ to learn prejudice. If an interaction fails, the agent analyses the interaction’s attributes and looks for unique attributes of other interactions that were previously encountered and found to be successful. These unique attributes are used to create a category to be used as a prejudice filter.

*Categorisation:* An agent creates various categories that are trusted. If an interaction request does not fall into a trusted category, the agent filters out that interaction in a prejudiced manner. This can also be implemented in a reverse manner where an agent creates categories that are distrusted and filters out communications that fall into those categories. Categories can also be created to represent various levels of trust. Any interactions falling into such categories are assigned the default trust value associated with that particular category.

*Policy:* Policies define the operational environment in which an agent exists and affect parameters of interactions that are regarded acceptable. Policy-based prejudice filters out interactions with agents whose policies differ from the agent doing the filtering. One way of doing this is to request data on the country an agent resides in. Such data defines the laws and culture that bind business interactions for that agent, as well as controls the means in which data and confidentiality are handled.

*Path:* Path-related prejudice allows an agent to refuse an interaction, simply because of the fact that the path of communication between two agents passes through a distrusted agent.

*Recommendation:* Agents that are trusted to make recommendations are known as recommender agents. Implementing prejudice by using recommendation allows a particular agent to only trust other agents that are trusted by the particular agent's recommender agents.

The above five filters can be incorporated into current trust models to extend their capability, while at the same time allowing for these filters to merge with a particular trust model's main philosophy. Just as some models use a combination of concepts to implement the concept of trust, interrelated filters can be implemented in different combinations in order to optimise their effectiveness.

3.2 Defining Interrelationships Between Filters

The five prejudice filters discussed above can be organised into a structure of relationships as shown in Figure 3. This structure depicts relationships that exist between these filters. The root node of a relationship between two prejudice filters indicates the dominant filter. The second filter can be incorporated into the workings of the dominant filter when the two are implemented together. The directional arrows in Figure 3 illustrate this. The dominant filter is situated at the tail of the directional arrows. Two prejudice filters emerge as more dominant than the others: learning and policy. These prejudice filters are always situated at the tail end of the arrows in Figure 3 and can be implemented in conjunction with all the other lesser filters.

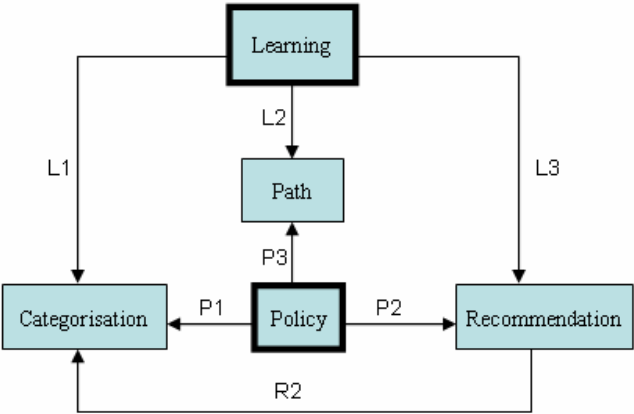


Fig. 3. Overview of the inter-relationships between prejudice filters

Due to space constraints, only one of the illustrated relationships is explored, leaving the rest for further discussion in future work. The relationship discussed has the learning filter as its root node and is labelled L1 – linking the learning and categorisation prejudice filters.

**Learning-Dominated Relationships.** The nature and success of learning is governed by the nature and variety of information and experience that an agent is exposed to [15]. Experiences and information are filtered to form templates unique to each agent. Templates are default rules that have been formed by experiences and that are subsequently used to evaluate other similar experiences.

When using learning, prejudice is not defined explicitly, and an agent relies on 'first impressions' to learn prejudice. Possible implementation of this concept allows an agent to begin with a basic set of rules that it uses to evaluate the success of an interaction. Initially, the agent will interact with any agent with which it comes into contact, under restricted conditions of trust. Each interaction instigates an analysis process by means of which the agent will identify parameters such as location of an agent, security of information required, and even factors such as an agent's reputation. These parameters become the characteristics of the particular interaction and should the interaction fail, they will be analysed in order to identify a means of filtering out future interactions of a similar nature.

Due to the fact that learning creates various forms of templates [16], learning various forms of prejudice can be accomplished. One of these is discussed below.

*Learning by Categorisation (L1).* Categorisation is an umbrella term that allows for objects or concepts with similar attributes to be grouped together. This allows for certain assumptions to be made in order to simplify analysis of such objects. The attributes that can be assumed are those that define a certain object or concept as belonging to a specific category. For instance, agents that belong to the same policy category are assumed to hold similar policy values, such as information privacy constraints. Only agents from acceptable categories will be sent for trust evaluation by an agent wishing to interact with another. Agents that are defined as unacceptable at the onset of the interaction are discarded before entering the comprehensive trust evaluation phase. This eases the processing load by filtering out undesirable categories before sending the interaction to the trust evaluation process which determines a trust value.

The process of learning prejudice relies heavily on categorisation. Learning analyses a transaction to determine its unique features. If the transaction fails, the agent uses this analysis process to create a category of failure to be used in future category-based prejudice decisions. Implementation of this concept relies on allowing an agent to form categories defined by the trust rules in place. For instance, if the trust rules in place require transactions to be analysed in order to determine the policies used by the agents in question, these agents can be categorised by their policies and characteristics. Agents can be categorised by their core services, products and policies [17].

An agent is required to either keep a list of categories that are trusted or categories that are not trusted. Whenever a new interaction is encountered, the interaction is analysed against the characteristics of the various categories in order to define the category the interaction belongs to. Once the category has been defined, the agent checks its list of trusted or distrusted categories in order to determine whether interactions of that nature are trusted. If the interaction type exists in the distrusted categories list or alternately does not exist in the trusted list, the interaction is seen as

distrusted and is then discarded. Unknown or undefined categories are by default considered to be distrusted.

Categorisation can also be used to define different levels of trust. This is accomplished by assigning a default trust value associated with a category to agents that fall into that category. The rights delegated to an interaction are consequently limited by the category to which it belongs [6]. An example of such a category is role. Various roles are given differing rights. An administrative role is given more access rights than a client role.

## 4 Discussion

The concept of implementing prejudice as discussed in this paper is a very new concept that still requires further experimentation and analysis. One of the shortcomings of these filters is related to the fact that they allow machines to deny access due to the values of prejudice that were obtained.

This can lead to a situation in which agents that are in actual fact trustworthy are seen as untrustworthy, simply because of the prejudice filter in place. A situation like this, however, can be controlled by allowing agents to interact with several agents with similar defined characteristics before deciding prejudice against them. Increasing the number of interactions in which an agent participates increases the risk an agent is exposed to. Thus, there is a trade-off between accuracy of prejudice prediction, and the risk an agent is willing to take.

## 5 Conclusion

This paper has introduced the concept of trust models and prejudice. Different means of incorporating prejudice include categories, policies, path, recommendation and learning. Several of these filters are related in such a manner that they may be implemented in conjunction with one another. One of these relationships, namely that between learning and categorisation, has been explored and defined by this paper.

The authors have explored this topic from a conceptual standing that requires implementation and testing. Since only one relationship was scrutinised in this paper, further work requires more detailed investigation of the other defined existing relationships. More in-depth work needs to be done on means to standardise the representation of trust-related data, thus allowing agents from various platforms and using various models to efficiently interact with one another.

## References

1. Hultkrantz, O., Lumsden, K., E-commerce and consequences for the logistics industry. In: Proceedings for Seminar on "The Impact of E-Commerce on Transport." Paris (2001)
2. Patton, M.A., Josang, A., Technologies for trust in electronic commerce. In Electronic Commerce Research, Vol 4. (2004) 9-21



3. Abdul-Rahman, A., Hailes, S., A distributed trust model: new security paradigms workshop. In Proceedings of the 1997 workshop on new security paradigms, Langdale, Cumbria, United Kingdom, (1998) 48-60
4. Ramchurn S.R., Sierra, C., Jennings, N.R., Godo, L., A Computational Trust Model for Multi-Agent Interactions based on Confidence and Reputation. In: Proceedings of 6th International Workshop of Deception, Fraud and Trust in Agent Societies, Melbourne, Australia, (2003) 69-75
5. Nooteboom, B., Trust: forms, foundations, functions, failures, and figures. Edward Elgar Publishing, Ltd., Cheltenham UK. Edward Elgar Publishing, Inc. Massachusettes, USA. ISBN: 1 84064 545 8 (2002)
6. Papadopou, P., Andreou, A., Kanellis, P., Martakos, D., Trust and relationship building in electronic commerce. In: Internet Research: Electronic Networking Applications and Policy, Vol 11. No. 4 (2001) 322-332
7. Carbone, M., Nielsen, M., & Sassone, V., A formal model for trust in dynamic networks. In: Software Engineering and Formal Methods. In: Proceedings of the First International Conference on 25-26 Sept. (2003) 54-61
8. Marx, M., Treur, J., Trust dynamics formalised in temporal logic. In: Proceedings of the Third International Conference on Cognitive Science, ICCS (2001) 359-362
9. Jonker, C.M., Treur, J., Formal Analysis of Models for the Dynamics of Trust based on Experiences. In: Proceedings of MAAMAW'99. LNAI (1999).
10. Damiani, E., De Capitani di Vimercati, S., Samarati, P., Managing Multiple and Dependable Identities. IEEE Internet Computing, November-December 2003.
11. Xiong L., Lui L., A Reputation-Based Trust Model for Peer-to-Peer eCommerce Communities. E-Commerce, IEEE International Conference on 24-27 June (2003) 275-284
12. S.R., Sierra, C., Jennings, N.R., Godo, L., Devising a trust model for multi-agent interactions using confidence and reputation. In: Applied Artificial Intelligence. , Vol. 18., (2004) 833-852
13. Wojcik, M., Venter, H.S., Eloff, J.H.P., Olivier, M.S., Incorporating prejudice into trust models to reduce network overload. In: Proceedings of South African Telecommunications and Networking Application Conference (SATNAC 2005). SATNAC, Telkom, CD ROM Publication. (2005)
14. Bagley, C., Verma, G., Mallick, K., Young, L., Personality, self-esteem and prejudice. Saxon House. , Teakfield Ltd, Westmead. Farnborough, Hants. England. ISBN: 0 566 00265 5 (1979)
15. Bowling, M., Manuela, V., Multiagent learning using variable rate. In: Artificial Intelligence. Vol. 136 (2002) 215-250
16. Dasgupta, D., Artificial neural networks and artificial immune systems: similarities and differences. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '97), Orlando, October 12-15 (1997)
17. Siyal, M.Y., Barkat, B., A novel Trust Service Provider for the Internet based commerce applications. In Internet research: electronic networking applications and policy, Vol. 12(1) (2002) 55-65

# How to Protect a Signature from Being Shown to a Third Party<sup>\*</sup>

Marek Klonowski<sup>\*\*</sup>, Przemysław Kubiak<sup>\*\*\*</sup>,  
Mirosław Kutylowski, and Anna Lauks

Institute of Mathematics and Computer Science  
Wrocław University of Technology,  
{Marek.Klonowski, Przemyslaw.Kubiak, Miroslaw.Kutylowski,  
Anna.Lauks}@pwr.wroc.pl

**Abstract.** Many attempts to controlling who and under which circumstances can verify our signatures have been made so far. For this purpose one can use undeniable signatures, designated confirmer signatures or designated verifier signatures. We introduce a model of new kind of signatures, called *dedicated digital signatures* (or *dds* for short). The core idea is that a designated verifier can present a standard signature of the signer derived from dds to a third party, but at the price of revealing the private key of the designated verifier or at the price of revealing the designated verifier's signature of a particular message. Therefore the verifier will show the signature only in very special situations. We present a construction of a dds based on ElGamal signatures and its modifications that allow to obtain additional important features.

## 1 Introduction

**Previous Work.** Classical digital signatures enable a holder of the signature to convince anybody about who has signed the document. The signer has no control on who and under which circumstances can see and verify his signature.

One of the attempts to control the flow of signatures are undeniable signatures proposed in [1] by Chaum and van Antwerpen. Undeniable signature can be distributed and verified any number of times, but its validity cannot be checked without interaction with the author of this signature. If he refuses to cooperate, then the recipient cannot verify validity of the signature. The problem appears when signer becomes unavailable. David Chaum introduced designated confirmer signatures [2] that solve the mentioned weakness of undeniable signatures. The designated confirmer protocol involves three parties - a signer, a recipient and a so-called confirmer. In this scheme the recipient cannot prove having a valid signature unless he gets some help from the confirmer.

---

<sup>\*</sup> Partially supported by KBN project 2003–2005, grant 0 T00A 003 23.

<sup>\*\*</sup> Contact author, Beneficiary of Domestic Grant for Young Scientists awarded by The Foundation for Polish Science.

<sup>\*\*\*</sup> On a leave from Adam Mickiewicz University.

The second disadvantage of undeniable signatures is that a witness of a verification procedure can become convinced about the outcome of the verification protocol. In [3] Jakobsson et al. introduced designated verifier proofs - a method which can be helpful to solve this problem. They showed how to change the traditional verification protocol in such a way that only a designated verifier can become convinced about signature's validity. Also in so-called chameleon signatures presented in [4] the receiver cannot prove validity of the signature.

**Our Contribution.** In this paper we provide a formal model of new kind of signatures – *dedicated digital signature*. In our scheme signer can construct his signature in such a way that the recipient (called designated verifier) cannot show this signature to third parties without being punished for that. Namely, in the protocol concerned the signer gives the recipient *dedicated digital signature* or *dds* for short. After receiving dds, the verifier derives a standard signature of the signer. So in principle, the verifier can present this signature to third parties. The point is that the signature together with the dds reveal the private key of the verifier or, depending on the protocol version, his signature of a particular message. We provide a formal model of this scheme as well as an example of construction such a protocol based on ElGamal digital signatures. Extensions of the basic scheme are designed so that a designated verifier can show only some number of such signatures to a third party without revealing his secrets.

## 2 Formal Model of Dedicated Digital Signature

Dedicated digital signatures can be regarded as a set of four algorithms that can be efficiently executed in terms of security parameters. We assume presence of signer, designed verifier and „regular” verifier in the protocols. We also assume that two pairs of keys are generated  $(pk, sk)$ ,  $(pk_{ver}, sk_{ver})$ , the public and private keys of the signer and designated verifier respectively.

**DDS Creation.** This protocol takes as an input message  $m$ , random parameter  $r$ , private signing key  $sk$  and public key of designated verifier  $pk_{ver}$ .

$$\sigma \leftarrow \text{DDSC}(m, sk, pk_{ver}, r)$$

Let  $\sigma$  be produced dds.

**Signature Retrieving.** Having dds  $\sigma$  of  $m$  under public key  $pk$  and secret key  $sk_{ver}$  one can find signature  $\sigma'$  of  $m$ .

$$\sigma' \leftarrow \text{RET}(\sigma, pk, sk_{ver}, m)$$

The intuition behind that is as follows: owner of secret key  $sk_{ver}$  (i.e. designated verifier) can retrieve signature of  $m$ .

**Verify.** Thanks to deterministic algorithm VERIFY one can check if signature  $\sigma'$  is a valid signature of  $m$  under secret key  $sk$ .

$$\text{VERIFY}(m, \sigma', pk) \rightarrow \{\text{TRUE}, \text{FALSE}\}$$

**Punish.** Having signature  $\sigma = \text{DDSC}(m, sk, pk_{ver}, r)$  and  $\sigma'$  one can retrieve secret key  $sk_{ver}$  corresponding to  $pk_{ver}$  using protocol PUNISH.

$$sk_{ver} \leftarrow \text{PUNISH}(m, pk, \sigma, \sigma')$$

### 3 Dedicated Digital Signatures Based on ElGamal Scheme

In this section we present dds scheme based on ElGamal signing scheme. We describe two scenarios that differ in the form of punishing the designated verifier for revealing the signature. The first form of punishment is revealing the verifier's private key, the second one is revealing the verifier's signature of a particular message.

In order to make the text more clear we hereafter skip in notation “mod” whenever it is obvious from the context.

**Preliminaries.** Let us assume that Alice would like to prepare a dds signature of a message  $M$  dedicated for Bob. Let  $g \in \mathbb{F}_p^*$  and  $\text{ord } g$  has no small prime factors. We assume for simplicity that  $\text{ord } g = q$ , where  $q$  is some very large prime divisor of  $p - 1$ . We assume that Alice and Bob use the same  $p$  and  $g$ . Let  $x$  and  $x_1$  denote randomly generated private keys of, respectively, Alice and Bob. Let  $y = g^x$ ,  $y_1 = g^{x_1}$  be the corresponding public keys.

#### 3.1 DDS Leaking Verifier's Private Key

**Creation of a Dedicated Signature.** In order to create a signature of a message  $M$ , dedicated to Bob, Alice first chooses  $k \in \{1, 2, \dots, q - 1\}$  uniformly at random. Afterwards she computes:

$$\begin{aligned} a &:= y_1^k, \\ b &:= k^{-1} (H(M) - ax) \bmod q, \end{aligned}$$

where  $H$  is an appropriate hash function. Then the dds of  $M$  is the pair  $(a, b)$ . So it is an ElGamal signature of  $M$ , except that instead of the assignment  $a := g^k$  from the original scheme we have the assignment  $a := y_1^k$ . Observe that  $y_1$  is the public key of Bob, to whom the signature of  $M$  is dedicated.

**Transformation of a Dedicated Signature.** Bob can easily get an ElGamal signature  $(\hat{a}, \hat{b})$  of Alice from  $(a, b)$ . Namely, he puts:

$$\hat{a} := a, \quad \hat{b} = x_1^{-1} \cdot b \bmod q.$$

One can easily see that  $(\hat{a}, \hat{b})$  is a standard ElGamal signature of  $M$ . Indeed,  $a = (g^{x_1})^k = g^{x_1 \cdot k}$ . So, the second part of the ElGamal signature with the first coefficient  $a$  equals

$$(x_1 \cdot k)^{-1}(H(M) - ax) = x_1^{-1} \cdot k^{-1}(H(M) - ax) = x_1^{-1} \cdot b \pmod{q}.$$

**Presenting a Signature to Other Parties.** Assume that Bob has shown the signature  $(\hat{a}, \hat{b})$ . Then anybody who has access to the dds  $(a, b)$  (for instance Alice) can retrieve Bob's private key  $x_1$  from equality  $\hat{b} = b \cdot x_1^{-1} \pmod{q}$ .

### 3.2 DDS Revealing a Verifier's Signature

We consider the following scenario:

- Bob prepares a pre-signature of a certain message, say  $M_1$ .
- Using the pre-signature of Bob, Alice constructs a dds for a message  $M$  of her choice.
- After getting the dds of  $M$ , Bob can transform it to an Alice's signature of  $M$ .
- If Bob shows this signature to Paul that knows the dds of  $M$ , then Paul can derive Bob's signature for  $M_1$ .

**Preparation of a Pre-signature of  $M_1$  by Bob.** In order to prepare a pre-signature of  $M_1$ , Bob creates a standard ElGamal signature  $(a_1, b_1)$  of  $M_1$ :

$$\begin{aligned} a_1 &:= g^{k_1}, \\ b_1 &:= k_1^{-1} \cdot (H_1(M_1) - a_1 \cdot x_1) \pmod{q}, \end{aligned}$$

where  $k_1 \in \{1, 2, \dots, q-1\}$  is chosen uniformly at random and  $H_1$  is an appropriate hash function. (Of course, we require that  $b_1 \neq 0$ , since otherwise private key  $x_1$  would be revealed.) Afterwards Bob reveals  $(a_1, h) = (a_1, g^{b_1})$  as a pre-signature of  $M_1$ .

Note that Alice is able to compute  $a_1^{b_1}$ . Indeed:

$$a_1^{b_1} = g^{H(M_1)} \cdot y_1^{-a_1}. \quad (1)$$

Bob must also provide a zero knowledge proof (ZKP for short) of equality of two discrete logarithms – i.e.

$$EQDL(a_1, a_1^{b_1}, g, h). \quad (2)$$

We see that in the equality (1) delivered from the standard signature verification condition Alice has been used the  $H(M_1)$ . Having  $(a_1, b_1)$  in the standard protocol she would be convinced by (1) of having a valid signature of  $M_1$ . But in this protocol the proof of (2) is needed to convince Alice, that  $h$  she got from Bob corresponds to  $a_1^{b_1}$ , and by (1) to  $M_1$ .

The protocol for *EQDL* (see [5]) is an extension of the ZKP of possessing the exponent  $b_1$  (see [6]). As a result, Bob also delivers a proof that he really knows  $b_1$ . This is crucial, since anyone might produce the pair  $(a_1, h)$  as follows:

$$\begin{aligned} a_1 &:= g^{k_1}, \\ a_1^{b_1} &:= g^{H(M_1)} \cdot y_1^{-a_1}, \\ g^{b_1} &:= (a_1^{b_1})^{k_1^{-1}}. \end{aligned}$$

**Creation of a Dedicated Signature.** In order to create a signature of a message  $M$ , dedicated to Bob, Alice first chooses a random value  $k \in \{1, 2, \dots, q-1\}$ . Afterwards she computes:

$$\begin{aligned} a &:= h^k, \\ b &:= k^{-1} \cdot (H(M) - a \cdot x) \bmod q. \end{aligned}$$

Then  $(a, b)$  is a dds of  $M$ .

**Transformation of a Dedicated Signature.** Having  $(a, b)$  Bob can easily get an ElGamal signature of Alice. Namely, like in the previous scenario, he computes:

$$\hat{a} := a, \quad \hat{b} := b_1^{-1} \cdot b \bmod q.$$

Since  $b_1$  is required in the above operations, Bob as a creator of  $(a_1, b_1)$  is the only person who can get Alice's signature of  $M$ .

**Presenting a Signature to Other Parties.** Assume that Bob shows signature  $(\hat{a}, \hat{b})$  to Paul. Then after getting access to the second component of Alice's original dds signature  $(a, b)$  of the message  $M$  (from Alice herself for example), Paul can retrieve  $b_1$ , the missing second component of Bob's signature of  $M_1$ , from equality  $\hat{b} = b \cdot b_1^{-1} \bmod q$ .

### 3.3 Difficulty of Preventing Disclosure

Bob would avoid disclosure of a key or of a signature provided that it is possible to construct another valid ElGamal signature  $(a', b')$  of the message  $M$  based on the signature  $(\hat{a}, \hat{b})$ . As far as we know, no method of this kind is known. Moreover, it is regarded as a nice feature that one can request a second signature of the same  $M$  (just like for a handwritten signature). Note also that during a search for a new  $(a', b')$  the right side of the equality

$$a'^{b'} \cdot y^{a'} = g^{H(M)}$$

is fixed, unlike in the case of forgery of Alice's signature of a message  $M'$  when an attacker has some influence on  $M'$  (for example by using various vocabulary or by introducing extra spaces). In the later case a kind of a meet-in-the-middle

attack can be mounted: the forger collects a set of pairs  $(a', b')$  and a set of messages  $M'$  with the same meaning as  $M$ . Then he seeks for a “collision” between the sets of values  $a'^{b'} \cdot y^{a'}$  and  $g^{H(M')}$ .

However, it should be noted that Bob can prove that it has a signature of Alice without transferring it. All standard signature schemes allow zero knowledge proofs.

## 4 Scheme Extensions

It should be noticed that the extensions presented below can be applied to the schemes presented in 3.2 as well as in 3.1 i.e. schemes with reviling private key as well as signature of particular message.

### 4.1 Revealing Private Key Without Cooperation with the Signer

According to the basic scheme presented in the previous section every verifier  $V$  can retrieve a private key or a signature of the designated verifier  $V_d$  provided that the component  $b$  is available. It happens when  $V_d$  has published signature  $(\hat{a}, \hat{b})$  and the signer  $S$  has access to these values. Another case is when signer  $S$  broadcasts the parameter  $b$ .

In many scenarios it would be very useful to enable getting a secret of the designated verifier automatically, without cooperation with the signer  $S$ . Below we present a simple modification of the scheme presented in Section 3.1 that meets this requirement. It is based on the idea that we force  $V_d$  to show  $b$ .

### Signature Creation

1. A signer  $S$  generates a dds  $(a, b)$  for a message  $M \otimes R$ , where  $R$  is a random bit sequence and  $\otimes$  denotes bitwise exclusive-OR operation.
2.  $S$  attaches a standard signature  $s$  of a message  $(R, b)$  to  $(a, b)$ .

**Signature Verification by a Third Party.**  $V_d$  shows the ElGamal signature of  $M \otimes R$ : to prove that this is actually signer’s  $S$  signature of  $M$  he has

1. to reveal parameter  $R$  and
2. to prove that  $R$  was actually signed by  $S$ . For this purpose  $V_d$  he must reveal signature  $s$ . However, to show it’s validity, presenting parameter  $b$  is also necessary.

### 4.2 Multi-key Scheme

Let us consider a scenario when a signer  $S$  wants two private keys, say  $x_1$  and  $x_2$  (belonging to possibly different verifiers  $V_{d,1}, V_{d,2}$ ) to be revealed only after publishing both ElGamal signatures corresponding to the dds’es addressed to

$V_{d,1}$  and  $V_{d,2}$ , respectively. We assume that  $y_1 = g^{x_1}$  and  $y_2 = g^{x_2}$ , and message  $M_i$  is addressed to  $V_{d,i}$ .

First we describe a simple protocol that has a couple of serious drawbacks. However, it will help us to understand the second solution.

- The first parameters of the signatures are determined as:

$$\begin{aligned} a_1 &:= (y_1 \cdot y_2)^{k_1}, \\ a_2 &:= (y_1 \cdot y_2^2)^{k_2}, \end{aligned}$$

and the second parameters as:

$$\begin{aligned} b_1 &:= k_1^{-1}(H(M_1) - x \cdot a_1) \bmod q, \\ b_2 &:= k_2^{-1}(H(M_2) - x \cdot a_2) \bmod q. \end{aligned}$$

for  $k_1$  and  $k_2$  chosen at random.

- Regular ElGamal signatures can then be computed as

$$(a_1, (x_1 + x_2)^{-1} \cdot b_1 \bmod q) \quad \text{and} \quad (a_1, (x_1 + 2x_2)^{-1} \cdot b_2 \bmod q)$$

Recall that  $q$  is a big prime number. Hence the probability that  $x_1 + x_2$  or  $x_1 + 2x_2$  is not invertible is negligible.

- After revealing both signatures it is possible to get both  $x_1$  and  $x_2$  by solving a simple set of equations.

If  $x_1$  and  $x_2$  belong to different parties, then the parties have to betray their private keys to each other to accomplish the protocol. Also, a single signature betrays either  $x_1 + x_2$  or  $x_1 + 2x_2$ . Below we present a version of that scheme that solves mentioned problems.

**Signature Creation.** The signer  $S$  chooses factors  $r_1, r_2$  at random and sends them as encrypted messages to verifiers  $V_{d,1}$  and  $V_{d,2}$ , respectively. Besides,  $S$  computes

$$a_1 := ((y_1^{r_1}) \cdot (y_2^{r_2}))^{k_1}$$

and

$$a_2 := ((y_1^{r_1}) \cdot (y_2^{r_2})^2)^{k_2}.$$

The numbers  $b_1$  and  $b_2$  are obtained as before.

Note that to encrypt a random  $r_i$  the sender  $S$  might use ElGamal cryptosystem and  $V_{d,i}$ 's public key  $y_i$ . Thus a complete dds of  $M_i$  takes the form

$$((g^{\ell_i}, (y_i^{\ell_i} + r_i) \cdot y_i^{\ell_i}), a_i, b_i),$$

where  $\ell_i \in \{1, 2, \dots, q-1\}$  is chosen uniformly at random. Due to addition  $y_i^{\ell_i} + r_i$  performed before encryption we avoid leaking some information about  $r_i$ : if the ciphertext would take the form of  $(g^{\ell_i}, r_i \cdot y_i^{\ell_i})$  then it would be easy to check whether  $r_i \in \langle g \rangle$  by examining if  $(r_i \cdot y_i^{\ell_i})^q = 1$ .



**Signature Transformation.** Verifiers  $V_{d,1}$  and  $V_{d,2}$  send each other the numbers  $r_1 \cdot x_1$  and  $r_2 \cdot x_2$ . Then each of them can retrieve

$$\begin{aligned}\hat{b}_1 &:= (r_1 \cdot x_1 + r_2 \cdot x_2)^{-1} \cdot b_1 \bmod q, \\ \hat{b}_2 &:= (r_1 \cdot x_1 + 2 \cdot r_2 \cdot x_2)^{-1} \cdot b_2 \bmod q.\end{aligned}$$

Having parameters  $b_1, b_2, \hat{b}_1, \hat{b}_2, r_1, r_2$  the signer can easily retrieve  $x_1$  and  $x_2$ .

**The Case of Multiple Keys.** The idea from the previous subsection can be easily extended for getting  $n$  private keys of verifiers  $V_{d,1}, V_{d,2}, \dots, V_{d,n}$  after publishing  $n$  signed messages. Indeed, it is enough to construct  $a_i$  as

$$\begin{aligned}(y_1^{r_1} \cdot y_2^{r_2} \cdot \dots \cdot y_{i-1}^{r_{i-1}} \cdot y_i^{r_i} \cdot y_{i+1}^{r_{i+1}} \cdot \dots \cdot y_n^{r_n})^{k_i} &\text{ for } i = 1, \\ (y_1^{r_1} \cdot y_2^{r_2} \cdot \dots \cdot y_{i-1}^{r_{i-1}} \cdot y_i^{2r_i} \cdot y_{i+1}^{r_{i+1}} \cdot \dots \cdot y_n^{r_n})^{k_i} &\text{ for } i > 1.\end{aligned}$$

It is easy to see that as before the signer can solve a set of  $n$  independent equations mod  $q$  and get all values  $x_1, x_2, \dots, x_n$ .

**A Single Signature for Several Private Keys.** Let us note that by merging techniques presented in this section we can enforce the designated verifier  $V_d$  to reveal  $n$  private keys when he shows a single signature to another parties. Namely, he signs  $M \otimes R_1 \otimes \dots \otimes R_n$  for random parameters  $R_1, \dots, R_n$  in a regular way (instead of  $M$ ). Moreover he also signs  $R_i$  in a manner that reveals  $x_i$  for  $i \leq n$ .

## 5 Threshold Schemes

In this section we consider dds( $k, n$ ) schemes. Suppose that a signer  $S$  sends a **group** of  $n$  dds'es to a designated Verifier  $V_d$ , and allows him to publish regular ElGamal signatures of at most  $k - 1$  messages. If more signatures will be published, then this would reveal the private key  $x_1$  of  $V_d$ . Moreover, the scheme below might be applied to the scenario in which instead of  $V_d$ 's public key the signature of some message  $M_1$  signed by  $V_d$  becomes known. For the sake of brevity we shall focus our attention only on the first scenario.

Let us begin with a simple case of two messages, namely the dds(2,2) case:

**Signing.**  $V_d$  sends to the signer  $S$  a value  $R = g^r$ . Than  $S$  generates dds'es of messages  $M_1$  and  $M_2$  as follows:

$$\begin{aligned}a_1 &:= (y_1 \cdot R)^{k_1}, \\ b_1 &:= k_1^{-1} \cdot (H(M_1) - x \cdot a_1) \bmod q, \\ a_2 &:= (R)^{k_2}, \\ b_2 &:= k_2^{-1} \cdot (H(M_2) - x \cdot a_2) \bmod q.\end{aligned}$$

Finally  $S$  sends to  $V_d$  the dds signatures  $(a_1, b_1), (a_2, b_2)$ .

**Transformation.**  $V_d$  derives the signatures:

$$\begin{aligned}\hat{a}_1 &:= a_1, \\ \hat{b}_1 &:= (x_1 + r)^{-1} \cdot b_1 \bmod q, \\ \hat{a}_2 &:= a_2, \\ \hat{b}_2 &:= r^{-1} \cdot b_2 \bmod q,\end{aligned}$$

and checks validity of the signatures  $(\hat{a}_1, \hat{b}_1)$ ,  $(\hat{a}_2, \hat{b}_2)$  for the messages  $M_1$ ,  $M_2$ . We see at once that

$$\begin{aligned}b_1/\hat{b}_1 &= x_1 + r \bmod q, \\ b_2/\hat{b}_2 &= r \bmod q.\end{aligned}$$

Consequently, if  $V_d$  publishes both signatures, then anyone who has access to  $b_1$  and  $b_2$  would be able to retrieve  $V_d$ 's private key  $x_1$ .

Our solution for  $\text{dds}(k, n)$  is based on a technique similar to the method used for the first time in [7].

**Preliminaries and Key Setup.** At the beginning a designated verifier  $V_d$  chooses random values  $f_1, \dots, f_{k-1}$  that constitute a polynomial  $F \in \mathbb{F}_q[X]$ , namely:  $F(y) = x_1 + f_1 y + f_2 y^2 + \dots + f_{k-1} y^{k-1}$ . Then  $V_d$  sends to the signer  $S$  the values  $g_j = g^{f_j}$  for  $j \leq k-1$ .

For  $i \leq n$  signer  $S$  computes:

$$z_i := y_1 \cdot g_1^i \cdot g_2^{i^2} \cdot \dots \cdot g_{k-1}^{i^{k-1}} = g^{F(i)}.$$

**Signing.** A dds of a message  $M_i$  addressed to  $V_d$  is generated as follows:

$$\begin{aligned}a_i &:= (z_i)^{k_i}, \\ b_i &:= k_i^{-1} \cdot (M_i - x \cdot a_i) \bmod q.\end{aligned}$$

Now  $S$  can send  $(a_i, b_i)$  as the dds of the message  $M_i$ .

**Signature Transformation.**  $V_d$  puts:

$$\begin{aligned}\hat{a}_i &:= a_i, \\ \hat{b}_i &:= b_i \cdot F(i)^{-1} \bmod q.\end{aligned}$$

At the last step  $V_d$  checks validity of the signature  $(\hat{a}_i, \hat{b}_i)$  for the message  $M_i$  as for the regular ElGamal scheme.

**Revealing Signature to Other Verifiers.** It is easy to see that having both  $\hat{b}_i$  and  $b_i$  one can retrieve  $F(i)$ . Using standard techniques having  $k$  different values of polynomial  $F$  one can easily get  $F(0) = x_1$ . Moreover, any  $k-1$  values reveal no information about  $x_1$ .

Note that if  $V_d$  had taken part in  $\ell$   $\text{dds}(k_i, n_i)$  schemes, for  $i \in \{1, 2, \dots, \ell\}$ , and has published  $\sum_{i=1}^{\ell} (k_i - 1)$  regular ElGamal signatures, then any attacker can collect  $(\sum_{i=1}^{\ell} k_i) - \ell$  linear equations with  $(\sum_{i=1}^{\ell} k_i) - \ell + 1$  unknowns. Thus  $x_1$  is still secure.

## 6 Applications

### Privacy Protection Area

1. The basic dds scheme can be used as an authentication scheme. In this scenario Alice authenticates herself to Bob with a standard signature software - installing extra systems is unnecessary. The only thing she has to do is to change the generator  $g$  to  $y_1$  in some configuration file. Note that the authentication protocol has the desirable feature that its transcript cannot be shown to a third party, therefore it automatically protects Alice's privacy.
2. A dds scheme can be used for preserving privacy of purchases. Alice can signed a special statemant, that she bought a certain service, so she can be billed later. But if the signature was created just like in 3.1 or 3.2 way, Bob as a service provider will not be interested to prove everybody that Alice was using his service. However in some special situations (if Alice do not want to pay) Bob will be able to prove it in the court.

### Business Transactions

1. Scheme from section 3.2 can be used for solving the problem of simultaneous revealing signatures of Bob and Alice. Assume that Bob has to present Paul an Alice's signature of  $M$  together with its own signature of  $M_1$ . Then Alice construct an appropriate dds. Once Bob shows the signature for  $M$ , a signature for  $M_1$  can be automatically deduced.
2. The multi-key scheme can be applied by Alice to business negotiations. Alice can give her negotiators a set of  $n$  dds-signed documents, each of them to be used in a different situation. Then the negotiators cannot use any two of the signed documents, even if they wish to do so.
3. A Threshold scheme may be used by Alice in the situation, when she wants her business representative to use some but not all documents signed by her. In the simplest case Alice for two different situations can prepar two different statements. The point is that Bob cannot present both of them without revealing his private key (or his signature of some unprofitable for him messeage).

## 7 Conclusions and Open Problems

Techniques presented provide simple conversions from ElGamal like schemes to schemes where a dedicated verifier cannot show a signature to third parties. Unlike in the previous schemes (where it is hard to convince a third party that such a signature is valid) the verifier loses his own secrets when exhibiting the signature.

The technique developed can be used in the cases when a signature is a kind of a token that should be presented to get some effect. However, as already mentioned the dedicated verifier may provide a zero-knowledge proof that it has a certain signature without presenting it. In many legal systems it would suffice

in the court to present such a proof in order to derive legal consequences of the signed document. Designing a dds system for standard digital signatures so that no such a zero-knowledge proof is possible is a challenging problem.

## References

1. Chaum, D., Antwerpen, H.V.: Undeniable signatures. In: CRYPTO'1989. Volume 435 of Lecture Notes in Computer Science. (1989) 212–216
2. Chaum, D.: Designated confirmer signatures. In: EUROCRYPT'1994. Volume 950 of Lecture Notes in Computer Science. (1994) 86–91
3. Jakobsson, M., Sako, K., Impagliazzo, R.: Designated verifier proofs and their applications. In: EUROCRYPT'1996. Volume 1070 of Lecture Notes in Computer Science. (1996) 143–154
4. Krawczyk, H., Rabin, T.: Chameleon signatures. In: NDSS. (2000)
5. Chaum, D., Evertse, J.H., van de Graaf, J.: An improved protocol for demonstrating possession of discrete logarithms and some generalizations. In: EUROCRYPT'1987. Volume 304 of Lecture Notes in Computer Science. (1987) 127–141
6. Chaum, D., Evertse, J.H., van de Graaf, J., Peralta, R.: Demonstrating possession of a discrete logarithm without revealing it. In: CRYPTO'1986. Volume 263 of Lecture Notes in Computer Science. (1986) 200–212
7. Feldman, P.: A practical scheme for non-interactive verifiable secret sharing. In: FOCS'1987. (1987) 427–437

# Security Analysis and Improvement for Key Issuing Schemes in ID-Based Cryptography

Saeran Kwon and Sang-Ho Lee

Dept. of Computer Science and Engineering,  
Ewha Womans University, Seoul, Korea  
`sranie@ewhain.net`

**Abstract.** In this paper, we analyze some serious weakness for security of existing key issuing schemes in identity(ID)-based cryptosystems which were proposed in order to eliminate key escrow property and remove the need of secure channel, and describe some attacks for them. In addition, we present the improved key issuing protocols for each scheme with weakness, which can resist the attack and overcome key escrow problem.

## 1 Introduction

In a traditional public key infrastructure, a main difficulty is to guarantee a user's public key is indeed linked to the right owner to hold the corresponding private key. A simple solution but suffering from management is that a user's public key is assured with certificate issued by a trusted certification authority (CA), essentially a signature by the CA on a public key.

Shamir introduced the concept of ID-based cryptography in 1984 [11], and Boneh and Franklin presented the first fully practical and secure ID-based encryption scheme (IBE) in 2001 [3]. In ID-based cryptography, an entity's public key is derived directly from its identity information such as name, E-mail address and IP address, etc. The corresponding private key for the entity is generated by a trusted third party called key generation center (KGC) and is handed to each user through a secure channel. The direct derivation of public keys in ID-based cryptography removes the need for certificates and some of the problems associated with them. However there are some disadvantages such that a user's private key is known to the authority, that is, the key escrow property, and also that the KGC must send user's private key over a secure channel, making private key distribution difficult.

To solve the key escrow problem in ID-based cryptography, several schemes have been proposed. One of approaches is to assume multiple KGCs. Schemes in Boneh and Franklin [3], Chen et al. [4], Hess [8] distributed a role of a master key of one KGC to multiple authorities. However, in such an approach, during a user's private key issuing, all multiple KGCs have to check user's identity independently, which is quite a burden.

Another approaches are by using some user-chosen secret information. In 2003, Gentry [6] proposed a certificate-based encryption (CBE) scheme that does not

require a secure channel for the delivery of the key issued by the trust authority called certification authority (CA). There, the key issued by CA is in fact the up-to-date certificate for his public key generated by the user with user-chosen secret information and it is only one part of a user's full decryption key. The other part of the full decryption key is user's personal secret key built by himself using the same secret information of his public key. Since the CA does not know the other part of full decryption key, key escrow problem related with the trust authority is avoided.

Al-Riyami and Paterson [1] also proposed a new scheme in 2003 called certificateless public key cryptography (CL-PKC), in which entity's secret key is constructed by scalar multiplying partial private key issued through secure channel from KGC by user-chosen secret value, while entity's public key is constructed by each user using the same secret value, which needs no certification by the trust authority. Certificateless public key cryptography (CL-PKC) shares some common features with the self-certificated keys of [7] and Gentry's proposed CBE [6] from the viewpoint of combining some user-chosen secret information with trust authority's master key for key extraction and offering implicit certification for a public key. However, both CBE and CL-PKC are not ID-based since their public keys are not determined exclusively by publicly known information of the user's identity.

In [9], a new secure ID-based key issuing protocol is proposed, in which a user's private key is issued by a single KGC and its privacy is protected by multiple key privacy authorities(KPAs). In the protocol, only the KGC checks a user's identity and then issues a user's partial private key through a blinded manner, whereas other KPAs just contribute service for key privacy by providing their signature sequentially in a blinded manner. The blinding technique to provide secure channel between users and authorities is used in pairing based cryptography, assuming the hardness of the bilinear Diffie-Hellman problem(BDHP)(Boneh & Franklin)[3]. Finally, in key retrieving stage, only a legitimate user who has the secure blinding parameter can unblind it and retrieve the real private key.

The scheme in [9] distributes the roles of user identification and key securing into the KGC and KPAs respectively, as effect of which the cost of user identification is reduced. Also the scheme provides a secure channel by blinding parameter between a user and the KGC or KPAs. However, there exists some weakness. First, if the KGC want to get a user's private key, it is possible because the KGC can impersonate any user easily when to request key privacy service to the KPAs and furthermore, impersonation of any user by the KGC is not detectable either, which means the scheme can not overcome the key-escrow property for the KGC of ID-based cryptography, contrary to the claim in [9]. Next, the scheme is vulnerable to denial-of-service(DoS) attacks since KPAs can not distinguish an entity's legitimate key securing request from an adversary's malicious request to disrupt service by overloading it.

Recently, in 2005, another key issuing scheme in ID-based cryptography is proposed in [5], which is similar to [9] in that a user's private key is issued by a key generation center(KGC) and its privacy is protected by multiple key privacy authorities(KPAs) and that only the single KGC check user's identity

and issues a user's partial private key, other KPAs just cooperate key privacy service, through blinding technique to avoid the necessity of secure channel. On account of similar structure with [9], it has also the same weakness as the one in [9] as follows. If the KGC intends to extract a user's private key, it is possible even without being detected like [9]. Accordingly, neither the scheme can overcome the key-escrow property for the KGC of ID-based cryptography. The scheme is also vulnerable to denial-of-service(DoS) attacks during key privacy service like the scheme in [9], since KPAs can not distinguish entity's legitimate key securing request from an adversary's malicious request.

In this paper, we present that the key issuing protocol proposed in [9] does not solve key escrow problem through concrete attack by the KGC with the intention of obtaining any user's private key without being detected in front of multiple KPAs' key privacy service, and illustrate a malicious adversary's denial-of-service(DoS) attack for KPAs. Next, we propose an improved key issuing protocol of [9] which can resist such attacks and prevent efficiently key escrow problem. Also we show that the key issuing protocol proposed in [5] as well does not solve key escrow problem, and is vulnerable to DoS attacks, through concrete attacks by the KGC and an adversary, respectively. similarly, we propose an improved key issuing protocol of [5].

The rest of this paper is organized as follows. In section 2, we review the key issuing scheme in ID-based cryptography proposed by Lee et al in [9], analyze its security through weakness and attack, and present some improvements on the scheme. In section 3, we also review the key issuing scheme in ID-based cryptography proposed by Gangishetti et al in [5], analyze its security through weakness and attack, and present some improvements on the scheme. Finally, we conclude the paper with some remarks.

## 2 Key Issuing in ID-Based Cryptography by Lee et al.

### 2.1 Review

The key issuing protocol proposed by Lee et al. [9] consists of five phases namely **System Setup**, **System Public Key Setup**, **Key Issuing**, **Key Securing** and **Key Retrieving**.

#### Stage 1. System setup(by KGC)

The KGC specifies two groups  $G_1$  and  $G_2$ , a bilinear map  $e : G_1 \times G_1 \longrightarrow G_2$  between them and an arbitrary point  $P \in G_1$  of order  $q$ . It also specifies two hash functions;  $H_1 : \{0, 1\}^* \longrightarrow G_1$  (extract a point on  $G_1$  from ID),  $H_2 : G_2 \longrightarrow Z_q^*$ . The KGC picks its master key  $s_0 \in Z_q^*$  at random and computes its public key  $P_0 = s_0 P$ . Then it publishes description of the groups  $G_1, G_2$ ,  $P$ , the bilinear map  $e$ , hash functions  $H_1, H_2$ , and the public key  $P_0$ .

#### Stage 2. System public key setup(by KPAs)

The  $n$  KPAs establish their key pairs. For all  $i = 1, \dots, n$ ,  $\text{KPA}_i$  chooses its master key  $s_i$  and computes its public key  $P_i = s_i P$ . Then KPAs cooperate

sequentially to compute the system public key  $Y = s_0 s_1 \cdots s_n P$ . More specifically, each  $KPA_i$  computes  $Y'_i = s_i Y'_{i-1}$  for  $i = 1, \dots, n$ , where  $Y'_0 = P_0$ . Then  $Y \equiv Y'_n = s_0 s_1 \cdots s_n P$  is published as the system public key. The correctness of this sequential processes can be verified by  $e(Y'_i, P) \stackrel{?}{=} e(Y'_{i-1}, P_i)$ .

### Stage 3. Key issuing (by KGC and user)

A user with identity  $ID$  chooses a random secret  $x$  and computes a blinding factor  $X = xP$ . He requests the KGC to issue a partial private key by sending  $X$  and  $ID$ . Then the KGC issues a blinded partial private key as follows.

- Checks the identification of the user.
- Computes the public key of the user as  $Q_{ID} = H_1(ID, KGC, KPA_1, \dots, KPA_n)$ .
- Computes a blinded partial private key as  $Q'_0 = H_2(e(s_0 X, P_0)) s_0 Q_{ID}$ .
- Computes the KGC's signature on  $Q'_0$  as  $Sig_0(Q'_0) = s_0 Q'_0$ .
- Sends  $Q'_0$  and  $Sig_0(Q'_0)$  to the user.

Here  $H_2(e(s_0 X, P_0))$  is a blinding factor; a secure channel between the user and the KGC. The user can unblind it using his knowledge of  $x$ , since

$$H_2(e(s_0 X, P_0)) = H_2(e(s_0 x P, P_0)) = H_2(e(P_0, P_0)^x).$$

### Stage 4. Key securing (by user and KPAs)

The user requests  $KPA_i$  ( $i = 1, \dots, n$ ) sequentially to provide key privacy service by sending  $ID$ ,  $X$ ,  $Q'_{i-1}$ , and  $Sig_{i-1}(Q'_{i-1})$ . Then  $KPA_i$

- Checks  $e(Sig_{i-1}(Q'_{i-1}), P) \stackrel{?}{=} e(Q'_{i-1}, P_{i-1})$ .
- Computes  $Q'_i = H_2(e(s_i X, P_i)) s_i Q'_{i-1}$  and  $Sig_i(Q'_i) = s_i Q'_i$ .
- Sends  $Q'_i$  and  $Sig_i(Q'_i)$  to the user.

The user proceeds this process to  $KPA_n$ . Finally he receives  $Q'_n = H_2(e(s_n X, P_n)) s_n Q'_{n-1}$ .

### Stage 5. Key retrieving (by user)

The user retrieves his private key  $D_{ID}$  by unblinding  $Q'_n$  as follows.

$$D_{ID} = \frac{Q'_n}{H_2(e(P_0, P_0)^x) \cdots H_2(e(P_n, P_n)^x)} = s_0 s_1 \cdots s_n Q_{ID}$$

The user can verify the correctness of his private key by  $e(D_{ID}, P) \stackrel{?}{=} e(Q_{ID}, Y)$ .

## 2.2 Weakness and Attack

**Key Escrow Problem.** In key issuing protocol of [9], when each  $KPA_i$  ( $i = 1, \dots, n$ ) receive  $ID$ ,  $X$ ,  $Q'_{i-1}$ , and  $Sig_{i-1}(Q'_{i-1})$  for sequential key privacy service from a user, it does not check the identification of the user but only checks such a validation that  $Q'_{i-1}$  is really signed by  $KPA_{i-1}$  with the signature  $Sig_{i-1}(Q'_{i-1})$ .



Accordingly, if the KGC replaces  $X$  with a blind factor  $Z = zP$  by a secret value  $z$  of its choice and requests sequentially key privacy service for the identity  $ID$  from  $KPA_i$   $i = 1, \dots, n$ , it achieves at the end of the protocol

$$Q'_n = H_2(e(s_0X, P_0))H_2(e(s_1Z, P_1)) \cdots H_2(e(s_nZ, P_n))s_0s_1 \cdots s_nQ_{ID}.$$

Then, because the KGC can compute the value  $H_2(e(s_0X, P_0))$  with its master key  $s_0$ , it can get the secret key  $D_{ID}$  of the identity  $ID$ , even though it does not have knowledge of the user's secret value  $x$ . In fact, during providing the privacy service sequentially, since each  $KPA_i$  authenticates neither user's identification nor a blinding factor, and does not use any information of user's identity  $ID$  when to compute  $Q'_i$  and  $Sig_i(Q'_i)$  from  $Q'_{i-1}$  presented, the KGC need not even masquerade as the user  $ID$  in front of  $KPA_i$ . It is sufficient that the KGC simply requests key privacy service of  $KPA_i$  ( $i = 1, \dots, n$ ) by sending a string  $ID'$  of any identity, any blinding factor  $Z = zP$  the secret value  $z$  of which it chooses at random, along with  $Q'_{i-1}$ , and  $Sig_{i-1}(Q'_{i-1})$  ( $i = 1, \dots, n$ ).

As we see, even if the scheme is assumed to have a trust level in the KGC such as the KGC does not replace a user's partial private key with one of its choice, the KGC can obtain a user's secret key proceeding as above scenario. Consequently, the scheme does not solve the key escrow problem.

**Denial-of-Service Attack.** On the other hand, the protocol has another weakness which is highly vulnerable to denial-of-service attacks. In fact, its signature  $Sig_i(Q'_i)$  made by  $KPA_i$  for a combined partial private key  $Q'_i$  does not really provide an assurance of such a claim that  $Q'_i$  is certified by  $KPA_i$ .

Let's consider a threat we might infer. Suppose an adversary tries to degrade a normal privacy service of KPAs. Here is how he execute this. He choose a fake value  $a \in Z_q^*$  randomly and he compute  $Q'' = aP$ . Then, through a scalar multiple of  $P_{i-1}$ , that is a public key of  $KPA_{i-1}$ , by the fake value  $a$ , he can make a forged signature  $Sig_{i-1}(Q'')$  of  $Q''$  to pretend to be signed by  $KPA_{i-1}$ , because

$$Sig_{i-1}(Q'') = s_{i-1}Q'' = s_{i-1}aP = aP_{i-1}.$$

With  $Q''$ ,  $Sig_{i-1}(Q'')$ , any identity  $ID''$ , and any blinding factor  $Z \in G_1$ , he request key privacy service to the target key privacy authority among  $KPA_i$  ( $i = 1, \dots, n$ ), here, let it be  $KPA_i$ . Then in stage 4 of the original scheme,  $KPA_i$  checks only whether  $e(Sig_{i-1}(Q''), P) \stackrel{?}{=} e(Q'', P_{i-1})$  or not. In fact, we see that

$$e(Sig_{i-1}(Q''), P) = e(aP_{i-1}, P) = e(s_{i-1}aP, P) = e(aP, s_{i-1}P) = e(Q'', P_{i-1}).$$

Consequently, it is difficult for KPAs to discern legitimate key privacy service request from malicious service-overloading disturbance, because neither a proper blinded partial private key  $Q'_i$  nor the forged key  $Q''$  is meaningful message.

**Complexity of System Public key Setup.** In addition, at the stage 2, when KPAs set up the system public key  $Y = s_0s_1 \cdots s_nP$ , correctness of the sequential process is verified by the equation  $e(Y'_i, P) \stackrel{?}{=} e(Y'_{i-1}, P_i)$ . However if

$KPA_{i+1}$  is about to verify correctness of  $Y'_i$  sent to it, it must have  $Y'_{i-1}$  assured of its integrity in advance. But, in order to verify integrity of  $Y'_{i-1}$ ,  $Y'_{i-2}$  must be as well assured of its integrity in advance. Consequently, in order to verify correctness of  $Y'_i$ ,  $KPA_{i+1}$  should succeed to check  $e(Y'_k, P) \stackrel{?}{=} e(Y'_{k-1}, P_k)$  for all  $k = 1, \dots, i$ . Thus, a more efficient way is required. For example, there might be an approach such as using secure signature scheme to authenticate integrity of each transmission and fixing one trusted authority to verify entire process among KGC or KPAs.

### 2.3 Improvement

We present some improvements; adding the so-called **key privacy service issue-list (KPSIL)** held by KPAs to original scheme and employing short signature scheme by Boneh, Lynn and Shacham [2] using gap Diffie-Hellman (GDH) group [10] whose security is based on the hardness of computational Diffie-Hellman problem (CDHP). Some improvements in key issuing stage and key securing stage are specifically described respectively, as follows.

**Key Issuing Stage by KGC and User.** In the stage of Key issuing, when the KGC computes its signature for a user's blinded partial private key  $Q'_0$ , we require the user's identity  $ID$ , the blinding factor  $X$  and a valid period  $T$  of  $Q'_0$  be appended in the message to be signed, in order to prevent an adversary (or the KGC) from altering a user's blinding factor in next key privacy service stage. So, the KGC's signature on  $Q'_0$  is such as  $Sig_0(Q'_0) = s_0 H_1(ID, X, T, Q'_0)$ , where the signature scheme by Boneh et al. [2] is applied on the message  $\{ID, X, T, Q'_0\}$ . The KGC sends  $T$ ,  $Q'_0$  and  $Sig_0(Q'_0)$  to the user.

**Key Securing Stage by KPAs and User.** We require the KPAs to maintain *key privacy service issue-list* which records key privacy service of users by a user's identity  $ID$ , a blinding factor  $X$  and a valid period  $T$  of a partial private key, if the user's request is valid. We assume that a user should request key privacy service to KPAs with same blinding factor during the set period if once a valid period is set for a partial private key in key issuing stage. If the KGC reissues a partial private key of a target user through a blinding factor of its choice in order to get the a user's private key, it can be easily checked by a valid request record with a different blinding factor for same user in KPAs' *key privacy service issue-list*. More precisely, since only the KGC is able to produce a signature  $Sig_0(Q'_0) = s_0 H_1(ID, X, T, Q'_0)$ , for the message  $\{ID, X, T, Q'_0\}$  by virtue of the security of short signature scheme [2], the existence of different blinding factors certified by a signature of the KGC for one user is a proof that the KGC has cheated. It means that our improved scheme reaches the same trust level 3, the frauds of the authority to be detectable, as traditional PKIs [7]. We describe this stage as follows. A user requests the  $KPA_i$  to provide key privacy service by sending  $\langle ID, X, T, Q'_{i-1}, Sig_{i-1}(Q'_{i-1}) \rangle$  sequentially for  $i = 1, \dots, n$ . Then the  $KPA_i$

- Computes  $H_1(ID, X, T, Q'_{i-1})$ .
- Checks  $e(Sig_{i-1}(Q'_{i-1}), P) \stackrel{?}{=} e(H_1(ID, X, T, Q'_{i-1}), P_{i-1})$ .
- Includes  $\langle ID, X, T \rangle$  in its *KPSIL* if the previous step is satisfied, the identity ID is not included yet in the *KPSIL* and the valid period  $T$  is currently available. But if the identity ID exist already with a different blinding factor, the  $KPA_i$  stops key privacy service, gives notice of this happening to the KGC and the user, and confirms which blinding factor is right. When the valid period  $T$  expires,  $\langle ID, X, T \rangle$  is discarded from the *KPSIL*.
- Computes  $Q'_i = H_2(e(s_i X, P_i))s_i Q'_{i-1}$  and its signature such as  $Sig_i(Q'_i) = s_i H_1(ID, X, T, Q'_i)$ .
- Sends  $Q'_i$ , its signature  $Sig_i(Q'_i)$  to a user.

**A further Remark.** Note that at above third step, the  $KPA_i$  should consider various situations. A legitimate user may not receive  $Q'_i$  or  $Sig_i Q'_i$  because of network problem or attack. Then, a user will again request a service with the same blinding factor  $X$  and same  $T$  as before if current time is in a valid period  $T$ . Hence, though the same  $\langle ID, X, T \rangle$  already exist in *KPSIL* of the  $KPA_i$ , the  $KPA_i$  performs a service. However, there might be a replay attack. If much more requests happen beyond normal scope by the same ID, X, and T, then the  $KPA_i$  can test that a user really knows a secret value  $x$  of  $X = xP$  by challenge-and-response protocol as follows, where  $H$  is a hash function such that  $H : G_2 \longrightarrow \{0, 1\}^l$ , where  $l$  is the length of a plaintext message.

- $KPA_i$  chooses a challenge,  $r$ , which is a random  $l$ -bit string.  $KPA_i$  sends  $r \oplus H(e(s_i X, P_i))$  to the user ID.
- The user can compute  $H(e(s_i X, P_i))$  if he is a legitimate user who knows secret value  $x$  because  $e(s_i X, P_i) = e(P_i, P_i)^x$ , hence can obtain  $r$ . Again he computes  $H(e(X, P_i)^x)$  and responds with  $r \oplus H(e(X, P_i)^x) \stackrel{let}{=} r'$  to  $KPA_i$ .
- $KPA_i$  computes  $r' \oplus H(e(X, X)^{s_i}) \stackrel{let}{=} r''$  and verifies that  $r \stackrel{?}{=} r''$ .

In fact  $r'' = r' \oplus H(e(X, X)^{s_i}) = r \oplus H(e(X, P_i)^x) \oplus H(e(X, X)^{s_i}) = r$  since  $e(X, P_i)^x = e(X, xP_i) = e(X, xs_i P) = e(X, xP)^{s_i} = e(X, X)^{s_i}$ .

If we assume that the bilinear Diffie-Hellman problem(BDHP), which asks to compute  $e(P, P)^{abc}$  for a given  $(P, aP, bP, cP)$ , is infeasible, this protocol will securely test whether a request is from a legitimate user or from a DoS attacker.

### 3 Key Issuing in ID-Based Cryptography by Gangishetti et al.

#### 3.1 Review

Similar to [9], the key issuing protocol proposed by Gangishetti et al. [5] consists of five stages namely **System Setup**, **System Public Key Setup**, **Key Issuing**, **Key Securing** and **Key Retrieving**.

**Stage 1. System setup(by KGC)**

The KGC specifies two groups  $G_1$  and  $G_2$  of prime order  $q$ , a bilinear map  $e : G_1 \times G_1 \rightarrow G_2$  between them and hash function  $H : \{0,1\}^* \rightarrow G_1$ . Let  $P \in G_1$  be a generator of  $G_1$ . The KGC picks a master key  $s_0 \in Z_q^*$  at random and computes its public key  $P_0 = s_0P$ .

**Stage 2. System public key setup(by KPAs and KGC)**

The  $n$  KPAs establish their key pairs. For all  $i = 1, \dots, n$ ,  $\text{KPA}_i$  chooses its secret key  $s_i$  and computes its public key  $P_i = s_iP$ . Then each  $\text{KPA}_i$  computes its share  $Y'_i = s_iP_0$  and send it to the KGC. The KGC computes the system public key as  $Y = \sum_{i=1}^n Y'_i = s_0(s_1 + s_2 + \dots + s_n)P$ . The KGC publishes the system parameters  $\langle G_1, G_2, q, e, H, n, P, P_0, P_1, \dots, P_n, Y \rangle$ . The correctness of the system public key can be verified by checking the equality  $e(Y, P) = e(\sum_{i=1}^n P_i, P_0)$ .

**Stage 3. Key issuing (by user and KGC)**

A user with identity  $ID$  chooses a random secret  $r \in Z_q^*$  and computes the user's public key  $Q_{ID} = H(ID)$ ,  $R = rP$  and  $D_{ID} = rQ_{ID}$ . He requests the KGC to issue a partial private key by sending  $ID$ ,  $R$  and  $D_{ID}$ . Then the KGC issues a blinded partial private key as follows.

- Checks validity of the user's  $ID$ .
- Computes the public key of the user  $Q_{ID} \in G_1$ .
- Validates the parameters  $(D_{ID}, R)$  by checking the equality  $e(D_{ID}, P) = e(Q_{ID}, R)$ .
- Computes a blinded partial private key as  $Q'_{0ID} = s_0D_{ID}$ .
- Sends  $Q'_{0ID}$  to the user over public channel.

Here,  $r$  is a blinding factor that eliminates the need of secure channel between the user and the KGC. The user can verify the issued blinded partial private key by checking the equality  $e(Q'_{0ID}, P) = e(D_{ID}, P_0)$ .

**Stage 4. Key securing (by user and KPAs)**

The user requests  $\text{KPA}_i$  ( $i = 1, \dots, n$ ) to provide key privacy service by sending  $ID$ ,  $R$ ,  $Q'_{0ID}$ , and  $D_{ID}$ . Then  $\text{KPA}_i$  does the follows.

- Checks  $e(Q'_{0ID}, P) = e(D_{ID}, P_0)$  to validate  $Q'_{0ID}$ .
- Computes  $Q'_{iID} = s_iQ'_{0ID}$ .
- Sends  $Q'_{iID}$  to the user over public channel.

**Stage 5. Key retrieving (by user)**

The user retrieves his blinded private key  $S'_{ID}$  by combining all blinded private key components issued by the KPAs. Then he unblinds  $S'_{ID}$  and gets his private key  $S_{ID}$  as follows.

$$S'_{ID} = \sum_{i=1}^n Q'_{iID}. \quad S_{ID} = r^{-1}S'_{ID} = s_0(s_1 + s_2 + \dots + s_n)Q_{ID}.$$

The user can verify the correctness of his private key by  $e(S_{ID}, P) \stackrel{?}{=} e(Q_{ID}, Y)$ .

### 3.2 Weakness, Attack and Improvement

The key issuing scheme proposed in [5] has also the same weakness as the one in [9] on account of a similar structure with [9]. We show the scheme does not overcome the key-escrow property for the KGC through the similar way, and it is as well vulnerable to denial-of-service(DoS) attack, as follows.

**Denial-of-Service Attack.** Suppose an adversary tries to degrade a normal privacy service of KPAs. The adversary chooses a fake value  $a \in Z_q^*$  randomly and compute  $D_{ID} = aP$ . Then, through a scalar multiple of a public key of KGC  $P_0$  by the fake value  $a$ , he can make a forged blinded partial private key  $Q'_{0ID}$  because  $Q'_{0ID} = s_0 D_{ID} = s_0 aP = aP_0$ . With any identity ID, any blinding factor R,  $Q'_{0ID}$  and  $D_{ID}$ , he requests key privacy service to the  $KPA_i$  for some  $1 \leq i \leq n$ . Then in stage 4 of the original scheme,  $KPA_i$  checks only whether  $e(Q'_{0ID}, P) = e(D_{ID}, P_0)$  or not. In fact, we see that

$$e(Q'_{0ID}, P) = e(aP_0, P) = e(s_0 aP, P) = e(aP, s_0 P) = e(D_{ID}, P_0).$$

Consequently, it is difficult for  $KPA_i$  to discern legitimate key privacy service request from malicious service-overloading disturbance.

To fix this problem, we require for each  $KPA_i$  to check the equality  $e(D_{ID}, P) = e(Q_{ID}, R)$  in the stage 4. It makes the scheme resist DoS attack by the hardness of the Computational Diffie-Hellman Problem to ask to compute  $abP$  for given  $(P, aP, bP)$ . Moreover, it provides the linkage between  $Q'_{0ID}$ , ID and R. But it needs  $2n$  times more pairing operations in key securing stage, when to be compared with the original scheme. To reduce times of pairing operations, we can redefine the blinding factor as  $R = rP_0$  with a user-chosen random secret  $r$ . In this case, the equality  $e(D_{ID}, P_0) = e(Q_{ID}, R)$  must be checked to validate the parameters  $(D_{ID}, R)$  in key issuing stage. In key securing stage, the  $KPA_i$  checks  $e(Q'_{0ID}, P) = e(D_{ID}, P_0) = e(Q_{ID}, R)$ , which reduces  $n$  times pairing operations against the above way of fixing.

**Key Escrow Problem.** When each  $KPA_i$  ( $i = 1, \dots, n$ ) receive ID, R,  $D_{ID}$  and  $Q'_{0ID}$  for key privacy service from a user, it does not check the identification of the user but only checks that the equality  $e(Q'_{0ID}, P) = e(D_{ID}, P_0)$  to validate  $Q'_{0ID}$ . Accordingly, if the KGC replaces a blind factor R with  $\hat{R} = r'P$  by a secret value  $r'$  of its choice, and requests a key privacy service for the identity ID to each  $KPA_i$   $i = 1, \dots, n$  by sending ID,  $\hat{R}$ ,  $\hat{D}_{ID} = r'Q_{ID}$  and  $\hat{Q}'_{0ID} = s_0 \hat{D}_{ID}$ , each  $KPA_i$  sends  $\hat{Q}'_{iID} = s_i \hat{Q}'_{0ID}$  after checking the equality  $e(\hat{Q}'_{0ID}, P) = e(\hat{D}_{ID}, P_0)$ . Hence, the KGC achieves at the end of the protocol  $\hat{S}'_{ID} = \sum_{i=1}^n \hat{Q}'_{iID}$ . Because the KGC knows the secret value  $r'$ , it can get the secret key  $S_{ID}$  of the identity ID without being detected like [9]. In fact, during providing the privacy service, since each  $KPA_i$  authenticates neither a user's identification nor a blinding factor, and does not use any information of user's identity ID, the KGC need not even masquerade as the user ID in front of  $KPA_i$ . Consequently, the scheme does not solve the key escrow problem.

To overcome the key escrow problem, we require the KPAs to maintain *key privacy service issue-list* which records key privacy service of users by a user's identity  $ID$ , a blinding factor  $R$  and a valid period  $T$  of a blinded partial private key  $Q'_{0ID}$  in the same way as described in improvement of the scheme [5], together with above remedy against DoS attack.

## 4 Conclusion

In this paper, we showed that Lee et al.'s key issuing protocol [9] and Gangishetti et al's [5] can not overcome the key escrow property in identity-based public key cryptography. Also we mounted a denial-of-service(DoS) attack on each scheme. In order to overcome the key escrow property and be secure against DoS attacks described in this paper, we proposed the improved secure key issuing schemes adding so-called *key privacy service issue-list* maintained by KPAs and using secure signature scheme or pairing operations.

## References

1. S. Al-Riyami and K. Paterson, "Certificateless Public Key Cryptography," Advances in Cryptology - Asiacrypt'03, **LNCS 2894**, pp.452-473, Springer-Verlag, 2003.
2. D. Boneh, A. Lynn and H. Shacham, "Short signatures from the Weil pairing," Advances in Cryptology - Asiacrypt'01. **LNCS 2248**, pp.514-532, 2001.
3. D. Boneh and M. Franklin, "Identity Based Encryption from the Weil Pairing," Advances in Cryptology - Crypto'01. **LNCS 2139**, pp.213-229, 2001.
4. L. Chen, K. Harrison, N. P. Smart and D. Soldera, "Applications of Multiple trust authorities in Pairing based Cryptosystems," Proc. of InfraSec'02. **LNCS 2437**, pp.260-275, 2002.
5. Raju Gangishetti, M. Choudary Gorantla and Manik Lal Das, "An Efficient Secure Key Issuing Protocol in ID-Based Cryptosystem," Proceedings of the International Conference on Information Technology: Coding and Computing(ITCC'05), IEEE Computer Society, 2005.
6. C. Gentry, "Certificate-Based Encryption and the Certificate Revocation Problem," Advances in Cryptology - Eurocrypt'03, **LNCS 2656**, pp.272-293, Springer-Verlag, 2003.
7. M. Girault, "Self-Certified Public Keys," Advances in Cryptology - Eurocrypt'91, **LNCS 547**, pp.490-497, Springer-Verlag, 1992.
8. F. Hess, "Efficient Identity based Signature Schemes based on Pairings," Proc. of SAC'02. **LNCS 2595**, pp.310-324, 2003.
9. B. Lee, C. Boyd, E. Dawson, K. Kim, J. Yang and S. Yoo, "Secure Key Issuing in ID-based Cryptography," Proc. of AISW'04. **vol. 32**, pp.69-74, 2004.
10. T. Okamoto and D. Pointcheval, "The gap-problems: a new class of problems for the security of cryptographic schemes," Proc. of PKC'01, **LNCS 1992**, pp.104-118, 2001.
11. A. Shamir, "Identity Based Cryptosystems and Signature Schemes," Advances in Cryptology - Crypto'84. **LNCS 0196**, pp.47-53, 1984.

# A Secure E-Tender Submission Protocol

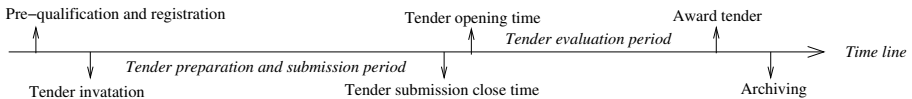
Rong Du, Colin Boyd, and Ernest Foo

Information Security Institute (ISI), Faculty of Information Technology,  
Queensland University of Technology, Australia  
r.du@qut.edu.au, c.boyd@qut.edu.au, e.foo@qut.edu.au

**Abstract.** There is fundamental difference between a simple e-tender box and a traditional physical tender box. Access to the e-tender box has become a private activity in contrast with the public access to a traditional tender box. A significant opportunity is therefore created for malicious business collusion by use of a simple e-tender box even though it may have cryptographic keys. This indicates that a different approach to the e-tender box is needed. This paper presents a secure e-tender submission protocol to address the advanced security requirements in e-tender submission. The principles of commitment schemes have been applied to the protocol design to prevent submission time dispute and collusion between favoured parties. The protocol is assumed to run under the condition that all tendering parties (principal and tenderers) are dishonest players. The security analysis shows that the protocol meets its security goals under well known colluding scenarios.

## 1 Introduction

Tendering is a process used in awarding government contracts. The tendering process is governed mostly by contract law. The basic components in the tendering process are performed in sequential order as shown in Fig. 1. The components are pre-qualification and registration, public invitation, tender preparation and submission, close of tender, opening tender, tender evaluation, award of tender, and archiving. Any tenderer has to ensure that its tender is submitted before the tender close time. The opening of tenders occurs after the tender close time.



**Fig. 1.** Tendering Process

An electronic tendering (e-tendering) system is usually considered to be more efficient and cost-effective than the traditional paper based system. Demand has generated a large number of e-tendering systems around the world. In general, most of the current e-tendering systems mirror some paper based process but

are untested with regard to both security and legal compliance. An e-tendering system with inadequate security provides opportunities for fraud and collusion by parties both inside and outside of the process. As an example, submitted tenders are confidential and are commonly the target of business collusion when a tenderer attempts to obtain its competitor's tender offer before opening time. To prevent this collusion requires implementing an advanced security protocol going beyond basic security services such as confidentiality and data integrity.

An electronic tender box has been included in most fielded e-tender systems to collect submitted tenders before the tender opening time. Various proprietary solutions have been used to protect the e-tender box but the common problem with these solutions is that the system administrator still has the full capacity to tamper with the submitted tenders. Although a secure e-contracting protocol has been proposed [4] to maintain the integrity of the e-tendering process, adequate security solutions for the e-tender submission phase remains undiscussed.

Our contribution is to provide a secure e-tender submission protocol, suitable as a stand-alone protocol or for integrating with more general protocols for e-tender security [4]. Our protocol addresses the special security requirements related to e-tender submission, including resistance to collusion between the principals and one or more tenderers. The next section discusses the security requirements related to e-tender submission. We then review the background technicalities in Section 3 before describing our new protocol in Section 4. Section 5 explains why the security goals are met by our protocol.

## 2 E-Tender Submission Security Requirements

Traditionally the tender submission process has been carried out by using a physical tender box placed in a public area. Tenderers submit their tenders into the tender box before the submission close time. The tender box is normally opened at the submission close time. Tenders that are submitted on time will be publicly recorded. Any later submission is considered as a non-conforming tender and will be rejected. Making the tender box publicly accessible increases the transparency of the process.

A simple e-tender box does not function in the same way as a traditional tender box. The simple e-tender box is typically a directory in a system server, which allows tenderers to upload their tender offer to that directory. The fundamental difference from the physical tender box is that access to the e-tender box has become a private activity and cannot be publicly monitored, thus removing transparency. A significant opportunity has been created for business collusion by using a simple e-tender box even though access to it may be protected by cryptographic keys. The server administrator typically has access to the e-tender box and is able to read its contents before the submission close time or alter those contents after the close time.

The integrity of time of receipt could be compromised by both receiver and senders if there were no security mechanism in place. This provides an opportunity for collusion and lead to unfair trading practices. Moreover, any late



submission should be identified as a non-conforming tender. The alteration of the submission or receiving time can raise a dispute as to whether a tender conforms or not. If the system clock is controlled by the local administrator there is scope for the time of submission to be changed at any time.

In a normal sales contract, seller and buyer try to maximize their own benefits. In tendering, when one of the tenderers (sellers) is the principal's favourite, there is a tendency to alter submitted tenders (price or other items) in order to win the contract.

Any e-tender system requires a number of basic security services to protect the confidentiality and integrity of tendering information and to authenticate the parties concerned [4]. However, e-tender submission faces some specific threats illustrated in the following three risk scenarios.

**Scenario 1.** The principal releases tender submissions to its favourite tenderer before tender submission opening time. The principal's favourite tenderer can then submit a competitive tender and win the tender project.

**Scenario 2.** The principal allows its favourite tenderer to alter its tender after the tender official opening time. This tender then becomes competitive and wins the tender project. This alteration is not only limited to price change.

**Scenario 3.** A dispute may occur between the principal and any of the tenderers over whether any tender submission happened before tender submission closing time. This may allow a tenderer to submit a late tender without risking rejection, thereby gaining an advantage over other tenderers.

Consideration of these threat scenarios leads to the following security requirements for electronic tender submission.

**Submission hiding** ensures that no party can reveal any electronically submitted e-tender document before the designated tender opening time. This is to prevent any party from gaining another party's tender strategy before tender close time.

**Submission binding** detects whether any party altered any tender submission after the tender closing time. This is to prevent business collusion between the principal and its favoured tenderer.

**Submission time integrity** service ensures that time of tender submission can be recorded in a reliable manner. This is to provide reliable evidence to determine whether a tender submission is on time.

### 3 Related Technologies and Application Issues

Digital signature schemes, commitment schemes and time stamping services are useful cryptographic technologies for the e-tender submission protocol. The commitment function will be used to generate a document integrity checksum for the tender submission process. The protection of the checksum is provided by using a digital signature during the tender submission process. Time stamps will be provided by a time stamp authority (TSA) to guarantee the submission time of all parties' commitments.

### 3.1 Commitment Scheme

A *commitment scheme* [3] is a protocol between two parties called the prover  $P$  and the verifier  $V$ . There are two phases.

1. In the *commitment phase*  $P$  provides  $V$  with the commitment  $C(m)$  to the message  $m$ .
2. In the *opening phase*  $P$  provides  $V$  with the value  $m$  and  $V$  can verify whether or not it is a correct opening of  $C$ .

Commitment schemes are typically constructed from one-way functions. For example, consider the commitment function  $C(m) = g^m$  where  $g$  is a generator of  $\mathbb{Z}_p^*$ , the integers modulo  $p$  for some prime  $p$ . Given the value  $m$  the verifier  $V$  can re-compute  $g^m$  in order to check the commitment. There are two basic but essential properties to any commitment scheme.

1. The **hiding** property prevents  $V$  from revealing the committed value  $m$  in the protocol commitment phase.
2. The **binding** property prevents  $P$  from changing its committed value  $m$  after commitment phase.

We can apply the concept of the commitment scheme to e-tender submissions to ensure that the principal cannot reveal a tenderers' tender before tender opening time, and tenderers cannot change their submissions after the tender close time.

A commitment function can provide either *unconditional* or *computational* assurance of hiding and binding. For example, consider again the commitment function  $C(m) = g^m$ . If  $1 < m < p$  then this function provides unconditional binding since there is only one possible value of  $m$  given  $C(m)$ . Therefore even with unlimited computational power, it would be impossible for the prover  $P$  to change its mind after committing. However, this same function provides only computational hiding since if  $V$  has sufficient computational power to take discrete logs then  $V$  can reveal  $m$  before it is opened by  $P$ . There are also commitment functions which in contrast have unconditional hiding and computational binding [7]. However, it is not hard to show that no commitment scheme can provide both unconditional hiding and unconditional binding. Likely candidates for the commitment function in our protocols include the following.

- $C(m) = g^m$  where  $g$  is described above. Such a function is suitable when unconditional binding is important.
- $C(m) = g^m h^r$  where  $g$  and  $h$  are independent generators of  $\mathbb{Z}_p^*$  and  $r$  is chosen randomly each time a commitment is made. This is Pedersen's scheme [7] and provides unconditional hiding.
- $C(m) = h(m)$  for some one-way hash function  $h$ . This may be suitable when efficiency is most important but this provides (at best) computational hiding and binding.

In choosing a suitable commitment scheme for our protocol we must balance the level of assurance for e-tender submission according to its legal purposes.

The submitted tender offer must be preserved over the long term as a requirement for archiving purposes. Therefore it is preferable to choose a scheme that provides unconditional binding assurance rather than one that provides unconditional hiding assurance (since we cannot have both). Although in the tender submission (commitment) phase, only computational hiding is provided, the period between submission and tender opening time will be within a few hours. This is significantly shorter than the period between tender opening and awarding time (which may be days or months), and particularly shorter than the document archiving requirements (which will usually be years). Therefore the binding property should be given the higher assurance. Note that even though computational assurances are less strong than unconditional assurances, we still expect them to hold within any reasonable lifetime.

**Role of Players.** All commitment schemes assume that the prover  $P$  and the verifier  $V$  are adversaries. The scheme will provide the hiding and binding properties only if no collusion occurs between the players  $P$  and  $V$ . When collusion happens, the principal is not a trustable verifier. One way to achieve this may be to distribute the role of  $V$  amongst multiple players.

It must be assumed that all tendering related parties are dishonest players, namely the principal  $A$  and the set of tenderers  $\mathcal{B}$ . In a real situation, it is very difficult to determine at what point which party is honest, therefore their commitments preferably should be held by trusted third parties. At this point, it will be a good strategy to introduce time stamping service.

In a colluding situation, colluding parties will be the prover  $P$  and their opponents are verifier  $V$ . It clearly indicates that non-colluding parties have to hold colluding parties' commitments. This will raise the credibility of a verifying process in the protocol.

### 3.2 Time Stamping Services

The function of a time stamping service is to reduce disputes over document generation time. Time-stamping services have been proposed and analysed by many researchers [2,6,8]. The definition of a time stamp is digital data intended to prove the existence of digital documents prior to or at a specified time.

In general, a time stamp service requires that a client send a request to a service provider through the Internet to gain a time stamp for a document. The service provider issues the time stamp of the document and sends it back to the requester. Other processes could be involved, such as the service provider publishing the time stamp to enhance the service integrity. If a dispute occurs at a later time, the integrity of the time stamp and related document will be verified through verification procedures associated with each time stamp scheme.

Time-stamping technology has been studied for more than a decade [5]. Traditionally it has been classified into two types according to its issuing process: conventional/simple and linking schemes. Haber [5] also proposed a distributed trust scheme by involving a trusted third party in issuing process. For systematic security analysis of time stamp schemes, Une and Matsumoto [9,8] performed

fine grained classification based on many aspects involved in time stamping other than just issuing process. Regardless of the significant body of research [1,2], all time stamping services require that a requester and issuer do not collude [9]. The hybrid time stamping service<sup>1</sup> with hardware support<sup>2</sup> and linked schemes will largely limit the capacity for collusion between requester and issuer.

## 4 Protocol Description

The secure e-tender submission protocol meets the special security requirements of a tender submission. It addresses the issues of time disputes and tender collusion between a principal and its favorite tenderer.

### 4.1 Notation

Commonly used notations in the protocol are listed in Fig 2. The function *commit()* will represent any suitable commitment function. Party *A* represents the principal in the tender process. Party *B* represents a tenderer in the set of all potential tenderers  $\mathcal{B}$ .

SYMBOL	NAME
$\longrightarrow$	One party sends another party a message eg. $B \longrightarrow A$ , $B$ send to $A$
$\longleftrightarrow$	two way communication
$\parallel$	Concatenation
$TSA$	Time Stamp Authority
$Priv_{ID}$	Private key of party ID eg. $Priv_B$ , $B$ 's private key
$Pub_{ID}$	Public key of party ID
$CT_{ID}$	Certificate of party ID
$commit$	commitment function
$Sig$	Signature generation function
$V$	Verifying function
$E_x$	Asymmetrical encryption function, $x$ represents input key
$D_x$	Asymmetrical decryption function, $x$ represents input key
$m_s$	Tender submission message in the contract negotiation
$TS_s$	TSA's time-stamp on tenderer's commitment on its $m_s$
$TS_{pl}$	TSA's time-stamp on concatenation of all tenderer's commitments received by Principal

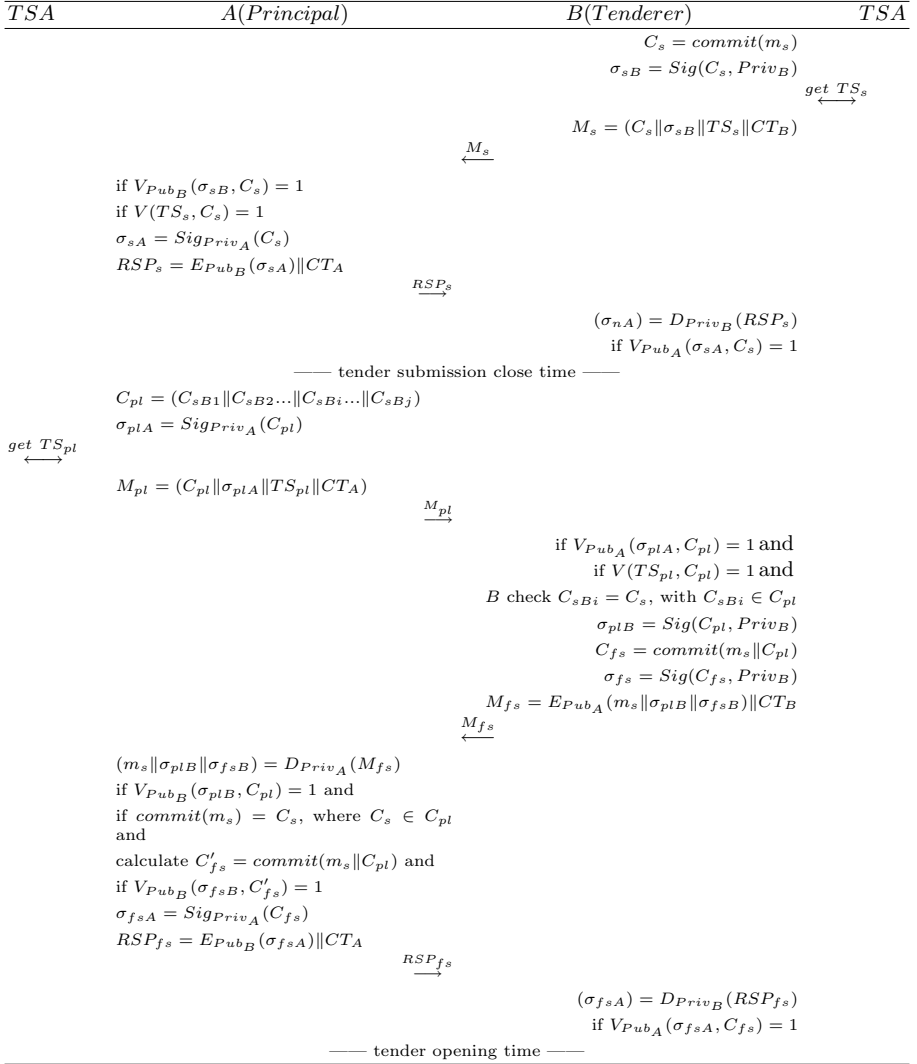
Fig. 2. Notation

### 4.2 E-Tender Submission Protocol

The e-tender submission protocol contains processes for tender submission, close of tender and opening of tender. E-tender submission protocol (Fig 3) contains the following steps:

<sup>1</sup> <http://www.e-timestamp.com/evidence.htm>

<sup>2</sup> <http://www-03.ibm.com/security/cryptocards/pcicc.shtml>

**Fig. 3.** Tender Submission Sub-Protocol

1. Every tenderer  $B$  requests a timestamp  $TS_B$  for its signed commitment  $\sigma_{sB}$  of its offer  $m_s$ , and sends  $(C_s \parallel \sigma_{sB} \parallel TS_s \parallel CT_B)$  to principal  $A$  before tender close time.  $A$  will verify  $TS_s$ ,  $\sigma_{sB}$  and send its confirmation  $RSP_s$  including  $\sigma_{sA}$  to each tenderer  $B$ .
2. At the tender closing time,  $A$  concatenates all received commitment and requests a timestamp for the concatenation ( $TS_{pl}$ ). It then sends timestamps  $TS_{pl}$  and concatenated commitments to all tenderers who have submitted offers (tenders). The message also acts as a call from principal  $A$  to all tenderers to submit their full document (offer).

3. On receiving  $M_{pl}$ , each tenderer verifies all commitments and signatures, and sends its encrypted full tender document  $m_B$  and other relevant values.
4.  $A$  extracts the documents  $m_B$  for each  $B$  and verifies the signatures, generates signature  $\sigma_{fsA}$  for the node  $C_{fs}$ , and send  $RSP_{fs}$  to each  $B$ .
5. Each  $B$  will verify the confirmation and tender can be opened.

Step 1 is the tender submission process, step 2 is the tender closing process, steps 3 and 4 are the tender opening process. Once all full tender documents are received, tenders can be officially opened.

## 5 Security Analysis

We summarise the required security goals of the protocol:

1. the principal  $A$  and its favourite tenderer  $B_{fav}$  cannot reveal other tenderers tender value  $m_s$  using  $C = \text{commit}(m_s)$  before tender opening time.
2. any alteration of  $m_s$  after tender opening time can be detected by the verification process.
3. any alteration of time stamping value  $TS_s$  and  $TS_{pl}$  can be detected by the verification process.

We assume that all players (principal and tenderers) are dishonest players. They have interception, insertion and alteration powers at any stage of the protocol run.

**Interception Power:** The power to intercept all parties' network messages in order to gain other parties tendering strategies.

**Insertion Power:** The power to insert malicious messages into the network during the protocol run. For example players can replay/relay intercepted messages or insert extra tender values during protocol run.

**Alteration Power:** The power to manipulate (alter, delete, and insert) all protocol generated elements belonging to them. It includes: the set of communicated messages, the set of signatures from message originator and receiver, the set of time-stamps from  $TSA$ , and the set of signatures from the trusted third party.

### 5.1 Protocol Assumptions

Protocol assumptions define a set of security conditions that a running environment should provide for e-tender submission protocol.

- $TSA$  is a trusted party and generates reliable time-stamp.
- Keys are securely stored and no party will intentionally release its private keys to any other (non-colluding) party participating in the tendering;
- No party will consciously sign anything that they do not agree upon.
- Verifying (challenging) procedures are transparent, run in the public by trusted third party, such as court and judges;

- Supplied verifying (challenging) elements are publicly available during the verifying process.
- Any party can challenge the process.

## 5.2 Analysis

A dishonest player may attempt to gain financial benefit through the defined attacks by using the powers described above. These attacks should be deemed successful only if they cannot be detected by the protocol verification procedures and the dishonest player gains financial benefit.

**Hiding Tender Submission.** The protocol prevents a dishonest principal and its favourite tenderer from gaining other tenderer's tender strategy during tender submission process, and before tender opening time. This will prevent  $B_{fav}$  submitting a more competitive tender, by knowing their tender price, than other tenderers  $B_{opp}$ .

During submission process only the commitment  $C_s$  is required. Tenderers do not need to submit their full tender documents. The colluding principal has the power to access every party's commitments and pass them to its favourite tenderer. However, if the commitment function provides the hiding property, no tender strategy can be obtained from the commitment value. The protocol uses the hiding property to prevent any colluding party from revealing opponents' tender value  $m_s$  before tender opening time.

**Binding Tender Submission.** The protocol also prevents colluding parties from successfully changing their commitment during the tender opening process by detecting the alteration through the protocol verification process.

In this situation, the colluding parties will change  $B_{fav}$ 's tender  $m_s$  to a competitive value  $m'_s$  after all tenderers have submitted their full tender documents. To cover this alteration, they would need to recalculate all related values to avoid the attack being detected. The colluding parties, however, cannot recalculate time stamps  $TS_s$  and  $TS_{pl}$ .

Therefore the verification process will detect that  $V(TS_s, C'_s) \neq 1$  and  $V(TS_{pl}, C'_{pl}) \neq 1$ , with  $TS_s$  and  $TS_{pl}$  supplied by time stamping authority  $TSA$ . The non-colluding tenderers  $B_{opp}$  also hold the principal's commitment  $C_{pl}$  and  $TS_{pl}$ .

The binding property of the commitment scheme is able to detect alteration of committed values  $m_s$ . The colluding parties cannot change  $m_s$  to  $m'_s$  without detection, therefore rendering the attack unsuccessful.

**Fixing Submission Time.** To prevent dispute over submission time of a commitment all parties are required to obtain a time stamp for their commitments. The protocol assumes that  $TSA$  is a trustworthy party - therefore time stamps  $TS_s$  and  $TS_{pl}$  are trustable values. The dispute can be resolved by examining whether  $TS_s \leq closetime \leq TS_{pl}$ , with  $TSA$  supplying the  $TS_s$  and  $TS_{pl}$ . The integrity of  $TS_s$  and  $TS_{pl}$  can also be verified.

## 6 Conclusion

There is fundamental difference between a simple e-tender box and the traditional physical tender box. This leads to a range of new security threats, particularly those including collusion between the principal and one or more tenderers. This paper has presented a secure e-tender submission protocol for providing advanced security service to prevent risks and collusions related in e-tender submission.

It will also be interesting to integrate the tender submission protocol with protocols for secure communications in tender negotiation [4].

## References

1. Ahto Buldas and Peeter Laud. New linking schemes for digital time-stamping. In *Information Security and Cryptology*, pages 3–13, 1998.
2. Ahto Buldas, Peeter Laud, Helger Lipmaa, and Jan Vilemson. Time-stamping with Binary Linking Schemes. In Hugo Krawczyk, editor, *Advances on Cryptology — CRYPTO '98*, volume 1462 of *Lecture Notes in Computer Science*, pages 486–501, Santa Barbara, USA, August 1998. Springer-Verlag.
3. Ivan Damgard. Commitment schemes and zero-knowledge protocols. *Lecture Notes in Computer Science*, 1561:63, Jan 1999.
4. R. Du, E. Foo, C. Boyd, and Kim-Kwang Raymond. Formal analysis of secure contracting protocol for e-tendering. In Safavi-Naini, Rei and Steketee, Chris and Susilo, Willy, editor, *Fourth Australasian Information Security Workshop (Network Security) (AISW 2006)*, volume 54 of *CRPIT*, pages 155–164, Hobart, Australia, 2006. Australian Computer Society Inc and ACM.
5. Stuart Haber and W. Scott Stornetta. How to time-stamp a digital document. *Journal of Cryptology*, 3(2):99–111, 1991.
6. Henri Massias, Xavier Serret, and Jean-Jacques Quisquater. Main issues on their use and implementation. In *Infrastructure for Collaborative Enterprises - Fourth International Workshop on Enterprise Security*, IEEE 8th International Workshops on Enabling Technologies, pages 178–183. ISBN 0-7695-0365-9, 1999.
7. T. Pedersen. A threshold cryptosystem without a trusted party (extended abstract). In D.W. Davies, editor, *Advances in Cryptology - EUROCRYPT '91*, volume 547 of *LNCS*, page 522. Springer-Verlag, 1991.
8. Masashi Une. The security evaluation of time stamping schemes: The present situation and studies. In *IMES Institute for Monetary and Economic Studies*, number No.2001-E-18 in IMES Discussion Paper Series. Bank of Japan, C.P.O BOX 203 Tokyo 100-8630 Japan, 2001.
9. Masashi UNE and TSutomu Matsumoto. A framework to evaluate security and cost of time stamping schemes. *IEICE TRANS. Fundamentals*, E85-A:125–139, 2002.



# A Sophisticated Solution for Revealing Attacks on Wireless LAN

René Neumerkel and Stephan Groß

Technische Universität Dresden  
Department of Computer Science  
Institute for System Architecture  
D-01062 Dresden, Germany  
{rene.neumerkel, stephan.gross}@tu-dresden.de

**Abstract.** The development of the WPA and IEEE 802.11i standards have vastly improved the security of common wireless LAN setups. However, many installations still use the broken WEP protocol or even run with no security settings enabled. Furthermore, several threats are only partially addressed by the new security standards, e.g. rogue access points or denial of service. Specialised wireless intrusion detection systems are promising means to protect wireless networks against these threats. They can further improve the reliability and security of these networks. In our contribution we present such a wireless IDS to reveal common attacks on wireless LAN. We describe the development and evaluation of our prototype solution that seamlessly integrates with approaches for traditional wired networks.

## 1 Introduction

During the last decade we were witnessing the breakthrough of wireless communication techniques. Today, the vision of seamless Internet access everywhere at anytime has almost become true, e.g. by the growing number of wireless access points in public and private places. Unfortunately, security has not been a design goal in the first place of the underlying technical foundations. Thus, many of the deployed techniques suffer from severe security drawbacks. For example, the Wired Equivalent Privacy (WEP) protocol has once been the standard security mechanism in IEEE 802.11 wireless LAN. In 2001 Fluhrer et al. described a way to utilise a design flaw in WEP's usage of the key scheduling algorithm RC4 to break the encryption [1]. Since then, several free available software tools empowered even unskilled users to penetrate their neighbour's wireless LAN. At least with the publication of statistical cryptanalysis attacks in 2004 like the KoreK [2] or the chopchop attack [3], the WEP protocol must be considered definitely defeated. These attacks no longer depend on millions of captured packets to crack a WEP key but combine several techniques like traffic injection or predefined password dictionaries to derive a key within several minutes. Today, the usage of recent security protocols like WPA [4] or WPA2 aka IEEE 802.11i [5] are indispensable. But even with these measures enabled, a wireless LAN can still be plagued by security issues like Denial of Service attacks or misconfigured access points. To overcome these threats one should consider an intrusion detection solution to monitor the current status

of a wireless network. As common IDS are only considering ISO/OSI layer 3 upwards they are not able to detect specific wireless threats taking place on the MAC layer. Furthermore, unlike wired networks, the special characteristics of mobile networks pose a number of new challenges to the security design we cannot solve using traditional approaches [6]. Thus, we need specialised Intrusion Detection Systems for wireless LAN.

There are several commercial as well as open source solutions available for that purpose. Examples for commercial competitors are Airdefense<sup>1</sup>, NetworkChemistry<sup>2</sup> or Internet Security Systems<sup>3</sup>. Prominent open source projects are Snort-Wireless<sup>4</sup>, Kismet<sup>5</sup> and WIDZ<sup>6</sup>. However, in our opinion all available solutions suffer from one or more of the following limitations:

- They require special infrastructural means. Thus, they can only be applied in a fixed environment with a centralised network structure (e.g. Airdefense Guard).
- They unnecessarily depend on specialized hardware even if the implemented features can be solved in software (e.g. NetworkChemistry RFprotect).
- They are not integrated with common Intrusion Detection and network management solutions (e.g. WIDZ).
- They do not allow the addition of user-defined detection strategies as they solely rely on built-in attack signatures (e.g. Kismet)
- They require changes in source code when adding more complex detection strategies that involve more than one packet (e.g. SnortWireless)

In this paper we present a solution to overcome these drawbacks. The remainder of this paper is organised as follows: First, we summarize related work in section 2. Section 3 deals with the design and implementation of our prototype wireless intrusion detection system. We utilise the Bro IDS for this purpose which was originally designed for monitoring high-speed wired networks and which we enhanced with several specific means to observe wireless networks. Section 4 is dedicated to a detailed report of our prototype's validation. Finally, we come up with some final remarks on our results and on directions for further investigations.

## 2 Related Work

We have already mentioned commercial and open source solutions in the previous chapter. In addition, there exist several contributions from the academic community from which we only mention some selected works. The need of reconsidering traditional network protection for mobile environments is addressed in [7]. [8] already presents a security architecture for mobile ad-hoc networks based on anomaly detection. However, this work tackles the problem from a more academic point of view and does not

<sup>1</sup> <http://www.airdefense.net/>

<sup>2</sup> <http://www.networkchemistry.com/>

<sup>3</sup> <http://www.iss.net/>

<sup>4</sup> <http://snort-wireless.org/>

<sup>5</sup> <http://www.kismetwireless.net>

<sup>6</sup> <http://freshmeat.net/projects/widz/>

address issues appearing in real-world scenarios. In [9] Lim et al. describe a prototype implementation of a wireless intrusion detection system. They modified an off the shelf wireless access point in order to detect common wireless attacks. In [10] they further enhance their prototype implementation with an active countermeasure capability and demonstrate the usefulness of their approach by a case study. Their approach is quite similar to ours. However, they fully concentrate on the wireless scenario and do not consider the integration with general intrusion detection solutions as we do. The Distributed Wireless Security Auditor (DWSA) presented in [11] works toward finding unauthorized wireless access points in large-scale wireless environments. The system utilises a centralised architecture and is, thus, only applicable in a fixed environment. The most interesting feature of the DWSA might be its ability to track down a potential adversary based on three-dimensional trilateration.

### 3 Developing a Wireless Intrusion Detection System

The starting point for our research work has been an intensive literature and Internet research on commonly known threats to IEEE 802.11 wireless networks. Today, there exist a multitude of basic techniques and ready to use exploits for compromising a wireless LAN. However, these techniques can all be traced back to three basic principles: violation of confidentiality, integrity or availability [12]. For the purpose of categorizing the acquired threats we used attack trees, a semi-formal method to analyse and document the security of a given system with respect to varying attacks [13]. Due to the lack of space we refer to [14] and [15] for a more or less complete description of commonly known threats to wireless LAN.

For developing our own wireless intrusion detection system we utilise an existing wired IDS called Bro<sup>7</sup> and added the necessary functionality for monitoring wireless networks. Thus, we are not only able to overcome the above mentioned missing integration of wireless solutions with conventional intrusion detection and network management systems but also improve the functional range of our system as we can fall back on the full range of mechanisms already implemented in the base system. In this section we first give a brief overview to the Bro IDS before we describe our modifications.

#### 3.1 Introducing the Bro IDS

Bro is a Unix-based open source Network Intrusion Detection System (NIDS) designed for monitoring high-speed, high-volume wired networks. The system detects intrusions in real-time by passively monitoring network traffic and comparing it to a set of user-defined rules describing security-related events such as the occurrence of known attacks or unusual network activities [16].

Bro uses *libpcap* to capture network traffic. The received packet stream is processed by *Analyzers* operating on protocols at OSI layer 3 and above. Analyzers exist, for example, for ICMP, TCP and HTTP. Besides that, Bro also uses a *Signature Engine* which matches the incoming packet stream against patterns of known attacks, thus realising basic misuse-based intrusion detection. In both cases built-in events are generated

---

<sup>7</sup> <http://www.bro-ids.org>

whenever interesting network activity (e.g. a failed TCP connection attempt) occurs. Generated events are processed by an *Event Engine* which is responsible for calling the associated *Event Handlers* located in *Policy Scripts* outside of Bro's core. Event handlers can execute various actions like logging, real-time notification or generation of user-defined events. For this purpose, Bro provides a special scripting language which allows end users to define their own policy scripts. Besides that, Bro comes with a pre-defined set of policy scripts that can be modified to reflect a site's actual security policy.

By putting the event handlers outside of Bro's core, a clear separation of built-in detection mechanisms from site-specific interpretation of events is achieved. This becomes especially useful in mobile environments with their frequently changing network topologies as it simplifies the adaptation to new conditions.

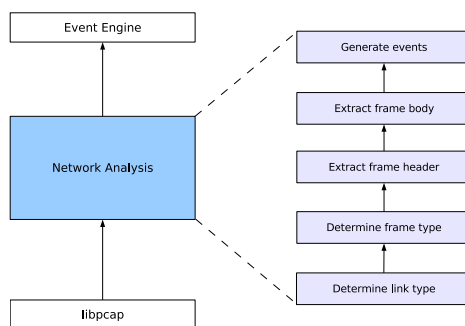
### 3.2 Enhancing Bro for the Wireless World

As mentioned before, Bro has been specifically designed for wired intrusion detection and thus does not provide any means for monitoring wireless networks. Therefore, several modifications to Bro's architecture were necessary which we will describe in the following.

**Capturing of 802.11 frames.** In order to monitor wireless traffic, Bro needs to access raw 802.11 frames. As mentioned before, Bro relies on libpcap to capture network traffic. Since current versions of libpcap already support the capturing of raw 802.11 frames, only minor changes were necessary to make Bro recognize the data link types identifying wireless links. However, capturing wireless frames alone is not enough. When monitoring 802.11 traffic it is also important to know on what channel the capture device is listening on. Therefore, we added the necessary code to determine the current channel of each active wireless device at the start-up of Bro. It is also possible to determine the current channel on a per-packet basis. For this purpose, some drivers include an optional pre-header before the regular 802.11 headers.

At this point it seems important to mention that the modified Bro system is still capable of monitoring wired networks. Furthermore, it is also possible to monitor both wired as well as wireless networks simultaneously.

**Processing of 802.11 frames.** Once our prototype was able to capture raw 802.11 frames, further functionality was added to allow processing of captured frames. We currently focus on the analysis of 802.11 management frames since the majority of wireless network attacks is based on this frame type. The basic steps are shown in figure 1 and include extracting all available information from header fields and frame body. Our implementation currently supports only field types defined by the original IEEE standard and ignores any unknown field types including vendor-specific information elements. Based on the information extracted from frame header and body, built-in events are generated which yet had to be defined (see next section). In addition to analyzing management frames, preliminary processing of data frames and control frames is already being done and could be further implemented when necessary.



**Fig. 1.** Processing of 802.11 frames

Note that our current solution does not handle fragmented frames, as this was not a requirement for our prototype. However, we believe that the missing functionality could easily be added anytime if needed.

**Definition of 802.11-Specific events.** Once we added the functionality to process 802.11 frames, we needed to define the events to be generated based on the network traffic. We found that there is more than one possible approach when defining new events. We eventually decided to implement a straight-forward approach where we simply map each management frame sub-type to a corresponding event, for example *ProbeRequest-Received*. This way, traffic analysis can be done completely at script level, giving us the highest possible flexibility. Note that, for our prototype, performance was not an issue at this point. When it comes to monitoring busy wireless networks, we probably would have to think of a more efficient solution than the one currently implemented.

**Extending Bro's script language.** Having defined new built-in events for 802.11, we would now begin writing corresponding event handlers. However, we found that the scripting language used by Bro to define policy scripts is not sufficient when it comes to writing policies for 802.11. In particular, the language does not provide a suitable data type for working with addresses used by the 802.11 protocol. We modified Bro's policy script interpreter and added built-in support for the 802.11 address type. This turned out to be the most time-consuming task during developing our prototype.

Now we are ready to start writing 802.11-specific security policies which is covered in section 4.1.

## 4 Prototype Validation

For the validation of our prototype we looked at three common threats to wireless networks and developed appropriate detection strategies which we incorporated into a series of policy scripts used by our prototype. The policy scripts were expressed using our extended version of the Bro script language. During an experiment we simulated different attack scenarios and verified the proper functioning of our prototype. In the

following section we will describe the strategies which we used for our prototype and present the results of our lab test.

#### 4.1 Policy Definition

The process of turning a detection strategy into a security policy that can be used by our prototype basically involves specifying the appropriate event handlers defining how certain events should be interpreted. The challenge, however, is to find a suitable detection strategy in the first place. We present some basic strategies for detecting three common wireless threats. They form the foundation of the policy scripts we used during our lab test.

**Rogue Access Points.** An important step towards detecting wireless attacks is locating so called *rogue access points*. Traditionally, the term *rogue access point* refers to wireless access points that have been attached to a wired network without explicit permission of the administrator. Such access points represent a direct threat to the respective network as they may circumvent existing security measures. However, it is important to note that also access points which are not connected to the wired network can be a security problem when they associate with authorized clients. Imagine an employee's laptop associating with an unknown access point across the street while being connected to the company's wired network. That is why we have to extend the notion of rogue access points to include any unknown access point within range of our sensors—whether or not it is attached to the wired network.

To detect rogue access points, we implement the strategy described in [9] where every discovered access point is matched against a list of trusted access points. In our case, this list includes the MAC address (Basic Service Set Identification, BSSID), the network name (Service Set Identifier, SSID) and the operating channel for each trusted access point. In order to discover new access points, the monitor listens to *Beacon Frames* and *Probe Response Frames*. Note, that an access point must send out at least one of those two frame types to be recognized by a client. When comparing the BSSID, SSID and channel information contained in the received frames with those stored in the list of trusted access points, we distinguish between the following situations:

1. Both BSSID and SSID are completely unknown to the system. In this case, the system would give a notice that an unknown access point has been detected and send an alert as soon as an authorized client associates with the rogue access point. Herefore, one would of course have to address the problem of how to distinguish between foreign and authorized clients.
2. BSSID, SSID and operating channel match an entry in the list of trusted access points. In this case, the system assumes that it has detected a trusted access point. Note however, that it is still possible that an attacker replaced the original access point by his own.
3. BSSID or SSID are known to the system, however, it cannot find a matching entry in the list of trusted access points. In this case, the system would assume that someone attempts to spoof a valid access point and immediately sends an alert.

This strategy may not be suitable for large wireless networks with hundreds of access points. However, in cases where it is possible to manage a list of trusted access points, detection of rogue access points can be implemented in a very straight-forward manner.

**Denial-of-Service.** There exist different techniques for launching DoS attacks against wireless networks. Most popular attacks use some form of management frame flooding. For example, an attacker could flood the network with spoofed *De-Authenticate Frames* which seem to originate from the legitimate access point. This way, it becomes impossible for any client station to stay connected with this access point. On the other hand, an attacker could flood an access point with spoofed *Authentication Frames*, simulating hundreds of individual client stations attempting to authenticate with the access point. Eventually, the access point will be unable to respond to legitimate client requests [17].

When we look at different forms of management frame flooding, we find that they usually have one or more of the following characteristics:

- exceptional high frequency of certain management frames
- exceptional large number of different source addresses
- destination address set to broadcast address when it should not
- use of invalid source addresses
- unrealistic number of unique network names (SSID) on a single channel

By applying basic sanity-checks on the received management traffic most known flooding attacks should be detected. One could, for example, define threshold values for the number of unique source addresses received during a certain period of time (*Authentication Flooding*), the number of unique network names per channel (*Beacon Flooding*) or just the plain frequency of certain management frame types. Additionally, one could check for invalid source addresses or the destination field set to multicast addresses where it is not appropriate. The actual difficulty is to find suitable threshold values. Setting them too low would cause too many false alarms while setting them too high could mean that we miss less aggressive attacks.

**Man-In-The-Middle.** Another very popular intrusion technique is the so called man-in-the-middle attack where the attacker positions himself between an authorized client station and an access point. This attack involves setting up a rogue access point, usually combined with spoofing the network name and sometimes even the BSSID of a legitimate access point. Once an authorized client associates with the fake access point, the attacker configures his wireless interface to use the MAC address of the client station and connects to the unsuspecting access point on behalf of the legitimate client. Optionally the attacker could send spoofed *De-Authenticate Frames* prior to the attack in order to force clients to disconnect from the access point.

In order to develop a strategy for detecting man-in-the-middle attacks, we look at them from the perspective of the monitor. From this point of view, a typical attack would look similar to this: When started, the monitor identifies a trusted access point on channel 3. At some point during runtime, it discovers a new access point on channel 11 which it identifies as rogue access point. Depending on the attackers strategy, the monitor could also witness some kind of denial-of-service attack on channel 3. Either

way, the monitor would, at some point, record a successful client association to the rogue access point on channel 11. Shortly after that it would witness the “same” client connecting to the legitimate access point on channel 3.

Based on these observations, a possible strategy could be as follows: first, the monitor has to detect an existing rogue access point. Furthermore, the monitor keeps a record of established connections between access points and clients. It sends an alert as soon as it detects simultaneous connections from one client to different access points, one of those being an identified rogue. Detecting the ongoing MAC spoofing would further improve the monitor’s level of confidence. This, however, requires some sophisticated detection techniques as described in [18] and is beyond the scope of this paper.

The strategies described here were incorporated into several policy scripts using Bro’s internal scripting language.

## 4.2 Laboratory Test

The setup used in our lab test is shown in table 1. Our prototype uses off-the-shelf hardware: a notebook with a 1 GHz processor and two wireless PCMCIA cards. Monitor, Attacker and Victim where situated in the same room. The access point was placed in a different room to give the attacking station the necessary gain in signal strength.

During our lab test we conducted the attacks described in section 4.1 using tools available from the Internet. The results of our test are shown in table 2. Except for the Association and Authentication Flooding, all of the attacks were successful. We

**Table 1.** Test setup

<b>Attacker</b>	Debian GNU/Linux 3.1 (Kernel 2.4.29) Netgear WG311 (Atheros chipset, madwifi driver 15.05.2005) Linksys WPC11 (Prism chipset, hostap driver 0.4.1, airjack driver 0.6.6alpha)
<b>Monitor</b>	Debian GNU/Linux 3.1 (Kernel 2.6.12) Netgear WAG511 (Atheros chipset, madwifi driver 15.05.2005) Netgear MA521 (Realtek chipset, driver by Andrea Merello version 0.21)
<b>Victim</b>	Windows XP Professional (Service Pack 2) Avaya Wireless Silver World Card (Hermes chipset, Windows driver)
<b>Access Point</b>	D-Link DI-624+

**Table 2.** Test results

<b>Attack</b>	<b>Tools</b>	<b>Successful</b>	<b>Detected</b>
Rogue Access Point	hostapd + aircsnarf	Y	Y
Association-Flooding	void11	N	Y
Authentication-Flooding	void11	N	Y
Deauthentication-Flooding	wlan_jack	Y	Y
Beacon-Flooding	fakeAP	Y	Y
Man-In-The-Middle	monkey_jack	Y	Y
Client MAC-Spoofing	no special tool	Y	N



believe that the reason for this is because our access point already had some built-in protection mechanism against this type of attack. More importantly however, except for the client MAC spoofing, all intrusion attempts have been detected by our prototype. As mentioned before, detecting MAC spoofing involves more sophisticated techniques than those described here. However, we believe that it should be possible to incorporate adequate strategies into our prototype.

## 5 Conclusion and Perspectives

In this paper we presented our prototype of a wireless intrusion detection system. We introduced an existing wired IDS and described our modifications in order to use it for monitoring wireless networks. Also, we analysed common wireless threats and came up with a strategy for detecting them. Finally, we incorporated those strategies into a series of policy scripts which we used to validate our prototype. We demonstrated that our prototype in fact realises basic wireless intrusion detection. Our approach does not require any special infrastructural means or specialised hardware. Thus, it can be easily adopted for any mobile environment. By utilising an existing intrusion detection solution for wired networks our system can easily be integrated into a comprehensive security solution.

However, a few issues remain unsolved. Being merely a proof-of-concept, the detection strategies currently used by our prototype are rather rudimental and have to be improved by further research. The same applies to performance issues. A promising approach for better detection strategies is the aggregation of several IDS into cooperating compounds. Besides that, there also remains a more fundamental problem when implementing wireless intrusion detection systems: in order for them to be effective, they must monitor all channels simultaneously. Since we only want to use standard hardware, this turns out to be a rather difficult task. Possible solutions for this include frequent channel hopping and the implementation of virtual network interfaces. Last but not least, we have concentrated on threat detection and yet neglected the need for (physical) response. We are also aware that our approach is only able to detect previously known attack scenarios. However, the handling of so called unknown attack vectors is beyond the scope of this paper and must be postponed for future research work.

In the near future we intend to include our wireless IDS solution as a basic building block in the collaborative architecture of a self-protecting mobile device [19]. Our main concern in doing so is the establishment of trustful communication paths in the absence of any central instances like trusted third parties.

## References

1. Fluhrer, S., Mantin, I., Shamir, A.: Weaknesses in the Key Scheduling Algorithm of RC4. In: Proceedings of the 8th Annual International Workshop on Selected Areas in Cryptography (SAC 2001). Volume 2259 of LNCS., Springer (2001) 1–24
2. KoreK: The KoreK attack – What FMS conveniently forgot to say. netstumbler.org forum (2004) <http://www.netstumbler.org/showthread.php?t=11869> last visited: February 9, 2006.

3. KoreK: chopchop – Experimental WEP attacks. netstumbler.org forum (2004) <http://www.netstumbler.org/showthread.php?t=12489> last visited: February 9, 2006.
4. Anonymous: Wi-Fi Protected Access: Strong, standards-based, interoperable security for today's Wi-Fi networks. Technical report, Wi-Fi Alliance (2003) [http://www.wifialliance.com/OpenSection/pdf/Whitepaper\\_Wi-Fi\\_Security4-29-03.pdf](http://www.wifialliance.com/OpenSection/pdf/Whitepaper_Wi-Fi_Security4-29-03.pdf) last visited: June, 28 2005.
5. Anonymous: 802.11i – Amendment 6: Medium Access Control (MAC) Security Enhancements. Technical report, Institute of Electrical and Electronics Engineers, Inc. (2004) <http://standards.ieee.org/getieee802/download/802.11i-2004.pdf> last visited: February, 24 2006.
6. Yang, H., Luo, H., Ye, F., Lu, S., Zhang, L.: Security in Mobile Ad Hoc Networks: Challenges and Solutions. *IEEE Wireless Communications* **11** (2004) 38–47
7. Buttyán, L., Hubaux, J.P.: Report on a Working Session on Security in Wireless Ad Hoc Networks. *ACM SIGMOBILE Mobile Computing and Communications Review* **7**(1) (2003) 74–94
8. Zhang, Y., Lee, W., Huang, Y.A.: Intrusion Detection Techniques for Mobile Wireless Networks. *Wireless Networks* **9** (2003) 545–556
9. Lim, Y.X., Schmoyer, T., Levine, J., Owen, H.L.: Wireless Intrusion Detection and Response. In: *Proceedings of the 2003 IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY, USA* (2003)
10. Schmoyer, T.R., Lim, Y.X., Owen, H.L.: Wireless Intrusion Detection and Response. A case study using the classic man-in-the-middle attack. In: *Proceedings of the IEEE Wireless Communications and Networks Conference, Atlanta, Georgia, USA* (2004)
11. Branch, J.W., Jr., N.L.P., van Doorn, L., Safford, D.: Autonomic 802.11 Wireless LAN Security Auditing. *IEEE Security & Privacy* (2004) 56–65
12. Welch, D.J., Lathrop, S.D.: A Survey of 802.11a Wireless Security Threats and Security Mechanisms. Technical Report IOTC-TR-2003-101, Information Technology and Operations Center, Department of Electrical Engineering and Computer Science, United States Military Academy, West Point, New York 10996, USA (2003)
13. Schneier, B.: Modeling security threats. *Dr. Dobbs's Journal* (1999)
14. René Neumerkel: Entwicklung eines Angriffssensors für Wireless LAN. Master's thesis, Technische Universität Dresden (2005)
15. Vladimirov, A., Gavrilenko, K.V., Mikhailovsky, A.A.: WI-FOO. The Secrets of Wireless Hacking. Addison-Wesley Professional (2004)
16. Paxson, V.: Bro: A System for Detecting Network Intruders in Real-time. *Computer Networks* **31**(23–24) (1999) 2435–2463
17. Bellardo, J., Savage, S.: 802.11 Denial-of-Service Attacks: Real Vulnerabilities and Practical Solutions. In: *Proceedings of the 12th USENIX Security Symposium, Washington D.C.* (2003) 15–28
18. Wright, J.: Detecting Wireless LAN MAC Address Spoofing. <http://home.jwu.edu/jwright/papers/wlan-mac-spoof.pdf> (2003) last visited: February, 28 2006.
19. Groß, S.: Selbstschützende mobile Systeme. In: *Sicherheit 2006, Beiträge der 3. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)*. (2006)

# Information Leakage in Ubiquitous Voice-over-IP Communications

Thorsten Neumann, Heiko Tillwink, and Martin S. Olivier

Information and Computer Security Architectures (ICSA) Research Group,  
Department of Computer Science, University of Pretoria, South Africa  
`tozzi@intdev.co.za`, `htillwink@cs.up.ac.za`, `molivier@cs.up.ac.za`

**Abstract.** In VoIP, proxies are used by end-devices to perform a number of tasks including call setup and routing. Setup and routing is achieved through the exchange of call control messages which are forwarded among all involved proxies as well as the communicating end-devices. This paper will explore the information exchanged in Voice-over-IP (VoIP) call control messages and any possible implications this has on personal privacy. We assess the explicit and implicit deductions that can be made from handling messages in transit and evaluate these with a conceptual anonymity model. We aim to show that profiling is a threat in current VoIP implementations and that this threat becomes increasingly relevant with the growing adoption of VoIP. We consider these facts in light of possible future scenarios whereby VoIP has the potential to become a truly ubiquitous technology.

## 1 Introduction

Many organisations are in pursuit to converge their communication networks, allowing for the provisioning of services over a single shared infrastructure. These services, such as voice, video and data, are being transported by packet-switched networks, extending the reach of our global communications infrastructure.

The motivating factors for convergence are the reductions in cost, the continuous innovations allowing for greater service integration and the potential for ubiquitous access and service delivery. However, with these advantages certain privacy concerns surrounding the unification of services into a single *global* network emerge [1].

The growing dependence on technology by society brings with it various privacy issues. More and more people make use of intelligent communication services when performing their day-to-day activities [2]. They knowingly (and unknowingly) transmit large amounts of personal information, thus putting themselves at risk of being monitored.

One technology that has the potential to considerably raise privacy concerns is VoIP, an emergent voice communications technology over the Internet. VoIP will eventually replace our current private-switched telephone network (PSTN).

VoIP is still in its infancy. The implementation of services has not yet matured sufficiently to address the multitude of privacy issues. Details about a call,

such as the participating individuals, are visible to various end-devices, proxies and unauthorised observers. Besides exposing potentially incriminating personal information, individuals also risk having their information being exploited by targeted marketing or insurance companies.

In business, the collection of information for customer relationship management (CRM) and business intelligence (BI) has developed into recognised disciplines. These activities support marketing initiatives in directing and focusing their efforts on particular user segments or individuals. Often, however, available information is averaged to summarise the activities of the collective for specific business purposes.

In VoIP, the analysis of captured private information could similarly be processed. Records can be aggregated to describe the behaviour of a group. Individuals can be monitored allowing users to be profiled. Such profiling makes it possible to determine an individual's activities and habits. Any exploitation of such sensitive information is an obvious infringement of privacy.

Various mechanisms exist that attempt to protect an individual's privacy. Some approaches include using pseudo-identities [3], encrypting sensitive data [4] and information hiding [5]. These privacy-enhancing technologies (PETs) attempt to provide individuals with an acceptable level of privacy.

However, adoption of PETs in VoIP services has received limited attention, largely due to more pressing technical challenges such as voice quality [6], seamless mobility [6] and call management protocols such as SIP [7]. A session management protocol is central to VoIP communication, and many of the privacy concerns relating to VoIP fall back on SIP. Because of SIP's popularity, this paper places specific focus on this protocol.

This paper highlights privacy implications when communicating using VoIP. More specifically, we discuss how private data is leaked when by SIP.

In Section 2 we present background on SIP. Section 3 takes an in-depth look at information leakage by applying the Freiburg Privacy Diamond [8] to show that the exchanged details reduce an individual's anonymity. We show that a communicating individual is not by action, device, location and identity independent. This leads onto section 5 which discusses profiling as an invasion of privacy. Finally, we will conclude with section 6.

## 2 Background

Voice over IP (VoIP) is a general term for any voice communication that is transmitted over the Internet Protocol. This effectively means that voice communication is available to anyone who has access to the Internet and who is using appropriate software.

VoIP commonly distinguishes between two types of a communication: a control channel and a data channel. The data channel is used to transfer the encoded audio stream between two remote parties. The channel is datagram-oriented by design and hence often uses UDP and not TCP. The data channel is set up according to instructions received from the control channel during session

initiation. The control channel, however, ensures that the data channel is established, maintained for the duration of the session and terminated at the end. It is used to exchange messages with the destined remote party, containing details about the source and destination, capabilities of the communicating devices and session information [7]. The control channel is used for, what in traditional telephony, is described as signalling. A protocol commonly used for the control channel is SIP [7]. SIP is the successor to H.323 [1] and has been adopted by the IEEE as the new signalling standard. A more detailed discussion of SIP is therefore appropriate.

An individual, wishing to communicate using a SIP-enabled device would instruct the device to *call* a remote party, identified by either a number or an alias. Gartner predicts (with a probability of 0.9) that users will continue to use traditional numbering in VoIP [9]. This numbering scheme allows for the use of the ITU-T's international public telecommunications numbering plan (E.164) [10] in VoIP. Since devices are no longer bound to physical locations, it allows for the smooth transition from traditional PSTN to VoIP, while ensuring that every device is reachable. Since SIP is designed to work as a distributed architecture, it requires the assistance of intermediaries. It would be impossible to *locate* the remote device without these intermediaries.

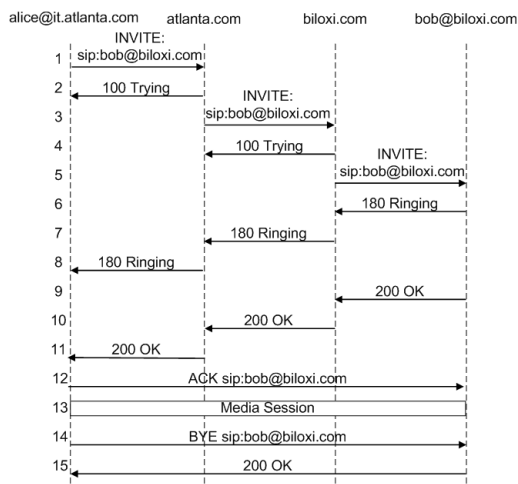


Fig. 1. SIP message exchange

We refer to Fig. 1 to illustrate the steps involved in setting up a SIP session between caller Alice and remote party Bob. SIP initiates a session by sending an INVITE request (step 3). This invite is forwarded by a number of proxies (steps 3,5) until the final proxy is reached. Every proxy is only responsible for its authoritative domain (e.g. biloxi.com), and messages not destined for it are passed on. This effectively allows for a hierarchical structure – for example:

calls destined for Bob working in the human resources (HR) department at a company called Biloxi can be directed to proxy biloxy.com which subsequently forwards the call to proxy hr.biloxi.com. Therefore, proxies dynamically map out a route, from one proxy to another, before the INVITE finally reaches the destination.

This “loose routing” establishes a path which is used for the exchange of subsequent messages. Responses are sent along this path but in the reverse direction. Every proxy, only knows the previous and next proxy. Optimal routes are created, which allow for efficient passing of messages and fail-over mechanisms to ensure sessions are maintained. Call status messages, such as ringing (steps 6–8) and answered (steps 9–11), are back to the calling proxy.

Once the call has been acknowledged (step 11), a data channel is established between the calling and final proxy (step 12). Each proxy will interface with the end-devices; which in our example are operated by Alice and Bob.

Various attributes are exchanged during a session. These attributes are useful to proxies, devices and users, and stored in SIP headers. Required headers are *To*, *From*, *Contact*, *Call-ID* and *Timestamp* values. The *To* and *From* headers are URIs identifying a device or user reachable a domain (e.g. bob@hr.biloxi.com). Additional headers can be used to convey location, alternate contact numbers or device capabilities such as codecs or firmware versions.

Whilst many individuals assume that voice conversations are private, few understand the implications that a signalling protocol has on their privacy. This is understandable as the VoIP environment bear little relation to existing PSTN networks. Telephony operators control the PSTN network, its interconnects to other networks and call routing; unlike the Internet environment.

We pay specific attention to the SIP headers, analyse what information can be acquired and subsequently retrieved from the headers. Furthermore, the inadequacies of the SIP protocol allow intermediate proxies to monitor as well as alter a SIP session. The method in which SIP operates raises concerns over the amount of personal information that is *leaked*.

We investigate the SIP message exchange, in particular SIP headers, in light of the mentioned privacy concerns. We explore what sensitive data is exchanged and how callers can be linked to a device or location. A proxy might have no knowledge about the source or destination, but consider the impact of aggregating messages from multiple proxies and different sources, which could lead to identifying and profiling users.

### 3 Information Leakage

In this section we discuss possible sources of information leakage and particulars visible to intermediaries.

For example, details about a user and his actions can be inferred. This argument is supported by the RFC 3261 [4] which states that “SIP messages frequently contain sensitive information about their senders”. It elaborates on the privacy of users and that it is possible to know with whom, when, how long

and from where a user communicates. While known security threats exist, this section highlights the privacy issues in SIP.

We first discuss the explicit and implicit attributes which are exchanged during a SIP session. We then examine how this can be used to compromise a users privacy in section 5.

### 3.1 Explicit Attributes

SIP exchanges many messages during a session, thus ensuring that engaging parties can continue to communicate. The messages contain explicit attributes which are defined in the protocol. These are connection properties which are exchanged among various entities and across networks. They are stipulated in SIP headers as shown in Fig. 2.

```

INVITE sip:01127117931486@atlas-east.vonage.net SIP/2.0
Via: SIP/2.0/UDP pc33.intdev.co.za;branch=z9hG4bK776asdhd;received=192.0.2.1
Record-Route: <sip:pl.vonage.net;lr>
From: "Thorsten Neumann" <sip:14169079479@atlas-east.vonage.net>;tag=122965585
To: <sip:01127117931486@atlas-east.vonage.net>;tag=28491840-EE2
Call-ID: a84b4c76e66710@192.168.0.120
CSeq: 314159 INVITE
Contact: <sip:tozzzi@intdev.co.za:5060>
User-Agent: <Motorola VT1000 mac: 000F9F466CD0 sw:VT20_1.1.16e ln:0 cfg:1097174/100282>
Content-Type: application/sdp
Content-Length: 142

SIP/2.0 183 Session Progress
Via: SIP/2.0/UDP logging.vonage.net
;branch=z9hG4bKnashds8;received=192.0.2.3
Via: SIP/2.0/UDP pc33.intdev.co.za
;branch=z9hG4bK776asdhd;received=192.0.2.1
From: "Thorsten Neumann" <sip:14169079479@atlas-east.vonage.net>;tag=122965585
To: <sip:01127117931486@atlas-east.vonage.net>;tag=28491840-EE2
Call-ID: a84b4c76e66710@192.0.2.1
CSeq: 2 INVITE
Content-Type: application/sdp
Content-Length: 160

```

**Fig. 2.** SIP Message with Headers

Each device requires an IP address to communicate giving some indication as to its location on the Internet. IP address information revealed in the SIP header does not tie to a particular location, but does bear on a user's locality. It can be established if a user is at work, communicating from corporate domain, on a mobile network or at home using a broadband connection. It can be argued that this information has carries little weight, yet tied to a users pseudo-identity has greater implications.

A user will assume a pseudo-identity, and use it to engage in a VoIP communication. This pseudo-identity is an address in the form of a SIP URI, and comparable an email address, denoted by an *alias@domain*. Devices and intermediaries assisting in the session resolve this address and communicate with the proxy responsible for the *domain*.

The SIP message can contain auxiliary headers that enhance the communication through informative attributes. A user might be reachable at more than one location and provide multiple contact points. This includes *sip*, *mailto* and

*tel* addresses. While the latter is not compulsory, a device must convey how it can be contacted directly [4].

Individuals might want to conceal their name, pseudo-identity or contact points. This becomes increasingly important when SIP messages are sent through numerous intermediaries. The communication for a realm is often controlled by an authoritative proxy, which a user has little control over what is communicated. In order to receive calls the user authenticates to this proxy, thereby confirming his identity and his presents.

Depending on vendor implementations, some devices might inform the proxies of additional device specific functionality. Since SIP is a generic implementation for session management, it allows remote parties to determine a devices capabilities. A device might want to provide additional functionality such as video support, presents information or mobility options. In our research we found that Vonage devices disclose the device model, its MAC address, software version and latest configuration.

Other more subtle deduction can be made by watching the transaction within a session. Next we identify how particulars about a user can be interred from these attributes.

### 3.2 Implicit Attributes

The above listed attributes are communicated in SIP headers. They are explicit and fact, while further implicit properties can be deduced from observing a complete session. Numerous messages are exchanged during a session, as illustrated in Fig. 1, and reveal subtle behavioral attributes. We agree with RFC 3261 which notes there are also less direct ways in which private information can be divulged.

Two important factors are those of time and the duration of a session. Observing SIP messages exchanged at a particular time has a bearing on the users location. A user could have left the office, yet still be communicating thus implying that he or she is possibly at home. Secondly, the progression of a session and its cumulative duration indicate the nature of the call. Many longer calls after work can be assumed to be personal, while those with a duration of less than a minute are most likely work related. This is comparable to the usage patterns found in fixed and mobile phone usage [11, 12] and instant messaging [13].

The final state about a call can be seen in the responses exchanged by devices. SIP response codes are consistent with, and extend, HTTP/1.1 response codes [4] and allow for both machine and human interpretation. These give insight as to how a session was directed or terminated. States such as *Redirected*, *Moved*, *Busy Here*, *Do Not Disturb* or *Rejected* are communicated in these system generated response. These indicate the state of a device or a conscious action of a user.

In the deprecated RFC 2543 (13.3) it is noted that “location and SIP-initiated calls can violate a callee’s privacy”. This includes revealing alternatives can infringe on privacy concerns of the user or the organization.

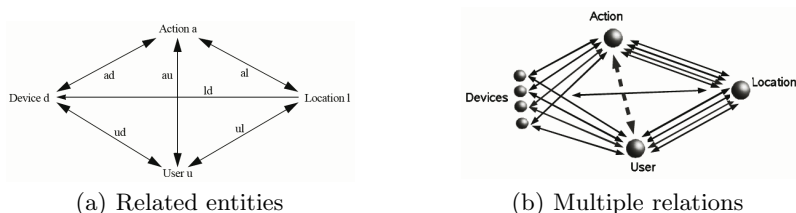
The SIP protocol does not provide sufficient security to protect these attributes transmitted during a session. Messages can be intercepted, inspected, stored or routed without the users consent.



In the following section we assess the implications of how the discussed attributes can be used to infer personal information. The Freiburg Privacy Diamond will be used as a model to show that an attacker can launch an inference attack on a user.

## 4 Freiburg Privacy Diamond

We apply the the Freiburg Privacy Diamond [14] which is a model that can be used to analyze anonymity. This model captures the essence of anonymity with regard to device and user mobility. It considers four entities which impact on the users level of privacy. They are: the action itself, the device, the location of the device, and the user, visually (see Fig. 3(a)). These different entities are related. The user performs an action, using a device at a particular location. In order to achieve anonymity, an attacker should not be able to link these entities when observing a single message, or complete session.



**Fig. 3.** The Freiburg Privacy Diamond

The model has been extended to describe the additional challenges faced in achieving anonymity in pervasive computing [8]. It shows how communicating devices must protect a users privacy through working together in achieving anonymity. Assessing these entities, an attacker would have to reveal the relations between a user and his action to deduce the identity of a user. Depending on the information captured, the attacker could correlate a user to a device or location. Any such relationship would breach the users privacy by revealing the action performed on a device and at a specific location.

A user would be reachable at, or perform actions using a set of possible devices. The user could make use of more than one device. A devices could be mobile (cellphone or softphone) or bound to a location (as is the case with traditional telephones). It is assumed that the user is in close proximity to the device. While this creates an immediate relation between the user and his location, it does not imply that the user can be identified. The semantics of the Freiburg Privacy Diamond require an attacker to determine which user or which device performed the action.

We apply the Freiburg Privacy Diamond to VoIP communication. It is well suited to our research of information leakage during SIP communication. The

flexibility of SIP allows for users to utilise any device and from any location. The SIP headers disclose private information and have notable implications on a users anonymity.

In order for a user to be *contacted*, it must be possible to locate the device being called. Considering that the utilised device will authenticate on the user's behalf, an implicit relation between the user and a device is created, contravening the Privacy Diamond entities. The exchanged information could also reveal the users location.

Two situations arise when a user is *contacted*. In the first scenario the user is contacted, and accepts the incoming call. The SIP session is initiated in which particulars about the session, therefore the user, are exchanged. This includes the users name, direct contact details and device used. Further, particulars about the users current location, presents and availability could be deduced. Redirection instructions such (181 Call Is Being Forwarded, 300 Multiple Choices, 301 Moved Permanently, 302 Moved Temporarily or 380 Alternative Service) communicate this information as part of the response.

An alternate scenario is when the user can not be located or does not accept the incoming call, thereby communicating back a state of a device or a conscious action of a user. If the user is not present at the time (480 Temporarily Unavailable), the resulting SIP headers would reveal alternate contact numbers or locations at which the user could be reached. In contrast, a conscious action would indicate that the user was contacted but unreachable (486 Busy here, 600 Busy everywhere) or declined the call (603 Decline).

Reverting to the Freiburg Privacy Diamond, the user can therefore be tied to the action, and can be associated with a device and possibly a location. Further assumptions can be made through observing the session, and the extracting the implicit attributes.

## 5 Profiling

We consider profiling of a VoIP user and the possible privacy implications thereof. The Freiburg Privacy Diamond provides a model through we we have shown that a users privacy is at risk. The Voice over IP Security Alliance [15] remarks that VoIP “faces different threats than other Internet applications, triggering unique security and privacy concerns.” Profiling in VoIP is the process of analysing personal information found in call data. We have introduced explicit and implicit attributes as two sources of personal information found in call data.

During the establishment of session, a proxy could unknowingly to the caller insert a *Record-Route* header. This instructs the participating devices to relay subsequent SIP messages through the proxy for the duration of the SIP session. The host specified in the *Record-Route* header need not be the proxy handling the SIP message. An simple example to illustrate the risk of information leakage is where *eve.com* would forward the SIP INVITE with this additional *Record-Route* header. While *eve.com* should no longer play an role in the session, the proxy will receive all messages and event updates exchanged between the communicating

parties. As indicated in Fig. 1, neither Alice nor Bob are aware of this intermediary.

SIP devices and proxies additionally rely on the *Route* directive to pass messages to specific hosts for processing and routing. A misconfigured or compromised proxy can manipulate messages without consent from the user, such as injecting additional headers. The SIP header will force the message to be forwarded to a specific intermediary before reaching the intended destination.

The possibility exists where *eve.com* inserts a *Route* instruction to have the current SIP message forwarded to *profiling.com*. This allows the next en route proxy to collect the Explicit Attributes described in Section 3.1. Further, one could consider this in combination with the aforementioned *Record-Route* header. This gives *profiling.com* the ability to monitor and profile the user, correlating the actions and ability to deduce the implicit attributes described in Section 3.2.

The users of a SIP session are not in control over the communication environment, often restricted to the interface of the device (or softphone). The communicating parties might not be aware of intermediaries logging and recording call control messages. While the mentioned records are specific to call control events, they expose a great amount of detail about a user.

With the growing adoption of VoIP profiling becomes an increasingly dangerous threat. Analysing a collection of calls performed or received by an individual could expose a substantial amount of information about a user's behaviour, habits or preferences. Whilst these threats are currently minor, one should consider a case where VoIP becomes a truly ubiquitous communications technology.

One could consider the case whereby many household, workplace and public devices are networked and support IP communications. Not every device needs to be a communications device. It could be used to inform an individual if his phone is ringing or if messages are available. If this were the case, more personal information would be available.

Further research is required to study the implication that a widespread acceptance of VoIP has on personal privacy. An interesting case is a probable future scenario whereby communication is possible from anywhere and by anybody using his own unique pseudo-identity or telephone number. We have only briefly touched on the implications hoping to stimulate ongoing discussions.

## 6 Conclusion

The aims to show that information leakage in VoIP, and specifically for SIP, affects a user's privacy. Personal details about the user are exposed, thus compromising a user's anonymity. Information about a user's action, the device used, location and identity can be correlated. A user is therefore not assured a sufficient level of privacy when communicating over the Internet.

We identified what the information is revealed when communicating with a remote device and discussed implicit attributes that can be deduced from this. The Freiburg Privacy Diamond [14] was used to support our argument.

Future research will assess viable methods of ensuring privacy and anonymity. Research surrounding security mechanisms to prevent the misdirection messages and manipulation of SIP Headers have been suggested [16]. This does, however, require the collaboration of multiple devices which must strive to protect the user's identity.

The trends indicate that the VoIP will increasingly dominate cable telephony, and start replacing traditional telephone lines [1]. This raises concerns about a users privacy as this pervasive technology starts replacing our existing communications infrastructure.

## References

- [1] Phil Sherburne and Cary Fitzgerald: You Don't Know Jack About VoIP. *Queue* **2**(6) (2004) 30–38
- [2] Weiser, M.: The Computer for the 21st Century. *Scientific American Ubicomp* **3** (1991) 94–104
- [3] Peterson, J., Jennings, C.: Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP), RFC 3323 (2003)
- [4] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.: SIP: Session Initiation Protocol, RFC 3261 (2002)
- [5] Peterson, J.: A Privacy Mechanism for the Session Initiation Protocol (SIP), RFC 3323 (2002)
- [6] Varshney, U., Snow, A., McGivern, M., Howard, C.: Voice over IP. *Commun. ACM* **45**(1) (2002) 89–96
- [7] Schulzrinne, H., Rosenberg, J. In: *The Session Initiation Protocol: Internet-centric signaling*. Volume 38., IEEE (2000) 134–141
- [8] Zugenmaier, A., Kreuzer, M., Müller, G.: The freiburg privacy diamond: An attacker model for a mobile computing environment. In: *KiVS Kurzbeiträge*. (2003) 131–141
- [9] Fraley, D.L.: Voice Over IP Communications Must Be Secured. Gartner, Inc. (G00124016) (2004) 5 of 6
- [10] Faltstrom, P.: E.164 number and DNS. RFC 2916 (1998)
- [11] Palen, L., Salzman, M., Youngs, E.: Going wireless: behavior & practice of new mobile phone users. In: *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work, USA*, ACM Press (2000) 201–210
- [12] Hindus, D., Schmandt, C.: Ubiquitous audio: capturing spontaneous collaboration. In: *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work, New York, NY, USA*, ACM Press (1992) 210–217
- [13] Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D.J., Kamm, C.: The character, functions, and styles of instant messaging in the workplace. In: *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work, NY, USA*, ACM Press (2002) 11–20
- [14] Zugenmaier, A.: The Freiburg Privacy Diamond - A Conceptual Model for Mobility in Anonymity Systems. In: *Proceedings of Globecom*. (2003)
- [15] Alfonsi, B.: Alliance addresses VoIP security. *IEEE Security & Privacy* **3**(4) (2005) 8
- [16] T. Neumann and Martin S Olivier: Enhancements to SIP to prevent abuse of Voice-over-IP services. In: *Southern African Telecommunication Networks and Applications Conference (SATNAC) Proceedings*. (2005)

# Author Index

- Alam, Muhammad 142
- Balopoulos, Theodoros 62
- Beyer, Anja 162
- Böhme, Rainer 31
- Boyd, Colin 213
- Brandi, Wesley 81
- Breu, Ruth 142
- Casassa Mont, Marco 1, 91
- Dritsas, Stelios 103
- Du, Rong 213
- Eloff, J.H.P. 182
- Fernández-Medina, Eduardo 51
- Foo, Ernest 213
- Gritzalis, Dimitris 103
- Gritzalis, Stefanos 62
- Groß, Stephan 223
- Gymnopoulos, Lazaros 62
- Hafner, Michael 142
- Höhn, Sebastian 114
- Jezierski, Juliusz 172
- Karyda, Maria 62
- Kataria, Gaurav 31
- Katsikas, Sokratis 62
- Klonowski, Marek 192
- Kokolakis, Spyros 62
- Koutrouli, Eleni 152
- Kubiak, Przemysław 192
- Kutyłowski, Mirosław 192
- Kwon, Saeran 203
- Lauks, Anna 192
- Lee, Sang-Ho 203
- Li, Zude 132
- Morzy, Mikołaj 172
- Neumann, Thorsten 233
- Neumerkel, René 223
- Nikolaou, Christos 11
- Nowey, Thomas 41
- Nützel, Jürgen 162
- Olivier, Martin S. 81, 123, 233
- Pearson, Siani 91
- Piattini, Mario 51
- Poursalidis, Vassilis 11
- Rodríguez, Alfonso 51
- Roßnagel, Heiko 71
- Schläger, Christian 41
- Thyne, Robert 91
- Tillwick, Heiko 233
- Tsalgatidou, Aphrodite 152
- Tsaparas, John 103
- Unterthiner, Stefan 142
- van Staden, Wynand 123
- Venter, H.S. 182
- Wang, Jianmin 132
- Wawrzyniak, Dariusz 21
- Wojcik, M. 182
- Ye, Xiaojun 132
- Zhan, Guoqiang 132