Steven Furnell
Costas Lambrinoudakis
Javier Lopez (Eds.)

# Trust, Privacy, and Security in Digital Business

LNCS 8058

10th International Conference, TrustBus 2013
Prague, Czech Republic, August 2013
Proceedings

Lecture Notes in Computer Science 8058

*Commenced Publication in 1973*
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Steven Furnell Costas Lambrinoudakis
Javier Lopez (Eds.)

# Trust, Privacy, and Security in Digital Business

10th International Conference, TrustBus 2013
Prague, Czech Republic, August 28-29, 2013
Proceedings

Springer

Volume Editors

Steven Furnell
Plymouth University
Centre for Security Communications and Network Research
Plymouth PL4 8AA, UK
E-mail: sfurnell@plymouth.ac.uk

Costas Lambrinoudakis
University of Piraeus
Department of Digital Systems
18532 Piraeus, Greece
E-mail: clam@unipi.gr

Javier Lopez
University of Malaga
Computer Science Department
29071 Malaga, Spain
E-mail: jlm@lcc.uma.es

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-40342-2 e-ISBN 978-3-642-40343-9
DOI 10.1007/978-3-642-40343-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013944915

CR Subject Classification (1998): D.4.6, K.6.5, E.3, K.4.4, H.2.7, C.2, H.4, J.1

LNCS Sublibrary: SL 4 – Security and Cryptology



*Typesetting:* Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

This book presents the proceedings of the 10th International Conference on Trust, Privacy and Security in Digital Business (TrustBus 2013), held in Prague, Czech Republic, during August 28–29, 2013. The conference continued from previous events held in Zaragoza (2004), Copenhagen (2005), Krakow (2006), Regensburg (2007), Turin (2008), Linz (2009), Bilbao (2010), Toulouse (2011) and Vienna (2012).

The recent advances in Information and Communication Technologies (ICT) have raised new opportunities for the implementation of novel applications and the provision of high-quality services over global networks. The aim is to utilize this "information society era" for improving the quality of life for all citizens, disseminating knowledge, strengthening social cohesion, generating earnings and finally ensuring that organizations and public bodies remain competitive in the global electronic marketplace. Unfortunately, such a rapid technological evolution cannot be problem free. Concerns are raised regarding the "lack of trust" in electronic procedures and the extent to which "information security" and "user privacy" can be ensured.

TrustBus 2013 brought together academic researchers and industry developers who discussed the state of the art in technology for establishing trust, privacy, and security in digital business. We thank the attendees for coming to Prague to participate and debate the new emerging advances in this area.

The conference program included six technical papers sessions that covered a broad range of topics, from access control and authentication to privacy and confidentiality management, and from identity and trust management to trust and privacy in mobile and pervasive environments. The conference attracted many high-quality submissions, each of which was assigned to four referees for review and the final acceptance rate was 34%.

We would like to express our thanks to the various people who assisted us in organizing the event and formulating the program. We are very grateful to the Program Committee members and the 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.

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.

August 2013

Steven Furnell
Costas Lambrinoudakis
Javier Lopez

# Organization

**General Chair**

| | |
|---|---|
| Javier Lopez | University of Malaga, Spain |

**Program Committee Co-chairs**

| | |
|---|---|
| Steven Furnell | Plymouth University, UK |
| Costas Lambrinoudakis | University of Piraeus, Greece |

**International Program Committee**

| | |
|---|---|
| Agudo, Isaac | University of Malaga, Spain |
| Casassa Mont, Marco | HP Labs Bristol, UK |
| Chadwick, David W | University of Kent, UK |
| Chu, Cheng-Kang | I2R, Singapore |
| Clarke, Nathan | University of Plymouth, UK |
| Cuppens, Frederic | ENST Bretagne, France |
| De Capitani di Vimercati, Sabrina | Università degli Studi di Milano, Italy |
| Fernandez, Eduardo B. | Florida Atlantic University, USA |
| Fernandez-Gago, Carmen | University of Malaga, Spain |
| Fischer-Huebner, Simone | Karlstad University, Sweden |
| Foresti, Sara | Università degli Studi di Milano, Italy |
| Fuss, Juergen | University of Applied Sciences Upper Austria at Hagenberg, Austria |
| Geneiatakis, Dimitris | University of Piraeus, Greece |
| Gritzalis, Dimitris | Athens University of Economics and Business, Greece |
| Gritzalis, Stefanos | University of the Aegean, Greece |
| Hansen, Marit | Independent Center for Privacy Protection Schleswig-Holstein, Germany |
| Jøsang, Audun | Oslo University, Norway |
| Kalloniatis, Christos | University of the Aegean, Greece |
| Karyda, Maria | University of the Aegean, Greece |
| Katsikas, Sokratis | University of Piraeus, Greece |
| Kesdogan, Dogan | University of Siegen, Germany |
| Kokolakis, Spyros | University of the Aegean, Greece |
| Lioy, Antonio | Politecnico di Torino, Italy |
| Markowitch, Olivier | Université Libre de Bruxelles, Belgium |
| Martinelli, Fabio | CNR, Italy |
| Matyas, Vashek | Masaryk University, Czech Republic |
| Mitchell, Chris | Royal Holloway, University of London, UK |

| | |
|---|---|
| Mouratidis, Haralambos | University of East London, UK |
| Olivier, Martin S. | University of Pretoria, South Africa |
| Oppliger, Rolf | eSecurity Technologies, Switzerland |
| Papadaki, Maria | Plymouth University, UK |
| Pashalidis, Andreas | Katholieke Universiteit Leuven, Belgium |
| Patel, Ahmed | Kingston University, UK) - University Kebangsaan, Malaysia |
| Pernul, Guenther | University of Regensburg, Germany |
| Posegga, Joachim | University of Passau, Germany |
| Preneel, Bart | Katholieke Universiteit Leuven, Belgium |
| Quirchmayr, Gerald | University of Vienna, Austria |
| Rajarajan, Muttukrishnan | City University, UK |
| Rizomiliotis, Panagiotis | University of the Aegean, Greece |
| Roman, Rodrigo | I2R, Singapore |
| Rudolph, Carsten | Fraunhofer Institute for Secure Information Technology, Germany |
| Ruland, Christoph | University of Siegen, Germany |
| Samarati, Pierangela | Università degli Studi di Milano, Italy |
| Schaumueller-Bichl, Ingrid | University of Applied Sciences Upper Austria at Hagenberg, Austria |
| Schunter, Matthias | Intel Labs Europe, Germany |
| Theoharidou, Marianthi | Athens University of Economics and Business, Greece |
| Tsohou, Aggeliki | Brunel University, UK |
| Teufel, Stephanie | University of Fribourg, Switzerland |
| Tjoa, A Min | Technical University of Vienna, Austria |
| Tomlinson, Allan | Royal Holloway, University of London, UK |
| Weippl, Edgar | SBA Research and Vienna University of Technology, Austria |
| Xenakis, Christos | University of Piraeus, Greece |

## External Reviewers

| | |
|---|---|
| Darra, Eleni | University of Piraeus, Greece |
| Drogkaris, Prokopios | University of the Aegean, Greece |
| Mylonas, Alexios | Athens University of Economics and Business, Greece |
| Ntantogian, Christoforos | University of Piraeus, Greece |
| Pitropakis, Nikolaos | University of Piraeus, Greece |
| Soupionis, Yannis | Athens University of Economics and Business, Greece |
| Vemou, Konstantina | University of the Aegean, Greece |
| Yfantopoulos, Nikolaos | University of Piraeus, Greece |

# Table of Contents

## Session 1: Access Control and Authentication



## Session 2: Identity and Trust Management



## Session 3: Pivacy and Confidentiality Management



## Session 4: Information Systems Security

## Session 5: Security Policies/Legal Issues



## Session 6: Trust and Privacy in Mobile and Pervasive Environments

# Improving Kerberos Ticket Acquisition during Application Service Access Control

Fernando Pereñiguez-Garcia, Rafael Marin-Lopez,
and Antonio F. Skarmeta-Gomez

Faculty of Computer Science, University of Murcia,
Murcia, E-30100, Spain
{pereniguez,rafa,skarmeta}@um.es

**Abstract.** Kerberos is one of the most deployed protocols to achieve a controlled access to application services by ensuring a secure authentication and key distribution process. Given its growing popularity, Kerberos is envisaged to become a widespread solution for *single sign-on* access. For this reason, the evolution of the protocol still continues in order to address new features or challenges which were not considered when initially designed. This paper focuses on the ticket acquisition process and proposes a new mechanism called *Kerberos Ticket Pre-distribution* that reduces the time required to recover tickets from the *Key Distribution Center* (KDC). We offer a flexible solution which is able to work in three different modes of operation, depending on what entity (the user, the network or both) controls the pre-distribution process. By employing the extensibility mechanisms available in Kerberos, we maintain interoperability with current implementations without compromising the security and robustness of the protocol. Using an implemented prototype, we evaluate our solution and demonstrate that our proposal significantly improves the standard Kerberos ticket acquisition process.

**Keywords:** Ticket pre-distribution, Kerberos, Access Control.

## 1 Introduction

Controlling subscribers' access to multiple application services by a single authentication process (commonly named *single sign-on* access) is gaining enormous interest in the Internet nowadays. In this sense, Kerberos [1] is becoming one of the most widely deployed standards for authentication and key distribution providing this feature [2,3]. Indeed, operating systems (Windows, Linux, etc.) and different network applications (FTP, SSH, HTTP, etc...) already integrate Kerberos to perform service access control. According to the Kerberos operation, a subscriber (Kerberos *client*) requests several credentials so-called *tickets* from a special server named *Key Distribution Center* (KDC) by means of a single initial authentication. These tickets are useful to access different application servers, which enforce a service access control based on Kerberos.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 1–12, 2013.

Despite Kerberos is a prominent candidate to support application service access control, it lacks an optimized mechanism to recover the credential (i.e. ticket) necessary to access a service. In practice, a Kerberized client is required to always contact the KDC per each accessed service. Since this KDC is typically placed in the core network managed by the service provider, which is far from the user's location, additional latency is added to the ticket recovery process and, consequently, to the overall service access time. This is a serious problem in some scenarios (e.g. those related to the network access service [4]) where reducing service access time is an essential requirement to increment the quality of the service experimented by users.

The credential acquisition optimization is a research area that has been successfully addressed in other research fields [5,6,7] by means of the *credential pre-distribution* concept. The idea is to pre-distribute several credentials (i.e. shared keys) by using a single contact with some network entity in charge of the credential distribution process. Nevertheless, to the best of our knowledge, there no exist previous works applying the credential pre-distribution concept to improve the ticket recovery process in Kerberos.

Thus, this paper describes a solution for this integration called *Kerberos Ticket Pre-distribution*, which reduces the latency and the signaling involved when a user wants to access several services using Kerberos. In particular, our solution defines a novel mechanism for Kerberos allowing to recover a set of service tickets to access multiple services in one single contact with the KDC. To accomplish this goal, we propose three modes of operation, which adapt themselves to different paradigms in current networks: either the user takes control of its actions or the network instructs the user; or both entities have an active participation. So that, we define *User-Controlled*, *Server-Controlled without Suggestions* and *Server-Controlled with Suggestions* modes of operation.

In the *User-Controlled* mode, the user selects beforehand the services and requests service tickets for all of them in a single exchange with the KDC. The KDC then will create as many tickets as requested services and send them back to the user in one single message. When the user may not know all the services which are available within a particular network, we propose the *Server-Controlled without Suggestions* mode. In this case, the user requests only one ticket to the KDC for a particular (known) service, and the KDC will answer not only with the service ticket for that service but also with a set of service tickets which are valid to access services related somehow with the requested service (e.g. DHCP service is related to network access service). Alternatively, in the *Server-Controlled with Suggestions* mode, the KDC may first suggest a list of services (instead of sending the set of service tickets) that could be interesting for the user, taking into account the initial service requested by the user. Thus, the user can decide whether to request a subset of the list or, even all of them, in an additional contact with the KDC. This mode adds more flexibility and provides an hybrid mode of operation.

The remainder of the paper is organized as follows: in section 2 we analyze Kerberos, since it is the basis where our solution stems from. Section 3 describes

the proposal in detail, how the modes operate and describe an example of use case. Section 4 shows some performance results and, finally, section 5 concludes the paper and gives some guidelines for future work.

## 2 Kerberos Authentication Protocol: General Overview

Kerberos [1] is a protocol for key distribution based on symmetric cryptography. There are three entities involved during a Kerberos protocol excecution; a *client* representing a user, an *application service* providing a service, and a *Key Distribution Center* (KDC). The KDC provides an *Authentication Server* (AS) and a *Ticket Granting Server* (TGS). Kerberos requires the pre-establishment of a shared secret key between client ⇔ AS and TGS ⇔ application service.

In Kerberos, a *session* key is generated by the KDC and distributed to the user in order to establish a security association with the application service. The user presents the session key contained in a *ticket* to the application service. A Kerberos ticket is a record generated by the KDC to help the user to authenticate itself against an application service. It contains the identity of the user, the session key, a timestamp and other information, which are encrypted by using the secret key shared between the KDC and the application service.

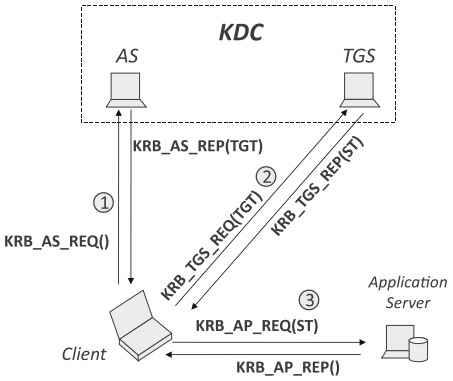



**Fig. 1.** Kerberos signalling

Kerberos operation involves three main exchanges, as shown in Fig. 1. Initially, the user requests a *Ticket Granting Ticket* (TGT), which is a special ticket used for generating other tickets, to the AS through a KRB_AS_REQ/KRB_AS_REP exchange (*1*). The AS generates a TGT, which contains a session key for the TGS (a *TGS session key*), and sends the TGT to the user together with a copy of the TGS session key, which is encrypted with the secret key shared by the KDC and the user.

Then, in a second KRB_TGS_REQ/KRB_TGS_REP exchange (*2*), the user sends the application service's identity and the TGT to the TGS , together with

a credential generated using the TGS session key, so that the TGS can verify that the user possesses the same TGS session key. After successful verification of the credential, the TGS generates a *service ticket* (ST) which contains a session key for the application service and sends the ticket to the user with a copy of the session key, which is encrypted with the TGS session key. In the third exchange KRB_AP_REQ/KRB_AP_REP (*3*), the user sends the ST, together with a message authentication code computed by the user, so that the application service can verify that the user possesses the same session key as the one contained in the ST. After successful verification of the credentials, the user and the application service are able to use the session key to protect their application protocol.

It is important to note that, KRB_AS_REQ/KRB_AS_REP is only performed when the user does not already have a valid TGT. If the user already owns a TGT, it only needs to perform the KRB_TGS_REQ/KRB_TGS_REP exchange to obtain a ST and KRB_AP_REQ/KRB_AP_REP exchange to access the service with the ST. Thus, both exchanges (*2* and *3*) are performed each time a user wants to access an application service for the first time.

## 3 Ticket Pre-distribution in Kerberos

In the following we present the ticket pre-distribution mechanism designed for optimizing access control to Kerberized services. Apart from presenting the operation modes supported by our solution, a example of use case is described in order to demonstrate the applicability of our proposal.

### 3.1 Design

In order to support our proposal for Kerberos ticket pre-distribution, we have extended the Kerberos protocol in different ways without impacting the standard. We employ the extensibility mechanisms already available in Kerberos, like definition of new flags and pre-authentication data types (hereafter *padata*), so that the proposed solution respects the core Kerberos specification [1] without affecting the security. In other words, the proposal does not change the semantics of messages or define new types of messages, which enables the interoperability with existing Kerberos implementations.

In particular, only the KRB_TGS_REQ and KRB_TGS_REP messages uses our new extensions. In particular, we extend the *kdc-options* field with a new *K*-flag to indicate that ticket pre-distribution is requested and supported. Additionally, we define a new set of *padatas* [1] which can be contained in the *padata* field of both messages.

- **PA-SERVICE-NAME**. It contains the name of a service for which the client requests the pre-distribution of a ticket.
- **PA-SERVICE-TICKET**. It contains the pre-distributed ticket as well as the information needed by the client to use it (e.g. key, ticket lifetime, etc.).

- **PA-FLAGS**. It contains three new flags: *M-flag* (User-Controlled support), *N-flag* (Server-Controlled without Suggestions support) and *NS-flag* (Server-Controlled with Suggestions support).
- **PA-ERROR**. It reports error situations when requesting a pre-distributed ticket. For example, if the user solicits the pre-distribution of a service ticket the KDC is not able to pre-distribute for some reason, the KDC includes a PA-ERROR to report the problem.

These are the basic components that we require to define the three modes of operation in our Kerberos ticket pre-distribution solution. In the following, we give details about how these modes operate and how these basic components are used. The following notation will be used for describing our solution:

- $TGT_X$: TGT for KDC X.
- $ST_X$: service ticket for service X.
- *[*$X_1$*,...,*$X_i$*,...,*$X_n$*]* : $X_i$ is padata type included in the padata field.
- *PA-SN[X]*: Abbreviation of PA-SERVICE-NAME with value X.
- *PA-ST[X]*: Abbreviation of PA-SERVICE-TICKET with value X.
- *PA-FL[X]*: Abbreviation of PA-FLAGS with *X*-flag activated.
- *PA-ER[X]*: Abbreviation of PA-ERROR, reporting an error for service X.

### 3.2 Modes of Operation

We give now the details about the three modes of operations. Without loss of generality, we provide an example with three services S1, S2 and S3 to describe each mode. In the examples, we assumed that the user already has a TGT for the KDC controlling the services. In other words, both user and KDC already participated in a KRB_AS_REQ/KRB_AS_REP exchange.

**User-Controlled Pre-distribution.** In this mode of operation, the user is in charge of selecting the set of services to which it desires access. In this sense, the user indicates the list of services in the extended version of KRB_TGS_REQ. In Fig. 2, we show an example where the user wants to access to services S1, S2 and S3 by using this mode. The process consists on the following steps.

By using the TGT, the user sends a KRB_TGS_REQ and requests not only a ST for S1 ($ST_{S1}$) as the standard Kerberos specifies, but also for the services S2 and S3. Specifically, the user activates the new defined *K*-flag; it includes a PA-FLAGS with the *M*-flag activated (PA-FL[M]); and two padatas PA-SN[S2], PA-SN[S3] which indicates the additional services for which a ST is being requested *(1)*. The KDC answers with a KRB_TGS_REP which includes the $ST_{S1}$. Moreover, it includes in the padata field a PA-FL[M] and two padatas PA-ST[$ST_{S2}$] and PA-ST[$ST_{S3}$], which contain the service ticket for S2 and S3, respectively *(2)*. Note that the PA-ST carries out both the ticket and the associated information needed by the client to employ the service ticket (e.g. session key), which is protected following the standard Kerberos operation (i.e. using the TGS session key). If for some reason, there is a problem with delivering,

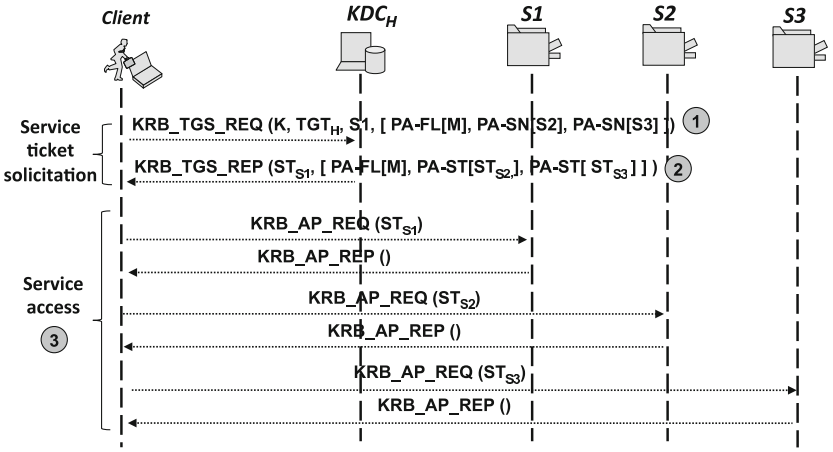



**Fig. 2.** User-Controlled Pre-distribution

for example, the $ST_{S3}$, the KDC includes a PA-ER[S3] by informing that it was not able to deliver a ticket for that service.

Once the user has $ST_{S1}$, $ST_{S2}$ and $ST_{S3}$, it can engage a standard KRB_AP_REQ/KRB_AP_REP with each service S1, S2 and S3 respectively *(3)*.

**Server-Controlled Pre-distribution without Suggestions.** In this mode, the KDC can deliver a set of service tickets for different services, apart from the service ticket requested by the user. The user does not really specify these extra services but just requests a service ticket for a single service, following the standard Kerberos. The Fig. 3 shows an example using the *Server-Controlled without Suggestions*. Basically, the user requests a $ST_{S1}$ for the service S1 by using the TGT by sending a KRB_TGS_REQ with the $K$-flag activated. Moreover, the user adds a PA-FL[N] to indicate that it is able to accept the use of *Server-Controlled without Suggestions* mode *(1)*.

When the KDC processes the KRB_TGS_REQ, it notices that user supports ticket pre-distribution ($K$-flag activated) and requests *Server-Controlled without Suggestions* (PA-FL[N]). Then the KDC provides the $ST_{S1}$ and determines that services S2 and S3 which could be accessed by the user in the near future[1]. The KDC can determine this based on, for example, the current point of attachment of the network and the nature of service S1. The policy used by the KDC to select the extra services is out of scope of this work.

Thus, the KDC adds two padatas PA-ST[$ST_{S2}$], PA-ST[$ST_{S3}$] in the padata field of the KRB_TGS_REP *(2)*. With $ST_{S1}$, $ST_{S2}$ and $ST_{S3}$, the user can access the services S1, S2 and S3, respectively *(3)*.

[1] The concrete policy followed by the KDC to select the services S2 and S3 is out-of-scope.

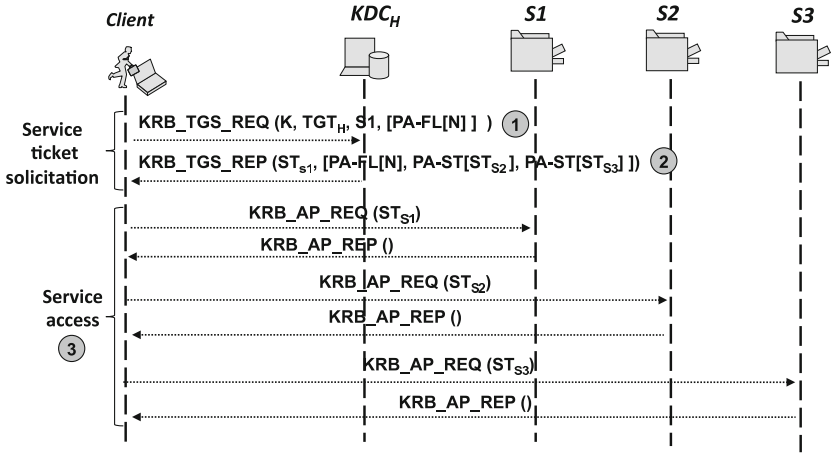



**Fig. 3.** Server-Controlled Pre-distribution without Suggestions

**Server-Controlled Pre-distribution with Suggestions.** This mode of operation represents an hybrid model between both modes previously presented. The KDC sends a list of available services (or *suggested* services) that could be useful to the user, but it does not send any service ticket. Then the user with that information can engage a *User-Controlled* mode with the KDC requesting a subset (or even all of them) of the suggested services.

Fig. 4 shows an example with this mode. Unlike previous two models, this mode involves two KRB_TGS_REQ/KRB_TGS_REP exchanges. In the first exchange, the user requests a service ticket $ST_{S1}$ for S1, enables the $K$-flag and adds a PA-FL[NS] to the padata field in the KRB_TGS_REQ *(1)*. In this way, the user informs that it solicits *Server-Controlled with Suggestions*. Therefore, the KDC will answer with a KRB_TGS_REP containing the $ST_{S1}$ but also adds two padatas PA-SN[S2], PA-SN[S3] which suggests the user about the presence of services S2 and S3 that the user may be interested to access *(2)*. Similarly to previous mode, the procedure used by the KDC to select the suggested services is out of scope of this work. With this information, the user sends another KRB_TGS_REQ with *User-Controlled* mode activated (PA-FL[M])) by soliciting a service ticket for S2 ($ST_{S2}$) and an additional ticket for S3 (PA-ST[$ST_{S3}$]) *(3)*. Then the KDC answers with the KRB_TGS_REP with the $ST_{S2}$ and a padata PA-SERVICE-TICKET containing the $ST_{S3}$ (PA-ST[$ST_{S3}$]) *(4)*. Finally, the user can access the services S1, S2 and S3 by using $ST_{S1}$, $ST_{S2}$ and $ST_{S3}$, respectively *(5)*.

Although this mode may involve an additional exchange with the KDC, it allows the user to select any of the suggested services provided for the KDC, providing a higher level of flexibility to the user. Note that, in certain cases, the user may not use any of the suggestions and the second KRB_TGS_REQ/KRB_TGS_REP is simply not required.

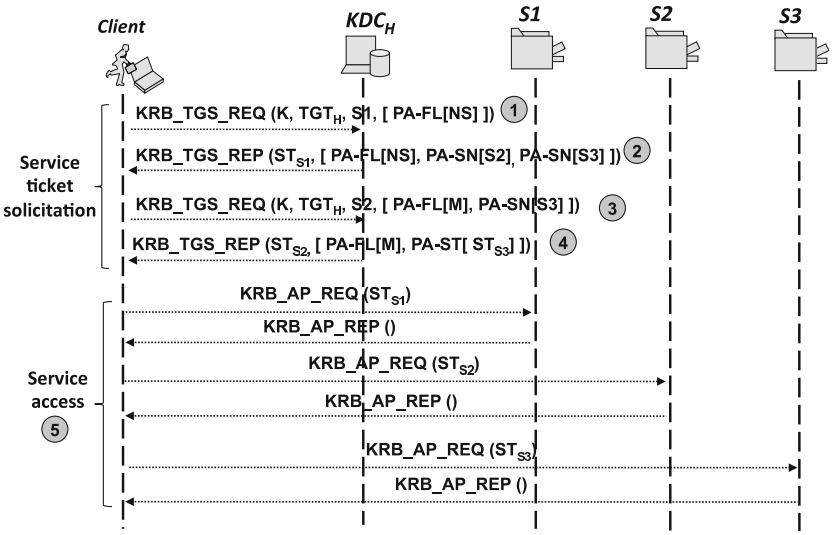



**Fig. 4.** Server-Controlled Pre-distribution with Suggestions

### 3.3 Example of Use Case: Network Access Service

One interesting scenario of applicability of our solution is related to the specific use of the network access service. In this case, the network service is provided by a point of attachment to the network, such as an access point or an access router. Generally, the point of attachment performs an authentication process before providing network access. A widely used protocol to carry out this authentication is the *Extensible Authentication Protocol* (EAP) [8]. However, EAP usually takes a significant time to complete, which can be specially counterproductive in scenarios where mobility and seamless handoffs are required.

Authors in [4] have proposed the use of Kerberos to perform a fast EAP re-authentication procedure. Basically, the mobile (the *user*) recovers from a KDC, service tickets to access different access points (APs), which act as application services. These service tickets are transported between the mobile and the APs by using a new EAP method named EAP-FRM [9]. As shown in Fig. 5, our extensions for Kerberos ticket pre-distribution can be applied to recover several service tickets with a single KRB_TGS_REQ/KRB_TGS_REP exchange *(1)*, performed for the first time that the mobile connects to the network. With these service tickets, the mobile can move through several APs (*2,3*) without contacting the infrastructure by just presenting the service ticket to the AP and reducing, as a consequence, the time to access the network. In the context of the Walkie-Talkie project [10], authors in [11] are using EAP over IKEv2 to provide IPsec access control in a pre-WiMax campus scenario, considering a vehicle as the mobile node. Specially in this case, fast re-authentication is even more important due to the higher velocity of the vehicle. The integration of Kerberos ticket pre-distribution has been envisaged

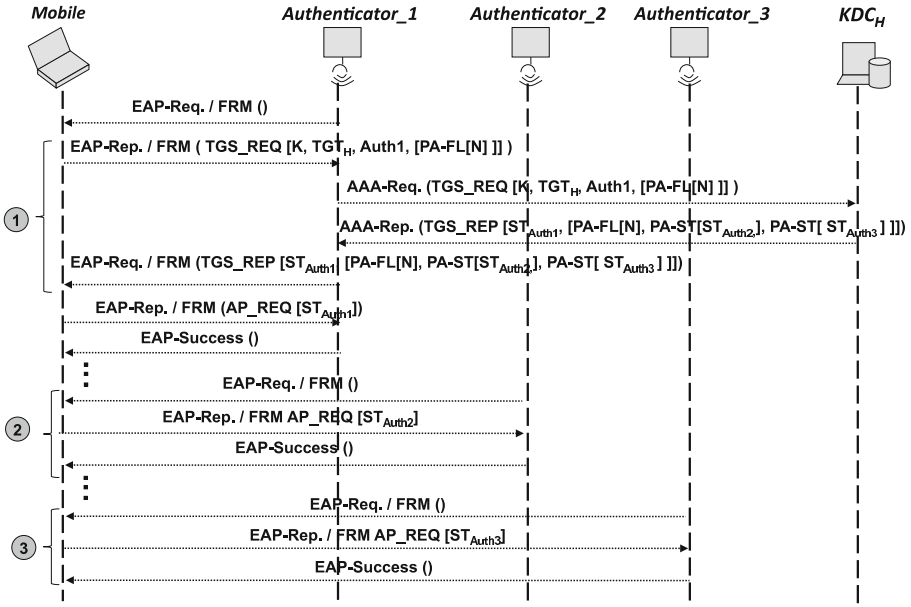



**Fig. 5.** Example of User-Controlled Pre-distribution for Network Access Service

as a promising solution to reduce the network access time in this type of scenario. The details about this integration are motive of future work.

## 4 Performance Results

In order to verify the feasibility of our approach, we have implemented the different modes and associated extensions required for the three mode of operations proposed. This implementation has been based on the MIT Kerberos implementation version 1.6.3 [12]. By using the API provided by the Kerberos implementation, we have created two programs which act as the user and the service, respectively. We have also deployed a KDC to support the kerberized authentication. It is worth noting though that we have mainly paid attention to the exchanges between the user and the KDC, since only KRB_TGS_REQ/KRB_TGS_REP exchanges require our extensions.

We have performed ≈500 Kerberos authentications with standard Kerberos protocol and each of our three modes of operation. In this way, we compare the impact of our extensions to support Kerberos ticket pre-distribution in contrast with standard operation of Kerberos. We have tested with the distribution of service tickets for 2, 5 and 10 services (when the user is just interested in accessing 1 service, standard Kerberos is enough and our modes are simply not used). We have used wireshark tool [13] to obtain different traces of the involved exchanges. With this information, we have obtained a mean value for $T_{AS}$, which is the time to perform KRB_AS_REQ/KRB_AS_REP; $T_{TGS}$, which is the time

to perform KRB_TGS_REQ/KRB_TGS_REP and $T_{AP}$, which is the time to perform KRB_AP_REQ/KRB_AP_REP exchange. Moreover, we have measured the number of bytes involved in the KRB_TGS_REQ/KRB_TGS_REP exchanges sent to the network to recover all the service tickets, which give us a hint about the bandwidth consumption.

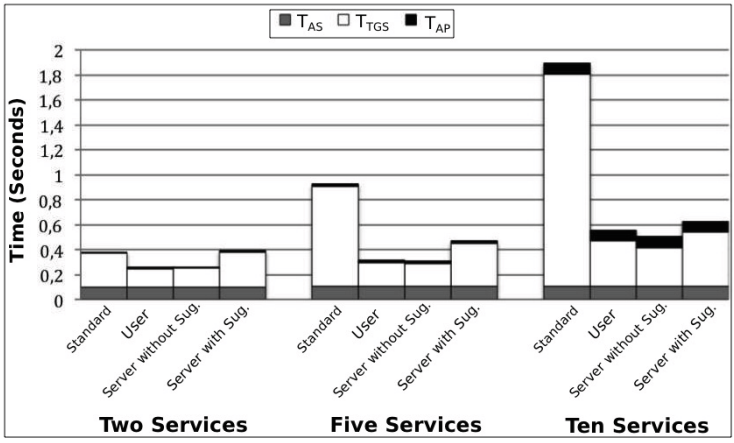



**Fig. 6.** Execution Times of Kerberos Exchanges for 2, 5 and 10 Services

Fig. 6 shows the mean time to execute all exchanges to access the services. As we can observe, $T_{AS}$ and $T_{AP}$ are similar regardless whether we use our extensions or not, since these exchanges do not need to be updated in our proposal. On the other hand, $T_{TGS}$ varies accordingly to the number of services tickets distributed and the mode of operation. With 2 services, standard Kerberos and our *Server-Controlled with Suggestions* mode obtain similar values since they both involved two KRB_TGS_REQ/KRB_TGS_REP exchanges, though a little difference can be observed. The reason is that *Server-Controlled with Suggestions* mode includes some of the new defined padatas in the first exchange in contrast with standard Kerberos (the second exchange can use the standard Kerberos since only one ticket is recovered). Nevertheless, the *User-Controlled* and *Server-Controlled without Suggestions* modes improve around a 50% due to they only involve one KRB_TGS_REQ/KRB_TGS_REP exchange. With the increment of the number of service tickets recovered, the differences between our modes and standard Kerberos augment. With 5 services, all our modes outperform the standard Kerberos reducing the time to obtain the service tickets in $\approx$3 times. As we can observe the *Server-Controlled without Suggestions* mode obtains better results in comparison with *User-Controlled* and *Server-Controlled with Suggestions*. The main reason is that *Server-Controlled without Suggestions* mode only includes PA-ST[$S_i$] in the KRB_TGS_REP but KRB_TGS_REQ does not include any additional padata.

However, the *User-Controlled* mode has to include several PA-SN, one per each service, whose number obviously increments with the number of services. In the *Server-Controlled with Suggestions* happens the same, since it uses a *User-Controlled* mode in the second KRB_TGS_REQ/KRB_TGS_REP exchange, besides involving an initial KRB_TGS_REQ/KRB_TGS_REP exchange. Finally, with 10 services, the differences generally increase up to ≈4 times with respect to standard Kerberos. The differences between the three modes are a consequence of the reasons explained for 5 services.

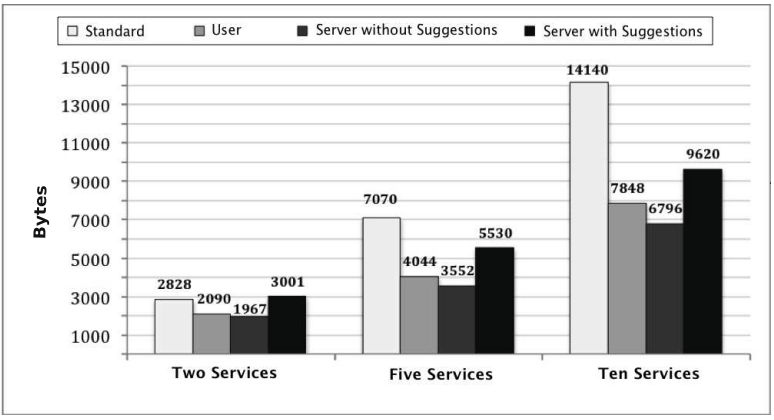



**Fig. 7.** Bytes Sent to the Network to Recover 2, 5 and 10 Service Tickets

Similar behaviour can be found in Fig. 7 with the number of bytes sent to the network in each analyzed case. As expected, the less consuming mode is the *Server-Controlled without Suggestions* for similar reasons as explained before.

## 5 Conclusions and Future Work

Standard Kerberos protocol mandates that a user accessing a service has to contact the KDC to obtain a service ticket. That is, if the user desires to access $n$ services, it has to communicate with the KDC $n$ times. In this paper, we present a solution for *Kerberos Ticket Pre-distribution*, which allows the user to obtain several tickets in a single contact with the KDC. In order to provide flexibility to our solution, we have provided three simple but effective modes of operation: *User-Controlled*, *Server-Controlled with Suggestions* and *Server-Controlled without Suggestions*. Through different experiments, we have proved that our solution outperforms the standard Kerberos protocol when requesting service tickets for several services. As an example, one area where we envisage the deployment of our contribution is related to scenarios involving the network access service and associated services, such as mobility and dynamic IP configuration. Nevertheless, we do not discard other applicability areas.

As future work, we are working on the deployment of our solution in the pre-Wimax scenario described in section 3.3, where reduction of the latency to obtain network access service is recommendable.

**Acknowledgements.** This work has been sponsored by the Ministry of Science and Innovation, through the Walkie-Talkie project (TIN2011-27543-C03), by the European Seventh Framework Program through the INTER-TRUST project (contract 317731) and the Seneca Funding Program for Research Groups of Excellence (04552/GERM/06). The authors also gratefully thank Delia Cadiz-Turpin for her valuable help in the implementation process.

## References

1. Neuman, C., Yu, T., Hartman, S., Raeburn, K.: The Kerberos Network Authentication Service (V5). IETF RFC 4120 (July 2005)
2. The MIT Kerberos Consortium, `http://www.kerberos.org` (last access date: May 20, 2013)
3. Information Technology Security: Governance, Strategy, and Practice, `http://net.educause.edu/ir/library/pdf/LIVE041.pdf` (last access date: May 20, 2013)
4. Marin Lopez, R., Pereniguez Garcia, F., Ohba, Y., Bernal Hidalgo, F., Gomez Skarmeta, A.F.: A Kerberized Architecture for Fast Re-authentication in Heterogeneous Wireless Networks. MONET 15(3), 392–412 (2010)
5. Mishra, A., Shin, M., Petroni, N., Clancy, C., Arbaugh, W.: Proactive Key Distribution Using Neighbor Graphs. IEEE Wireless Communication 11, 26–36 (2004)
6. Pack, S., Choi, Y.: Fast Inter-AP Handoff using Predictive-Authentication Scheme in a Public Wireless LAN. In: Proc. of IEEE Networks 2002 (Joint ICN 2002 and ICWLHN 2002) (August 2002)
7. Ohba, Y., Wu, Q., Zorn, G.: Extensible Authentication Protocol (EAP) Early Authentication Problem Statement. IETF RFC 5836 (April 2010)
8. Dantu, R., Clothier, G., Atri, A.: EAP methods for wireless networks. Elsevier Computer Standards & Interfaces 29, 289–301 (2007)
9. Marin-Lopez, R., Pereniguez, F., Ohba, Y., Bernal, F., Skarmeta, A.F.: A Transport-Based Architecture for Fast Re-Authentication in Wireless Networks. In: Proc. of IEEE Sarnoff Symposium 2009, Princeton, USA. IEEE Computer Society Press (2009)
10. Project Walkie-Talkie: Vehicular Communication Systems to Enable Safer, Smarter, and Greener Transportation (TIN2011-27543-C03), `http://www.grc.upv.es/walkietalkie/index.html`
11. Fernandez-Ruiz, P.J., Nieto-Guerra, C., Gómez-Skarmeta, A.F.: Deployment of a Secure Wireless Infrastructure Oriented to Vehicular Networks. In: AINA, pp. 1108–1114 (2010)
12. MIT Kerberos Distribution, `http://web.mit.edu/Kerberos/` (last access date: May 20, 2013)
13. WIRESHARK, `http://www.wireshark.org` (last access date: May 20, 2013)

# A Better Time Approximation Scheme for e-Passports

Charalampos Petrou, Christoforos Ntantogian, and Christos Xenakis

Department of Digital Systems, University of Piraeus
Piraeus, Greece
{petrou,dadoyan,xenakis}@unipi.gr

**Abstract.** E-passports are the new means of identification documents in border control points, where special reader devices named inspection terminals are installed to authenticate travelers. The authentication of e-passports to inspection terminals is based on biometric data stored in the formers, while the authentication of inspection terminals to e-passports is based on digital certificates. To check the expiration date of certificates, e-passports maintain an internal variable named *effective date*, which provides only an estimation of the current time. This introduces a serious threat on e-passports' privacy. Specifically, e-passports may accept expired certificates, considering them as non-expired, due to the time difference between the effective dates of e-passports and the current time. Thus, in case an adversary obtains an expired certificate, he/she may impersonate a fake inspection terminal and compromise sensitive personal information (e.g., biometric data) from e-passports. This paper proposes a scheme that enables e-passports to update their effective dates based on the effective dates of other, more recently updated e-passports, in a secure and effective manner. In this way, more e-passports have a better estimation of the current time, reducing the time window in which an attacker can use an expired certificate. The proposed scheme minimizes the deployment complexity, since it does not require extensive modifications to the existing infrastructure, while at the same time maintains compatibility with the legacy system.

**Keywords:** privacy, e-passports, proxy signatures, biometric, time approximation.

## 1 Introduction

E-passports are the new type of international identification travel documents that come to substitute the traditional passports, containing also biometric data (i.e., face, fingerprints, and iris). They are hybrid documents that combine the paper form with an embedded chip and antenna, allowing digital processing and wireless communication with special reader devices named inspection terminals (IS), installed at the border control points, as well as providing travelers' authentication. The extended access control (EAC) mechanism [1] describes the authentication procedure that takes place between an e-passport and an IS. However, because of some acknowledged security weaknesses of EAC, an enhanced version named EACv2 [2] has been released by the Bundesamt fur Sicherheit in der Informationstechnik - Germany. In EACv2, an IS and an e-passport, first, execute the password authenticated connection establishment

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 13–23, 2013.

(PACE) protocol, which verifies that the former has authorized access to the latter. After PACE, the terminal authentication protocol is executed, which authenticates IS to the e-passport, using a challenge-response mechanism and a three-level public key infrastructure (PKI) hierarchy. On top of this hierarchy, there is the country verifier certification authority (CVCA) with a root certificate $C_{CVCA}$, which is also stored in all e-passports of the country. Moreover, the CVCA issues certificates for domain and foreign document verifiers (DVs), $C_{DV}$; while DVs issue certificates for ISs, $C_{IS}$. During the terminal authentication protocol, the IS conveys to the e-passport a certificate chain ($C_{IS}$, $C_{DV}$, $C_{CVCA}$), and the latter using its stored $C_{CVCA}$ authenticates $C_{DV}$ and $C_{IS}$. After that, the e-passport sends to IS the stored biometric data for holder's authentication. In the final step, the chip authentication procedure is performed that protects the e-passport from cloning, as well as provides a new session key for secure data transfer.

An interesting question that arises from the above hierarchy is how the certificate of an IS is canceled. A certificate revocation list cannot be applied, since e-passports cannot be online with a public directory that maintains this list. Therefore, the limited time period validity is the only way for canceling an IS's certificate. Following this, all certificates in the employed PKI hierarchy are valid for a specific time period: (i) CVCA certificates from 6 months to 3 years; (ii) DV certificates from 2 weeks to 3 months; and (iii) IS certificates from 1 day to 1 month [3]. The lifetime of e-passports also vary from 5 to 10 years. Before the expiration of a certificate, the responsible entity requests for a new one from the upper layer of the hierarchy (i.e., an IS from a DV, and a DV from a CVCA). However, a CVCA, which resides at the top layer of the hierarchy, updates its certificate by itself using forward certificate chains [2].

Nevertheless, checking the expiration date of an IS's certificate cannot be effectively performed, since e-passports are passive RFID devices that cannot maintain an internal clock. For this reason, e-passports sustain an internal variable named *effective date*, which provides an estimation of the current time for checking certificates' expiration date. Initially, the effective date is set up equal to the time the e-passport is created, and as the e-passport passes through ISs, its effective date is updated with the most recent time value of the certificates that it receives from ISs in the certificates' chains ($C_{IS}$, $C_{DV}$, $C_{CVCA}$). However, this scheme provides only an approximation of the current time, introducing a serious threat on e-passports' privacy [4-11]. More specifically, e-passports may accept expired certificates, considering them as non-expired, due to the time difference between the effective dates of the e-passports and the current time. Thus, in case that an adversary obtains an expired certificate, he/she can exploit it to impersonate a fake IS and compromise sensitive personal information (e.g., biometric data) from e-passports.

This paper proposes a scheme that enables e-passports to update their effective dates based on the effective dates of other, more recently updated e-passports, in a secure and effective manner. In this way, more e-passports have a better estimation of the current time, reducing the time window in which an attacker can use an expired certificate to impersonate a fake terminal. To achieve this, the interacting e-passports and ISs exchange and store the most recent effective dates that they possess using the following rules: (i) if the e-passport has a newer effective date compared to this the

IS, then the latter updates its effective date with the effective date of the former, or (ii) if the e-passport has an older effective date than the IS's one, then the e-passport updates its effective date with the effective date of the IS. The security of the proposed scheme is based on proxy signatures [12]. In particular, the e-passports and ISs verify proxy signatures, created on behalf of a trusted CVCA, before updating their effective dates. The proposed scheme minimizes the deployment complexity, since it does not require extensive modifications to the existing infrastructure, while at the same time maintains compatibility with the legacy system.

The rest of the paper is organized as follows. In section 2 the related work is presented. Section 3 elaborates on the proposed scheme by analyzing its key components and functionality. Section 4 evaluates our proposal and finally, section 5 concludes the article.

## 2 Related Work

Recently, a few solutions to protect e-passports from the fake terminal attack have been proposed. In [7], the use of trusted time servers has been proposed to update e-passports' current time, using digitally signed timestamps. However, the servers' source of time is not defined, enabling the occurrence of a far-in-the-future denial of service attack. That is, the time of an e-passport is updated with an effective date far in the future. As a result, the e-passport will deny all the received certificates, because it considers them expired. In [6], the enhancement of e-passports with displays and buttons has been proposed. Based on these, the critical decision for an expired date will be taken by the e-passport's holder, who stops or allows the procedure using the button. However, semi-automated procedures may lead to users' dissatisfaction, making this solution inacceptable. Moreover, in some cases the owners give their e-passports to professionals for authentication purposes, e.g., hotel reception, bank cashier, etc., where they do not have the full control of them.

In [8], a new protocol, called on-line secure e-passport protocol (OSEP) is introduced. OSEP provides an active monitoring system, at the level of IS, that attempts to detect criminal behaviors. Additionally, OSEP includes a mutual authentication protocol between e-passports and ISs, enhancing the security of EAC. A variation of OSEP is proposed in [9] that uses elliptic curves, instead of Diffie-Hellman key agreement. An important weakness of OSEP (using either Diffie-Hellman key agreement or elliptic key cryptography) has to do with the prerequisite of online connectivity between ISs and DVs, which cannot be implemented, for example, in cases of cross-border trains and ships.

In [10], an identity based cryptography scheme is proposed, where the public keys are the users' identities. It avoids the complexity of a PKI deployment and maintenance, but it requires extensive modifications to the legacy system. Finally, in [11], a key management infrastructure is proposed, which allows dynamic update of the access keys used in EACv1. It requires less time and memory, compared to the legacy system; and the authors have implemented a prototype of this, using open-source

tools. However, many important issues have not been analyzed yet, such as the required complexity for keys' synchronization among servers. Moreover, there is no recovery process, which means that if a list of keys is compromised, all e-passports should be recalled.

A common limitation of the aforementioned solutions is that their deployment requires extensive modifications to the existing infrastructure. In particular, they propose the replacement of EAC with new protocols, which are not compatible with the existing PKI infrastructure. Moreover, they apply cryptographic functions (e.g., identity based cryptography), which have not been applied in real environments, and, therefore, their practical acceptance is limited.

## 3 Proposed Scheme

The proposed scheme enables e-passports to update their effective dates based on the effective dates of other, more recently updated e-passports, in a secure and effective manner. A key characteristic of this scheme is that its deployment does not require extensive modifications to the existing infrastructure, while at the same time maintains compatibility with the legacy system. To achieve this objective, it does not introduce any new protocol or entity, but rather extends the functionality of the existing ones in the legacy system. Thus, as in the legacy system, the proposed scheme consists of: (i) the e-passports, (ii) the inspection-update terminals (ISU) that interacts with the e-passports, (iii) the CVCAs, and (iv) the update and extended access control (UEAC) procedure. For the security of the exchanged effective dates, the proposed scheme applies proxy signatures, where e-passports sign their effective dates on behalf of trusted CVCAs. The aforementioned components are enhancements of their counterparts in the legacy system. In particular, the e-passports are enhanced to store the proxy key and the related certificate (see sect. 3.2). The ISU is an extension of the IS, maintaining also the most recent effective date received from e-passports. The CVCAs are enhanced to store and provide a backward certificate chain. Finally, the UEAC procedure, which is an extension of the legacy EACv2 procedure, is used for the mutual authentication between the e-passports and ISUs, as well as also for updating their effective dates.

### 3.1 Proxy Signatures

An e-passport updates its effective date by interacting with an ISU that stores updated effective dates of other e-passports. A question that arises is how the e-passport can verify the validity of the effective date that receives from the ISU. A possible solution would be the CVCA entity to sign the effective dates, before they are stored in the ISU. In this way, the e-passport or the ISU could verify the signature of an effective date using the public key of the CVCA. Although this solution seems to be effective and secure, it cannot be directly applied, because the CVCAs do not participate in the communication between the e-passports and ISUs.

To overcome this limitation, the proposed scheme applies proxy signatures [12] and specifically the proxy-unprotected mono-signature scheme [13], which is based on the RSA-based key pair. This scheme allows maximum compatibility with the legacy system, which also uses the RSA algorithm. Generally speaking, the main objective of proxy signatures is to delegate a *proxy signer* to sign on behalf of the *original signer*. To achieve this, the original signer using his/her private key and a random value, creates a proxy key, which is securely delivered to the proxy signer. The latter can sign messages, on behalf of the original signer, using a proxy signing algorithm and the proxy key. On the other hand, for verifying proxy signatures, only the original signer's public key is required.

In the proposed scheme, the original signer is a CVCA that generates proxy keys using its private key. On the other hand, the proxy signer is an e-passport that uses a proxy key to generate and verify the proxy signatures of the effective dates. More specifically, assume that the CVCA's certificate includes the public key $e$, while the corresponding private key is $d$. This public-private RSA key pair ($e$, $d$) satisfy $ed = 1 mod \varphi(n)$, where $\varphi(n)$ is the Euler-Totient function and $n = pq$ where $p$, $q$ are large primes randomly selected. The CVCA generates a proxy key $u$ as follows:

$$u = h(CVCA\_id, SN)^{-d} \bmod n \quad (1),$$

where CVCA_id is an identifier of the CVCA, SN is a sequence number and h() denotes a hash function. The CVCA_id, SN and the public modulus $n$ are all included in the CVCA certificate, which also contains the public key $e$. This CVCA certificate is defined as *signer CVCA certificate* and is denoted as $C_{signer}$.

### 3.2 E-passports

An e-passport stores the most recent certificate received from the interacting ISUs and, additionally, the proxy key $u$, as well as the signer CVCA certificate $C_{signer}$. The proxy key $u$ and the $C_{signer}$ do not change for the lifetime of the e-passport and are stored in a tamperproof and read-only memory area of it. For the creation of a proxy signature on an effective date (denoted as Eff.Date), the e-passport first selects an integer $t \in [1, n]$. Next, using the public key e, which is retrieved from the signer CVCA certificate $C_{signer}$, it produces the value $r$ as follows:

$$r = t^e \bmod n \quad (2).$$

Next, it generates the values $k$ and $y$ as follows:

$$k = h(Eff.Date, r) \quad (3),$$

$$y = t\, u^k \bmod n \quad (4).$$

The pair ($k$, $y$) constitutes the proxy signature.

In order to verify a proxy signature, an e-passport, first, computes $r'$ as follows:

$$r' = y^e\, h(CVCA\_id, SN)^k \bmod n \quad (5),$$

and, then, it verifies that:

$$h(Eff.Date, r') = k \quad (6).$$

This verification holds because:

$$r' = y^e \, h(CVCA\_id, SN)^k$$

$$= t^e \, u^{k\,e} \, h(CVCA\_id, SN)^k$$

$$= t^e \, h(CVCA\_id, SN)^{-k} \, h(CVCA\_id, SN)^k$$

$$= t^e = r \bmod n \quad (7).$$

### 3.3 CVCA

A CVCA generates and maintains both a forward and backward certificate chains. When the CVCA generates a new public-private key pair, it issues two different certificates: one for the forward CVCA certificate chain and another for the backward CVCA certificate chain. More specifically, assume that the CVCA has the public-private key pair ($e_i$, $d_i$) and generates a new key pair ($e_{i+1}$, $d_{i+1}$). In this case, two certificates are created. The first certificate is created for the forward CVCA certificate chain and includes the public key $e_{i+1}$ signed by the old private key $d_i$. The second certificate (i.e., backward CVCA certificate chain) includes the old public key $e_i$ signed by the new private key $d_{i+1}$.

To better understand the above notions, we use the following example: Assume that a CVCA has generated four public - private key pairs (see Fig. 1). That is, ($e_1$,$d_1$), ($e_2$,$d_2$), ($e_3$,$d_3$), ($e_4$,$d_4$), where ($e_1$,$d_1$) is the first generated pair and the ($e_4$,$d_4$) the last. In this case, the certificates $C_1$, $C_2$, $C_3$ constitute the forward CVCA certificate chain. For example, the certificate $C_2$, which includes the public key $e_3$ (with corresponding private key $d_3$), has been signed by the private key $d_2$. On the other hand, the certificates $C_4$, $C_5$, $C_6$ constitute the backward CVCA the certificate chain. For example, certificate $C_5$, which includes the public key $e_2$ (with corresponding private key $d_2$), has been signed by the private key $d_3$.

As mentioned previously, the CVCA generates the proxy keys that are used from e-passports to create the proxy signatures of their effective dates. A proxy key is generated using the private key of the CVCA certificate (see eq. 1). Note that the CVCA certificate can be either a forward or a backward CVCA certificate. In this paper, we arbitrary choose that all proxy keys are generated by forward CVCA certificates.

### 3.4 ISU

ISUs are installed at the border control points and inspect the passing e-passports using the UEAC procedure. Apart from the inspection functionality, the ISUs update also the effective dates of the e-passports. To support this additional functionality, the ISUs store for each country: (i) the most updated effective date, (ii) the corresponding proxy signature of the effective date, (iii) the related signer CVCA certificate, (iv) the

forward CVCA certificate chain, and (v) the backward CVCA certificate chain. The signer CVCA certificate stored in an ISU will be denoted as $C_{ISU\text{-}signer}$. Note that whenever a new public-private key is generated from a CVCA, the latter delivers to the ISUs both the forward and backward CVCA certificates to update accordingly their CVCA certificate chains.

### 3.5 UEAC

Similarly to EACv2, the UEAC includes the PACE, terminal authentication and chip authentication protocols. The extra functionality of UEAC is the update procedure, which is executed after the successful completion of the chip authentication. The aim of this procedure is to effectively and securely update the effective dates between ISUs and e-passports. All messages exchanged for this purpose are protected by the session keys derived from the chip authentication.

Since the PACE, terminal authentication and chip authentication protocols are performed as in the legacy EACv2, we do not analyze them. In the proposed update procedure, the involved ISU, first, delivers to the e-passport an *Update Info message* that includes the following: (i) the proxy signature, (ii) the related effective date; (iii) the forward CVCA certificate chain; (iv) the backward CVCA certificate chain; and (v) the signer CVCA certificate $C_{ISU\text{-}signer}$ that is required for the verification of the proxy signature. Upon receiving this message, the e-passport checks the validity of the received proxy signature, by verifying the received signer certificate $C_{ISU\text{-}signer}$. We identify two possible scenarios for verification of $C_{ISU\text{-}signer}$: a) the signer CVCA certificate $C_{ISU\text{-}signer}$ is older than the signer CVCA certificate $C_{signer}$ of the e-passport, and b) the signer CVCA certificate $C_{ISU\text{-}signer}$ is newer than the signer CVCA certificate $C_{signer}$ of the e-passport.

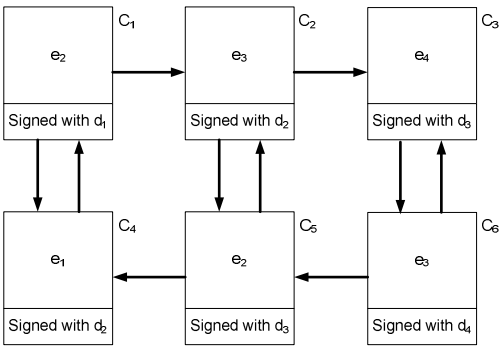



**Fig. 1.** Forward and backward CVCA certificate chains

In the first case, the e-passport should use the forward CVCA certificate chain for the verification of $C_{ISU\text{-}signer}$. That is, starting from its signer CVCA certificate $C_{signer}$, it uses the public keys of old certificates to verify the next certificates, until it reaches and

verifies the signer CVCA certificate $C_{ISU\text{-}signer}$. For example (see Fig. 1), assume that the certificate $C_3$ is the signer CVCA certificate $C_{ISU\text{-}signer}$ and the certificate $C_1$ is the signer CVCA certificate $C_{signer}$ of the e-passport. In this case, the e-passport first verifies the certificate $C_2$ using $C_1$ and, subsequently, verifies $C_3$ (i.e., the signer CVCA certificate $C_{ISU\text{-}signer}$) using $C_2$. In the second scenario, the e-passport uses the backward CVCA certificate chain to verify the signer CVCA certificate $C_{ISU\text{-}signer}$. For example (see Fig. 1), assume again that the certificate $C_2$ is the signer CVCA certificate $C_{signer}$ and the certificate $C_1$ is the signer CVCA certificate $C_{ISU\text{-}signer}$. The e-passport first verifies the certificate $C_5$ using $C_2$ and then, it verifies $C_4$ using $C_5$. Finally, the e-passport verifies $C_1$ (i.e., the signer CVCA certificate $C_{ISU\text{-}signer}$) using $C_4$.

After the successful verification of the signer certificate $C_{ISU\text{-}signer}$, the e-passport extracts from it the necessary values *e, n, SN, CVCA_id* (see sect. 3.1). If the proxy signature is valid, then the e-passport compares the received effective date with its own one. If the effective date of the ISU's certificate is more recent, then the e-passport updates its own effective date. In this case, the e-passport simply sends to ISU an *Update End* message with empty content, finalizing the procedure. On the other hand, if the effective date of the e-passport is more recent, then the e-passport signs its effective date using its stored proxy key. Next, the e-passport sends to the ISU an *Update End* message that includes the effective date, the related proxy signature and the signer certificate $C_{signer}$. Upon receiving the *Update End* message, the ISU obtains the appropriate values from the signer certificate $C_{signer}$ and proceeds with the verification of the proxy signature (see eq. (6) and (7)). If it is successful, the terminal checks that the effective date is indeed more recent from its stored one. If yes, the ISU updates its effective date and stores the proxy signature, as well as the e-passport's signer certificate (i.e., $C_{ISU\text{-}signer} = C_{signer}$).

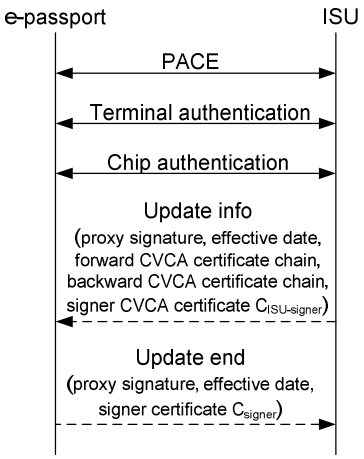



**Fig. 2.** UEAC execution

## 4 Evaluation

The proposed scheme mitigates the threat of compromised expired certificates, since an adversary can use them for a more limited time period to impersonate a fake terminal. This happens because the proposed scheme allows e-passports to update their effective date based on the effective date of other e-passports. In this way, more e-passports have a better time approximation compared to the legacy system. This can be justified as follows. Assume the owner of an e-passport with an updated effective date plans to travel. During traveling, the e-passport interacts with ISUs, which update their effective dates with the effective date of the updated e-passport. The ISUs in turn will update the effective dates of other e-passports (i.e., not updated) that interact with. In other words, the updated effective date of one e-passport propagates to other e-passports through ISUs. On the other hand, in the legacy system the e-passports update their effective dates using only the effective dates found in the certificate chains ($C_{IS}$, $C_{DV}$, $C_{CVCA}$).

One can argue that in case an ISU has not interacted with any e-passport for a long time, then it may be possible that its effective dates are not updated. However, assuming that in each country there is a critical mass of frequent travelers, the majority of ISUs in a country will have updated effective dates. The approximation of the effective dates of the e-passports with the current time depends on the time that the e-passports will interact with the ISUs. That is, if an ISU has just received a newly issued certificate and an e-passport happens to interact with the specific ISU, then the effective date of this e-passport will have a very good approximation to the current time. Note that the validity period of PKI certificates depends on the configuration of each national PKI [3].

The possibility of a fake terminal attack is also mitigated by the fact that an adversary, in order to perform this attack, should not only compromise an ISU certificate, but also possess a valid proxy signature. However, proxy signatures can be produced only by an authentic e-passport or a CVCA, as these two entities are the only authorized proxy key owners. However, it is considered that these keys in CVCAs are securely generated and stored, while in e-passports they are stored in a tamperproof read/write protected area. Moreover, a proxy key is never conveyed during the UEAC execution, eliminating the possibility an attacker to eavesdrop and obtain it. Even if an adversary obtains a valid certificate of an ISU, it cannot force an e-passport to sign a chosen effective date, since the proxy signature is produced only after the e-passport verifies that the ISU possesses also a valid signature.

One of the key advantages of the proposed scheme is that its deployment does not require extensive modifications to the existing infrastructure. The functionality of the e-passports, ISU and the UEAC protocol are extensions of the e-passports, IS and EAC, respectively, of the legacy system. The CVCAs are additionally required to store and maintain the backward CVCA certificate chain for the verification of the proxy keys. Moreover, the proposed scheme uses the same PKI hierarchy of the legacy system.

Finally, the communication overhead caused by the execution of the update procedure in UEAC is negligible, since it includes only one message exchange round

(see Fig. 2). On the other hand, the computational overhead of the proposed scheme depends on the number of certificates in the forward and backward CVCA certificate chains that an e-passport should examine to reach and verify a signer CVCA certificate. In the base case scenario, the e-passport should verify only one (1) certificate in the forward certificate chain to reach the signer CVCA certificate. On the other hand, the worst case scenario happens when the e-passport has been issued long time ago, and the validity period of the CVCA certificates is the minimum one, which is six months. In this case, assuming that the e-passport has a lifetime of 10 years and the $C_{ISU\text{-}signer}$ is the most recently issued CVCA certificate, the e-passport should verify14 different forward CVCA certificate chains to reach the signer CVCA certificate.

## 5 Conclusions

This paper proposes a scheme that enables e-passports to update their effective dates based on the effective dates of other, more recently updated e-passports, in a secure and effective manner. In this way, the e-passports have a better estimation of the current time, reducing the time window in which an attacker can use an expired certificate to impersonate a fake terminal. In the proposed scheme, an ISU and an e-passport execute the UEAC procedure to update their effective dates. To verify the authenticity of the effective dates and protect against malicious actions, the ISU and the e-passport verify proxy signatures, created on behalf of a trusted CVCA. Finally, the proposed scheme minimizes the deployment complexity, since it does not require extensive modifications to the existing infrastructure, while at the same time maintains compatibility with the legacy system.

## References

1. Bundesamt fur Sicherheit in der Informationstechnik (BSI), Germany: Advanced Security Mechanisms for Machine Readable Travel Documents – Extended Access Control (EAC), version 1.0, TR-03110 (2006)
2. Bundesamt fur Sicherheit in der Informationstechnik (BSI), Germany: Advanced Security Mechanisms for Machine Readable Travel Documents - EAC, PACE and RI, version 2.0 TR-03110 (2008)
3. Commission Decision C (2006) 2909, EU – E-passport Specification (June 28, 2006)
4. Nithyanand, R.: A Survey on the Evolution of Cryptographic Protocols in e-passports. University of California – Irvine (2009)
5. Sinhahttp, A.: A survey of system security in contactless electronic e-passports. International Journal of Critical Infrastructure Protection 4(3-4), 154–164 (2011), http://www.sciencedirect.com/science/article/pii/S187454821100045X-af000005
6. Nithyanand, R., Tsudik, G., Uzun, E.: Readers Behaving Badly Reader Revocation in PKI-Based RFID Systems. In: Gritzalis, D., Preneel, B., Theoharidou, M. (eds.) ESORICS 2010. LNCS, vol. 6345, pp. 19–36. Springer, Heidelberg (2010)

7. Ullmann, M., Vögeler, M.: Contactless Security Token Enhanced Security by Using New Hardware Features in Cryptographic-Based Security Mechanisms" from "Towards Hardware-Intrinsic Security Information" Security and Cryptography, ch. 4.4, pt. 5, pp. 259–279 (2010)
8. Pasupathinathan, V., Pieprzyk, J., Wang, H.: An on-line secure E-passport protocol. In: Chen, L., Mu, Y., Susilo, W. (eds.) ISPEC 2008. LNCS, vol. 4991, pp. 14–28. Springer, Heidelberg (2008)
9. Abid, M., Afifi, H.: Secure e-passport protocol using elliptic curve diffie-hellman key agreement protocol. In: 4th International Conference on Information Assurance and Security (2008)
10. Li, C.H., Zhang, X.F., Jin, H., Xiang, W.: E-passport EAC scheme based on Identity-Based Cryptography. Information Processing Letters 111, 26–30 (2010)
11. Najera, P., Moyano, F., Lopez, J.: Security Mechanisms and Access Control Infrastructure for e-passports and General Purpose e-Documents. Journal of Universal Computer Science 15(5), 970–991 (2009)
12. Mambo, M., Usuda, K., Okamoto, E.: Proxy signatures for delegating signing operation. In: Proceedings of the 3rd ACM Conference on Computer and Communications Security. ACM (1996)
13. Shao, Z.: Proxy signature schemes based on factoring. Information Processing Letters 85, 137–143 (2003)

# Trust Evaluation of a System for an Activity

Nagham Alhadad[1], Patricia Serrano-Alvarado[1], Yann Busnel[1], and Philippe Lamarre[2]

[1] LINA/Université de Nantes – France
[2] LIRIS/Université de Lyon – France

**Abstract.** When users need to perform a digital activity, they evaluate available systems according to their functionality, ease of use, QoS, and/or economical aspects. Recently, trust has become another key factor for such evaluation. Two main issues arise in the trust management research community. First, how to define the trust in an entity, knowing that this can be a person, a digital or a physical resource. Second, how to evaluate such value of trust in a system as a whole for a particular activity. Defining and evaluating trust in systems is an open problem because there is no consensus on the used approach. In this work we propose an approach applicable to any kind of system. The distinctive feature of our proposal is that, besides taking into account the trust in the different entities the user depends on to perform an activity, it takes into consideration the architecture of the system to determine its trust level. Our goal is to enable users to have a personal comparison between different systems for the same application needs and to choose the one satisfying their expectations. This paper introduces our approach, which is based on probability theory, and presents ongoing results.

## 1 Introduction

Everyday digital activities like chatting, mailing, blogging, buying online, and sharing data are achieved through systems composed of physical and digital resources (*e.g.,* servers, software components, networks, data, and personal computers). These resources are provided and controlled by persons (individual or legal entities) on whom we depend to execute these activities. The set of entities and the different relations between them form a complex system for a specific activity.

When users need to choose a system to perform an activity, they evaluate it considering many criteria: functionality, ease of use, QoS, economical aspects, *etc*. Trust is also a key factor of choice. However, evaluating this trustworthiness is a problematic issue due to the system complexity.

Trust has been widely studied in several aspects of daily life. In the trust management community [1,2,3,4,5,6], two main issues arise: *(i) How to define the trust in an entity, knowing that entities can be persons, digital and physical resources?* Defining the trust in each type of entity naturally is different but mainly depends on *subjective* and *objective* properties [6]. *(ii) How to evaluate such value of trust in a system under a particular context?* This point embodies the main focus of our study. We argue that studying trust in the separate entities that compose a system does not give a picture of how trustworthy a system is as a whole. Indeed, the trust in a system depends on its architecture. Several types of trust have been proposed with different meanings, which

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 24–36, 2013.

are strongly context-dependent. Defining and evaluating trust is still an open problem; there is no consensus on the approach applicable to systems in general. The aim of our work is to propose an approach applicable to any kind of system.

Inspired by probability theory, the goal of this paper is to evaluate the trust value in a system for an activity that a person wants to perform. The system definition is based on SOCIOPATH [7] which allows to model the architecture of a system by taking into account entities of the social and the digital world involved in an activity. To focus on the trust in the system, the SOCIOPATH model is abstracted in a graph-based view. Levels of trust are then defined for each node in the graph. By combining trust values, we are able to estimate two different granularities of trust, namely, *trust in a path* and *trust in a system*, both for an activity to be performed by a person. Our contribution is named SOCIOTRUST, to evaluate it, we conducted several experiments to analyze the impact of different characteristics of a system on the behavior of the obtained trust values. Experiments realized on both synthetic traces and real data sets allow us to validate the accuracy of our approach.

This paper is organized as follows. Section 2 gives a quick overview of SOCIOPATH. In Section 3, we propose SOCIOTRUST to compute the trust value in a system for an activity. Section 4 presents the experiments that validate the proposed approach. Section 5 presents some related works. Finally, we conclude in Section 6.

## 2 Overview of SOCIOPATH

The SOCIOPATH meta-model [7] describes a system in terms of the entities that exist in *(i)* the *social world*[1], where *persons* own *physical resources* and *data*, and in *(ii)* the *digital world*, where *instances of data* (including application programs) are stored and *artifacts* (software) are running. SOCIOPATH also describes the relations between the different entities of the two worlds. Enriched with deduction rules, the SOCIOPATH meta-model allows to underline and discover chains of *access* relations between *artifacts*, and *control* relations between *persons* and *digital resources* in a system. The main concepts defined in SOCIOPATH are:

- *minimal path* ($\widehat{\sigma}$); a list that begins with an *actor*, ends with a *data instance* and contains *artifacts* in between. Between each two consecutive elements in this list, there is a relation *access*. A *minimal path* describes a straight way an *actor* achieves an *activity* without passing through cycles.
- *activity* ($\omega$); a task like editing a document, where some restrictions are considered to impose the presence of particular elements in the path. For instance, if a user wants to read a `.doc` document, she must use an *artifact* that can *understand* this type of document (*e.g.,* Microsoft Word or LibreOffice Writer).

Each *artifact* in the path is controlled by at least one *person* and supported by at least one *physical resource*. In SOCIOPATH, the persons who *control* an *artifact* are the persons who *own* a *physical resource* that *supports* the *artifact* or who own some *data* represented by a *data instance* that *supports* the *artifact* (the *providers*).

[1] The words in italic in this section refer to keywords in the SOCIOPATH meta-model `http://hal.archives-ouvertes.fr/hal-00725098`

Figure 1 presents a graphical representation of a simple system drawn using SOCIOPATH. Consider that a person `John` wants to achieve the activity "accessing the document `toto` using `GoogleDocs`". In the social world, the person `John` owns some `Data`, a `PC` and an `iPad`. `Microsoft`, `Google` and `Apple` are legal entities which provide resources and artifacts. `Renater`, `Orange` and `SFR` are French telecom companies. John's `iPad` is connected to `SFR Servers` and `Renater Servers` and John's `PC` is connected to `Orange Servers`. In the digital world, the operating system `Windows` is running on John's PC. `Windows` supports `IExplorer`. John's `iPad` supports the running `iOS`, which supports the application `Safari`. John's `data` are represented in the digital world by the document `toto` that is supported by the physical resources owned by Google. We consider `Google Cloud` as the storage system used by the application `GoogleDocs`. By applying the SOCIOPATH rules on this example, we obtain the relations of *access* and *control* shown in Figure 1 where Jonh has the following minimal paths to access `toto`:

$\widehat{\sigma}_1$ ={John, Windows, IExplorer, ADSL Network, Google Cloud, GoogleDocs, toto}.

$\widehat{\sigma}_2$ ={John, iOS, Safari, SFR Network, Google Cloud, GoogleDocs, toto}.

$\widehat{\sigma}_3$ ={John, iOS, Safari, Professional Network, Google Cloud, GoogleDocs, toto}.

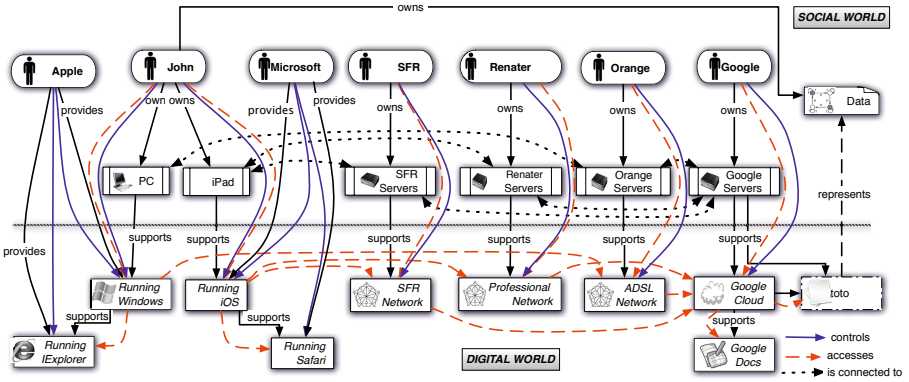



**Fig. 1.** Graphical representation of a system for the activity "John *accesses* a document `toto` on `GoogleDoc`" using SOCIOPATH

For simplicity sake, in the current paper we voluntary limit the digital activities to those that can be represented using a straight path. We do not consider activities that need multiple paths in parallel to be achieved. Most of the popular activities can be illustrated this way, such as connecting to a search engine, consulting a web page, publishing a picture, editing a document, *etc.* In the next sections, "accessing a document" is our illustrative activity.

## 3 Inferring the Trust Value of a System for an Activity

In order to evaluate the trust level of a particular user in a system for a particular activity, we first obtain a coarse-grained view of the system, from a SOCIOPATH model, as a

weighted directed acyclic graph (WDAG) (*cf.* Section 3.1). This graph represents the system allowed to perform the digital activity of the user. We then apply a probabilistic approach on this graph (*cf.* Section 3.2) to calculate the trust level of a user in a system for an activity achieved through the different paths in the graph.

### 3.1 A SOCIOPATH Model as a Weighted Directed Acyclic Graph

We simplify the representation of SOCIOPATH by using only *access* and *control* relations derived from SOCIOPATH rules. We combine an artifact, the set of persons controlling it and the set of physical resources supporting it into one unique component. These merged components are represented by the nodes in the WDAG. The edges in the WDAG represent the relations' *access*. A user performs an activity by passing through successive *access* relations of the graph, so-called a *path*[2]. A user who wants to achieve an activity associates each node with a trust value. To summarize, a system that enables a user to achieve an activity can be formally modeled as a tuple:
$\alpha_{\omega,P} =< \mathbb{N}_\omega, \mathbb{A}_\omega, t_\omega >$ where:

- $P$: the user who wants to achieve an activity.
- $\omega$: the activity the user wants to achieve.
- $\mathbb{N}_\omega$: the set of nodes in a system for an activity. Each node aggregates one artifact, the persons who control it and the physical resources that support it.
- $\mathbb{A}_\omega \in \mathbb{N}_\omega \times \mathbb{N}_\omega$: the set of edges in a system. From the rules of SOCIOPATH and the aggregation we made for a node, our WDAG exhibits only the relation *access*.
- $t_\omega : \mathbb{N} \rightarrow [0,1]$: a function that assigns to each node a trust level, which we assume to be within the interval [0,1], where 0 means not trustworthy at all and 1 means fully trustworthy. The evaluation of these values differs from one person to another. There are several ways to construct this trust level. We can figure out different objective and subjective factors that impact this trust level like the reputation of the persons who control the artifact, their skills, the performance of the physical resource that supports the artifact or the personal experience with this artifact. We thus have $t_\omega(N) = f(t_\omega^F, t_\omega^\mathcal{P}, t_\omega^{\mathcal{PR}})$, where $t_\omega^F$, $t_\omega^\mathcal{P}$, $t_\omega^{\mathcal{PR}}$ are the trust value assigned to an artifact $F$, the set of persons $\mathcal{P}$ who control $F$, the set of physical resources $\mathcal{PR}$ which support $F$ respectively for a given activity $\omega$. The meaning of the resulting trust value in a node depends on the employed function $f$ to compute this value [8]. For instance, if Bayesian inference is employed to evaluate it as is done in [9], the node trust value is considered as *the probability by which a user believes that a node can perform an expected action for a given activity* [10]. However, in this work, we do not address the issue of computing this value. Moreover, in this study, we suppose that the edges are trustworthy, and we do not assign a level of trust to the edges.

Figure 2 shows the system presented in Figure 1 as a merged WDAG where each node represents an artifact with all additional information as physical resources it depends on and persons who control it, and each edge represents the relation *accesses*. The associated value on the node represents the level of John's trust in this node. The paths

[2] If there is no ambiguity, we denote a minimal path through the WDAG by simply a path $\sigma$.

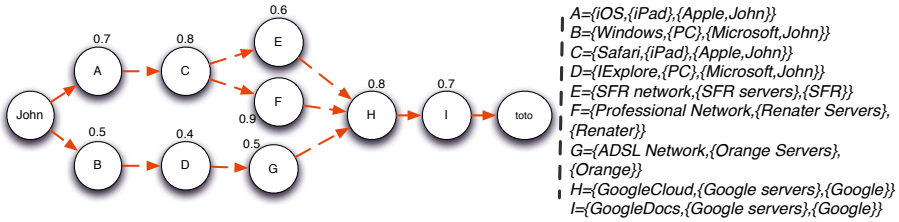



**Fig. 2.** The activity "John accesses a document `toto` on `GoogleDoc`" as a WDAG

that enable John to access toto become: $\sigma_1 =$`{A, C, E, H, I}`; $\sigma_2 =$`{A, C, F, H, I}`; $\sigma_3 =$`{B, D, G, H, I}`.

### 3.2 SocioTrust: A Probabilistic Approach to Infer the System Trust Value

Gambetta in [10] argues that: *When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him.* According to this argument, we can consider the trust value as the probability, by which one party believes that another party can perform an expected action in a certain situation [4].

We consider thus the following (Table 1 summarizes the notations used here).

- **Trust in a node**: The trust value of a user $P$ in a node $N$ for an activity $\omega$ is the probability, by which $P$ believes that $N$ provides her the expected services for $\omega$. Then, we have $t(N) = \mathbf{P}(\lambda^N)$.
- **Trust in a path**: The trust value of a user $P$ in a path $\sigma$ for an activity $\omega$ is the probability, by which $P$ believes that $\sigma$ enables her to achieve $\omega$. Then, we have $t(\sigma) = \mathbf{P}(\lambda^\sigma)$.
- **Trust in a system**: The trust value of a user $P$ in a system $\alpha$ for an activity $\omega$ is the probability, by which $P$ believes that $\alpha$ enables her to achieve $\omega$. Then, we have $t(\alpha) = \mathbf{P}(\lambda^\alpha)$.

We consider trust in a node, a path or a system as a value of probability. Hence, probability theory is the used tool to obtain the formula of these probabilities, as we show in the next section [11].

**3.2.1 Trust in a Path (Formal Evaluation):** The trust level of a person in a path for an activity is the probability that all the nodes that belong to this path provide the expected services for the activity. Let $\sigma = \{N_1, N_2, \ldots, N_n\}$ be a path that enables a person $P$ to achieve an activity $\omega$. The trust level of a person $P$ to achieve an activity through $\sigma = \{N_1, N_2, \ldots, N_n\}$ is the probability that all the nodes $\{N_i\}_{i\in[1..n]}$ provide the expected services for the activity. Thus $\mathbf{P}(\lambda^\sigma)$ is computed as follows:

$$\mathbf{P}(\lambda^\sigma) = \mathbf{P}(\lambda^{N_1} \wedge \lambda^{N_2} \wedge \ldots \wedge \lambda^{N_n})$$

**Table 1.** List of symbols and notations

| Concept | *Notation* | Concept | *Notation* | Concept | *Notation* |
|---|---|---|---|---|---|
| an activity | $\omega$ | a user who wants to achieve an activity | $P$ | the probability of an event | $\mathbf{P}(\lambda)$ |
| a node | $N$ | a path | $\sigma$ | a system | $\alpha$ |
| trust in a node value for an activity | $t(N)$ | trust in a path value for an activity | $t(\sigma)$ | trust in a system value for an activity | $t(\alpha)$ |
| the event "$N$ provides the expected services for an activity" | $\lambda^N$ | the event "$P$ achieves an activity through the path $\sigma$" | $\lambda^\sigma$ | the event "$P$ achieves an activity through the system" | $\lambda^\alpha$ |
| For a given activity $\omega$ achieved by a person $P$, the symbols $\omega$, $P$ are omitted for simplicity if there is no ambiguity | | | | | |

The event $\lambda^{N_i}$ means that $N_i$ provides the expected services for an activity. Since the graph is acyclic, then the nodes $N_1, \ldots, N_n$ are different in the path, thus each $\lambda^{N_i}$ is independent from all others. Hence, we can rewrite the trust in a path as follows:

$$\mathbf{P}(\lambda^\sigma) = \mathbf{P}(\lambda^{N_1}) \times \mathbf{P}(\lambda^{N_2}) \times \ldots \times \mathbf{P}(\lambda^{N_n}) = \prod_{i=1}^{n} \mathbf{P}(\lambda^{N_i}) \tag{1}$$

**3.2.2 Trust in a System (Formal Evaluation):** The trust level of a person $P$ in a system $\alpha$ to achieve an activity is the probability that she achieves her activity through one of the paths in the system. To evaluate the trust in a system for an activity, two cases have to be considered: (1) the paths are independent *i.e.,* they have no nodes in common and (2) the paths are dependent *i.e.,* there exists at least one node in common. The following shows how we use probability theory for these two cases.

1. **Independent paths:**
   Let $\{\sigma_i\}_{i\in[1..m]}$ be independent paths that enable a person $P$ to achieve an activity. The probability of achieving the activity through a system, $\mathbf{P}(\lambda^\alpha)$, is the probability of achieving the activity through one of the paths $\sigma_i$. Thus $\mathbf{P}(\lambda^\alpha)$ is computed as follows:
   $$\mathbf{P}(\lambda^\alpha) = \mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2} \vee \ldots \vee \lambda^{\sigma_m})$$
   Since the paths are independent then the equation can be rewritten as follows:
   $$\mathbf{P}(\lambda^\alpha) = 1 - \prod_{i=1}^{m}(1 - \mathbf{P}(\lambda^{\sigma_i}))$$
   For instance, if a person has two independent paths to achieve an activity then:
   $$\begin{aligned}\mathbf{P}(\lambda^\alpha) = \mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2}) &= 1 - (1 - \mathbf{P}(\lambda^{\sigma_1})) \times (1 - \mathbf{P}(\lambda^{\sigma_2})) \\ &= \mathbf{P}(\lambda^{\sigma_1}) + \mathbf{P}(\lambda^{\sigma_2}) - \mathbf{P}(\lambda^{\sigma_1}) \times \mathbf{P}(\lambda^{\sigma_2})\end{aligned} \tag{2}$$
2. **Dependent paths:** To evaluate the trust through dependent paths, we begin from a simple case where a system has two paths before generalizing.
   2.1. ***Two dependent paths with one common node:*** Let $\sigma_1$, $\sigma_2$, be two paths that enable a person $P$ to achieve an activity. $\sigma_1 = \{N, N_{1,2}, \ldots, N_{1,n}\}$, $\sigma_2 = \{N, N_{2,2}, \ldots, N_{2,m}\}$. These two paths have a common node, which is $N$ and so they are dependent. Thus the probability that a person $P$ achieves the activity $\omega$ through the system $\alpha$ is computed as follows:

$$\mathbf{P}(\lambda^{\alpha}) = \mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2}) = \mathbf{P}(\lambda^{\sigma_1}) + \mathbf{P}(\lambda^{\sigma_2}) - \mathbf{P}(\lambda^{\sigma_1} \wedge \lambda^{\sigma_2})$$

The probability $\mathbf{P}(\lambda^{\sigma_1} \wedge \lambda^{\sigma_2})$ can be rewritten using conditional probability as the two paths are dependent.

$$\begin{aligned}\mathbf{P}(\lambda^{\alpha}) &= \mathbf{P}(\lambda^{\sigma_1}) + \mathbf{P}(\lambda^{\sigma_2}) - \mathbf{P}(\lambda^{\sigma_2}) \times \mathbf{P}(\lambda^{\sigma_1}|\lambda^{\sigma_2}) \\ &= \mathbf{P}(\lambda^{\sigma_1}) + \mathbf{P}(\lambda^{\sigma_2}) \times (1 - \mathbf{P}(\lambda^{\sigma_1}|\lambda^{\sigma_2}))\end{aligned}$$

We have to compute $\mathbf{P}(\lambda^{\sigma_1}|\lambda^{\sigma_2})$ which is the probability that $P$ achieves the activity through $\sigma_1$ once it is already known that $P$ achieves the activity through $\sigma_2$. Thus it is the probability that $N$, $\{N_{1,i}\}_{i\in[1..n]}$ provide the expected services for this activity, once it is known that $N$, $\{N_{2,i}\}_{i\in[1..m]}$ provided the expected services. Thus $N$ has already provided the expected services. Hence, $\mathbf{P}(\lambda^{\sigma_1}|\lambda^{\sigma_2}) = \prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}})$, where $\lambda^{N_{1,i}}$ is the event "$N_{1,i}$ provides the necessary services for the activity".

$$\begin{aligned}\mathbf{P}(\lambda^{\alpha}) &= \mathbf{P}(\lambda^{N}) \times \prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}}) + \mathbf{P}(\lambda^{N}) \times \prod_{i=2}^{m} \mathbf{P}(\lambda^{N_{2,i}}) \times (1 - \prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}})) \\ &= \mathbf{P}(\lambda^{N}) \times \left[\prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}}) + \prod_{i=2}^{m} \mathbf{P}(\lambda^{N_{2,i}}) \times (1 - \prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}}))\right] \\ &= \mathbf{P}(\lambda^{N}) \times \left[\prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}}) + \prod_{i=2}^{m} \mathbf{P}(\lambda^{N_{2,i}}) - \prod_{i=2}^{m} \mathbf{P}(\lambda^{N_{2,i}}) \times \prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}})\right]\end{aligned}$$

From Equation 2 we can note that the term:

$$\prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}}) + \prod_{i=2}^{m} \mathbf{P}(\lambda^{N_{2,i}}) - \prod_{i=2}^{m} \mathbf{P}(\lambda^{N_{2,i}}) \times \prod_{i=2}^{n} \mathbf{P}(\lambda^{N_{1,i}})$$

is the probability that $P$ achieves the activity through $\sigma_1' = \{N_{1,2}, \ldots, N_{1,n}\}$ or $\sigma_2' = \{N_{2,2}, \ldots, N_{2,m}\}$ which are the paths after eliminating the common nodes. Thus the previous equation can be rewritten as follows:

$$\mathbf{P}(\lambda^{\alpha}) = \mathbf{P}(\lambda^{N}) \times \mathbf{P}(\lambda^{\sigma_1'} \vee \lambda^{\sigma_2'})$$

2.2. ***Two dependent paths with several common nodes:*** Let $\sigma_1$, $\sigma_2$, be two paths that enable a person $P$ to achieve an activity. These two paths have several common nodes. By following the same logic as in 2.1., we compute the probability that a person $P$ achieves activity $\omega$ through system $\alpha$ as follows:

$$\mathbf{P}(\lambda^{\alpha}) = \prod_{N \in \sigma_1 \cap \sigma_2} \mathbf{P}(\lambda^{N}) \times \mathbf{P}(\lambda^{\sigma_1'} \vee \lambda^{\sigma_2'}) : \sigma_1' = \sigma_1 \setminus \sigma_2, \sigma_2' = \sigma_2 \setminus \sigma_1.$$

2.3. ***Several dependent paths:*** A person may have several paths $l$ with common nodes.

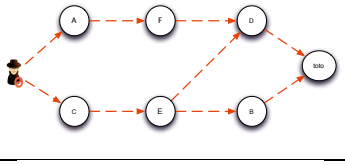
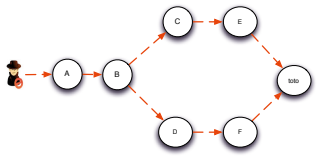
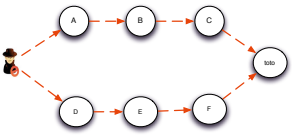
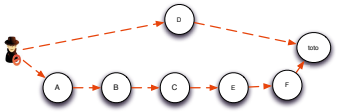
$$\mathbf{P}(\lambda^{\alpha}) = \mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2} \vee \ldots \vee \lambda^{\sigma_l}) =$$

$$\mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2} \vee \ldots \vee \lambda^{\sigma_{l-1}}) + \mathbf{P}(\lambda^{\sigma_l}) - \mathbf{P}(\lambda^{\sigma_l}) \times \mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2} \vee \ldots \vee \lambda^{\sigma_{l-1}}|\lambda^{\sigma_l}) \tag{3}$$

Let us discuss these terms one by one:

- $\mathbf{P}(\lambda^{\sigma_l})$ can be computed directly from Equation 1.
- $\mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2} \vee \ldots \vee \lambda^{\sigma_{l-1}})$ can be computed recursively using Equation 3.
- $\mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2} \vee \ldots \vee \lambda^{\sigma_{l-1}} | \lambda^{\sigma_l})$ needs first to be simplified. If we follow the same logic we discussed in Section (2.1.), the term $\mathbf{P}(\lambda^{\sigma_1} \vee \lambda^{\sigma_2} \vee \ldots \vee \lambda^{\sigma_{l-1}} | \lambda^{\sigma_l})$ can be replaced by the term $\mathbf{P}(\lambda^{\sigma'_1} \vee \lambda^{\sigma'_2} \vee \ldots \vee \lambda^{\sigma'_{l-1}})$ where we obtain each $\lambda^{\sigma'_i}$ by eliminating the nodes in common with $\sigma_l$.
- $\mathbf{P}(\lambda^{\sigma'_1} \vee \lambda^{\sigma'_2} \vee \ldots \vee \lambda^{\sigma'_{l-1}})$ can be computed recursively using Equation 3, and recursion is guaranteed to terminate when the number of paths is finite.

**Table 2.** Different systems and their trust value

| $\alpha$ | | $t_\omega(\alpha)$ | $\alpha$ | | $t_\omega(\alpha)$ |
|---|---|---|---|---|---|
| $\alpha_1$ |  | 0.4409 | $\alpha_2$ |  | 0.0144 |
| $\alpha_3$ |  | 0.507 | $\alpha_4$ |  | 0.9003 |

## 4 Experimental Evaluations

This section presents different experiments, their results, analysis and interpretation. The main objectives are *(i)* to study the influence of the system organization on the trust values, and *(ii)* to confront this approach with real users. The first two experiments are related to the first objective while the third experiment is devoted to the second.

### 4.1 Influence of the System Architecture on the Trust Value

SocioTrust is motivated by the hypothesis that studying trust in the separate nodes that construct a system does not give an accurate picture of the trustworthiness of the system as a whole. To validate this hypothesis, we apply our equations on different systems that have the same number of nodes `A,B,C,D,E,F` and the same values of trust assigned to each node, but assembled in different topologies as presented in Table 2. The values of trust associated to nodes `A,B,C,D,E,F` are $0.1, 0.2, 0.3, 0.9, 0.8, 0.7$ respectively. We calculate the trust value $t_\omega(\alpha)$ of each system. We obtain very divergent results varying from $0.0144$ to $0.9003$ as illustrated in Table 2. Collecting the values of trust in each separated node in a system is not enough to determine if the system is trustworthy or not for an activity. One must also know how the system is organized. For example, in $\alpha_2$, all the paths contain the nodes `A` and `B` and the trust values in these nodes is quite low, $0.1$ and $0.2$ respectively, so the system trust value is also low due to the strong dependency on these two nodes.

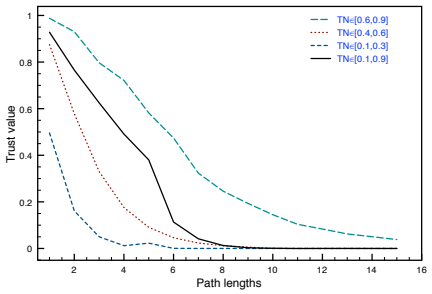



**Fig. 3.** System trust value according to the length of paths

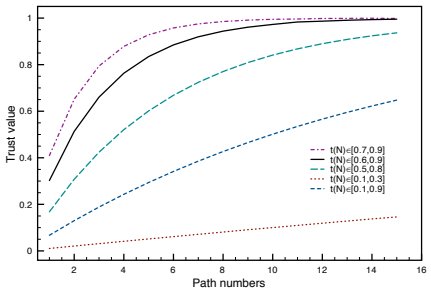



**Fig. 4.** System trust value according to the number of paths

### 4.2 Influence of the Path Length and the Number of Paths on the Trust Value

We conducted several simulations to observe the evolution of the trust value for an activity according to some characteristics in the graph. As a dataset, we considered random graphs composed of 20 to 100 nodes, and of 1 to 15 paths. Each node in the graph is associated with a random value of trust from a predefined range.

Firstly, the evolution of trust values according to the path lengths in a graph is evaluated. Each simulated graph is composed of 5 paths with lengths varying from 1 to 15 nodes. Different ranges of trust values towards the nodes were simulated, namely: $[0.1, 0.4]$, $[0.4, 0.6]$, $[0.6, 0.9]$ and $[0.1, 0.9]$. Figure 3 illustrates the impact of the path lengths on the trust value. Note that, the system trust value decreases when the length of paths increases. This reflects the natural intuition that the measure of trust in a path falls as the path gets longer, which is coherent with most of the existing results [12,13,14].

Secondly, we set the path lengths to 5 nodes and we increased the number of paths from 1 up to 15 in order to observe the variation of the trust values. Again, different node trust values were simulated: $[0.1, 0.3]$, $[0.5, 0.8]$, $[0.6, 0.9]$, $[0.7, 0.9]$ and $[0.1, 0.9]$. Simulation results are reported in Figure 4 which show that the trust value increases as the number of paths increase. This reflects the intuition that the measure of trust in a system for an activity rises when the number of ways to achieve this activity increases.

### 4.3 Social Evaluation: A Real Case

In order to evaluate SOCIOTRUST in a real use case, we modeled a subpart of the LINA research laboratory system[3] using SOCIOPATH. We applied the rules of SOCIOPATH on this system for the activity "a user accesses a document `toto` that is stored on the SVN server at LINA". Due to space constraints and privacy issues, Figure 5 presents only the WDAG of LINA for this activity, with anonymous nodes. We recall that each node represents a software that is controlled by persons and supported by physical resources. For the sake of clarity, we simplify the underlying graph as much as possible. Based on this context, we conducted an opinion survey among twenty members of LINA including, PhD students, professors and computer technicians about their level of trust in

[3] `https://www.lina.univ-nantes.fr/`

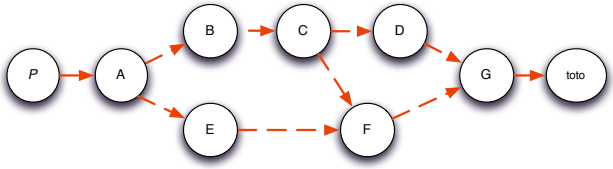



**Fig. 5.** LINA's WDAG for the activity "accessing a document `toto` on the SVN"

each node. The survey allows to infer values of function $f$ *(cf. Section 3.1)* given by real users. For each person, we have computed the system trust value according to the methodology presented in Section 3. Table 3 presents The survey data and the computed trust value in the system according to LINA members. Over a second phase, we asked each user for feedback about the system trust values computed with respect to their level of trust in the nodes. The last column of Table 3 shows this feedback, where ✓ means that they are satisfied with the value, and × means that they are not satisfied. 75% of the users are satisfied with the computation. Unsatisfied users argue that they expect a higher trust value. The node trust values of the unsatisfied users, have relatively low values (around 0.5 or 0.6) compared to the other users. These users explain that the lack of knowledge about some nodes leads them to vote with a neutral value (0.5 or 0.6) which for them considered neither trustworthy, nor untrustworthy. Clearly, such behavior is not compatible with a probabilistic interpretation where 0.5 is no more *neutral* than any other possible value. The explanations provided by users reveal an interesting point; even in the case of a local environment and even considering advanced users, not everyone is in possession of all the information necessary for an informed assessment. To conform to this reality and model this phenomenon, it requires to use a formalism allowing to express uncertainty related to incompleteness of available information. Classical probability theory is limited in expressing ignorance or uncertainty while subjective logic [15] was proposed to deal with this issue. In our future work we plan to extend SOCIOTRUST to use subjective logic.

## 5 Related Work

This paper proposes SOCIOTRUST, an approach to evaluate the system trust value for an activity as a combination of several trust values through a graph. This work is related to two domains: social networks and service-oriented computing (SOC).

Usually, a social network is represented as a graph where the nodes are persons and the edges reflect the relations between these persons. The values associated to the edges represent the value of trust between these persons. Trust propagation problem in social network focuses on finding a trust value toward a defined person or resource through the multiple paths that relate the trustor with the trustee. A lot of metrics have been proposed to calculate this trust value like the one of Richardson et al. [16], the TidalTrust [14], the SUNNY algorithm [17], the work of Agudo *et al.* [18]. SOCIOTRUST converges with these works on some points like navigating on the graph between the source and the target to collect the values of trust and combining these values to obtain the general trust value but it diverges in other points like, the values of

**Table 3.** User's trust value in the system SVN in LINA

| | A | B | C | D | E | F | G | System trust value | User's feedback about the system trust value |
|---|---|---|---|---|---|---|---|---|---|
| $P_1$ | 0.5 | 0.5 | 1 | 0.5 | 0.5 | 1 | 1 | 0.4375 | ✓ |
| $P_2$ | 0.7 | 1 | 1 | 0.7 | 0.7 | 1 | 1 | 0.847 | ✓ |
| $P_3$ | 0.5 | 0.5 | 1 | 0.7 | 0.5 | 1 | 1 | 0.4375 | × |
| $P_4$ | 0.6 | 0.6 | 0.8 | 0.7 | 0.6 | 0.8 | 0.6 | 0.3072 | × |
| $P_5$ | 0.8 | 0.8 | 1 | 0.8 | 0.8 | 1 | 0.9 | 0.8202 | ✓ |
| $P_6$ | 0.9 | 0.9 | 1 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9043 | ✓ |
| $P_7$ | 0.6 | 0.6 | 0.7 | 0.6 | 0.6 | 0.6 | 0.7 | 0.2770 | × |
| $P_8$ | 0.8 | 0.6 | 1 | 0.9 | 0.8 | 0.8 | 1 | 0.7416 | ✓ |
| $P_9$ | 0.7 | 0.5 | 1 | 0.4 | 0.7 | 0.6 | 0.9 | 0.4407 | ✓ |
| $P_{10}$ | 0.8 | 1 | 0.7 | 0.8 | 0.8 | 0.9 | 0.8 | 0.6975 | ✓ |
| $P_{11}$ | 0.5 | 0.5 | 0.9 | 0.5 | 0.5 | 0.5 | 0.9 | 0.2473 | × |
| $P_{12}$ | 0.95 | 0.95 | 0.8 | 0.8 | 0.95 | 0.95 | 0.8 | 0.8655 | ✓ |
| $P_{13}$ | 0.8 | 0.9 | 0.8 | 0.7 | 0.95 | 0.8 | 0.7 | 0.6433 | ✓ |
| $P_{14}$ | 0.8 | 0.7 | 0.9 | 0.7 | 0.9 | 0.8 | 0.8 | 0.6652 | ✓ |
| $P_{15}$ | 0.9 | 0.8 | 0.8 | 0.9 | 0.9 | 0.9 | 0.8 | 0.7733 | ✓ |
| $P_{16}$ | 0.7 | 0.6 | 0.6 | 0.6 | 0.8 | 0.7 | 0.6 | 0.337 | ✓ |
| $P_{17}$ | 0.5 | 0.9 | 0.8 | 0.7 | 0.9 | 0.5 | 0.8 | 0.3807 | × |
| $P_{18}$ | 0.7 | 0.7 | 1 | 0.7 | 0.6 | 0.7 | 1 | 0.6088 | ✓ |
| $P_{19}$ | 0.8 | 0.8 | 1 | 1 | 1 | 0.8 | 1 | 0.8704 | ✓ |
| $P_{20}$ | 0.9 | 0.9 | 0.8 | 0.9 | 0.9 | 0.9 | 0.8 | 0.7971 | ✓ |

trust associated to each node in our work are values attributed by the source node which represent her trust in these nodes. In their works, the values associated to the edges represent the trust between the nodes related with this edge. Hence, these works discuss the problem of trust propagation through a graph, while SOCIOTRUST focuses on finding a trust value toward the whole graph that reflects an activity performed through it.

In SOC, a service invokes other services forming a composite service, so the composite service can be represented as a graph where the nodes represent the service components and the edges represent the relation of invocation. In [13,19,9], authors evaluate the trust toward the composite service by considering the value of trust as probability depending on the definition presented in [4]. They calculate a global trust value toward the separated services and they use the theory of probability to evaluate the global trust value of the composite services. These works are similar to our proposal in some points. Firstly, the value associated to a node in the graph is represents the value of trust toward a service. Secondly, they consider this value as a probability that the node performs a given action that enables a user to achieve her activity. However, they diverge from SOCIOTRUST in a main point. In their work, the computed trust value is toward a certain choice (path) of the composite services where in our work, it is toward the whole system including all the paths that enable a user to achieve an activity.

The trust evaluation proposed in these two domains cannot be straightly adopted in our work due to the difference in the graph nature between their works and ours.

## 6 Conclusion and Perspectives

In this paper, we present a new notion of trust: trust in a system for an activity. We propose SOCIOTRUST, a probabilistic approach to calculate the system trust value. We conduct some experiments to illustrate that the system construction is a key factor in

evaluating the user trust value in a system. Finally, we confront our approach with real user opinions based on a real modeled system to extract the limitations of this proposition. A serious limitation of our study is that trust values have been considered as a traditional probability where expressing ignorance or uncertainty is not possible. Subjective logic which is an extension of probability theory can deal with this issue. We are currently extending SOCIOTRUST to use subjective logic.

## References


1. Marsh, S.: Formalising Trust as a Computational Concept. PhD thesis, University of Stirling (1994)
2. Moyano, F., Fernandez-Gago, C., Lopez, J.: A Conceptual Framework for Trust Models. In: Fischer-Hübner, S., Katsikas, S., Quirchmayr, G. (eds.) TrustBus 2012. LNCS, vol. 7449, pp. 93–104. Springer, Heidelberg (2012)
3. Viljanen, L.: Towards an Ontology of Trust. In: Katsikas, S.K., López, J., Pernul, G. (eds.) TrustBus 2005. LNCS, vol. 3592, pp. 175–184. Springer, Heidelberg (2005)
4. Jøsang, A., Ismail, R., Boyd, C.: A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems 43(2), 618–644 (2007)
5. Zhang, P., Durresi, A., Barolli, L.: Survey of Trust Management on Various Networks. In: International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2011), pp. 219–226 (2011)
6. Yan, Z., Holtmanns, S.: Trust Modeling and Management: from Social Trust to Digital Trust. In: Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions (2007)
7. Alhadad, N., Lamarre, P., Busnel, Y., Serrano-Alvarado, P., Biazzini, M., Sibertin-Blanc, C.: SOCIOPATH: Bridging the Gap between Digital and Social Worlds. In: Liddle, S.W., Schewe, K.-D., Tjoa, A.M., Zhou, X. (eds.) DEXA 2012, Part II. LNCS, vol. 7447, pp. 497–505. Springer, Heidelberg (2012)
8. Mcknight, D.H., Chervany, N.L.: The Meanings of Trust. Technical report, University of Minnesota, Carlson School of Management (1996)
9. Li, L., Wang, Y.: Subjective Trust Inference in Composite Services. In: 24th Conference on Artificial Intelligence (AAAI), pp. 1377–1384 (2010)
10. Gambetta, D.: Can We Trust Trust?. In: Gambetta, D. (ed.) Trust: Making and Breaking Cooperative Relations. Department of Sociology, pp. 213–237. University of Oxford (2000)
11. Carminati, B., Ferrari, E., Morasca, S., Taibi, D.: A Probability-Based Approach to Modeling the Risk of Unauthorized Propagation of Information in on-Line Social Networks. In: ACM Conference on Data and Application Security and Privacy (CODASPY), pp. 51–62 (2011)
12. Hang, C.W., Wang, Y., Singh, M.P.: Operators for Propagating Trust and their Evaluation in Social Networks. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pp. 1025–1032 (2009)
13. Li, L., Wang, Y.: A Subjective Probability Based Deductive Approach to Global Trust Evaluation in Composite Services. In: IEEE International Conference on Web Services (ICWS), pp. 604–611 (2011)
14. Golbeck, J.A.: Computing and Applying Trust in Web-Based Social Networks. PhD thesis, University of Maryland, College Park, College Park, MD, USA (2005) AAI3178583
15. Jøsang, A.: A Logic for Uncertain Probabilities. International Journal of Uncertainty Fuzziness and Knowledge-Based Systems 9(3), 279–311 (2001)

16. Richardson, M., Agrawal, R., Domingos, P.: Trust Management for the Semantic Web. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 351–368. Springer, Heidelberg (2003)
17. Kuter, U., Golbeck, J.: SUNNY: A New Algorithm for Trust Inference in Social Networks Using Probabilistic Confidence Models. In: Proceedings of the National Conference on Artificial Intelligence (AAAI), pp. 1377–1382 (2007)
18. Agudo, I., Fernandez-Gago, C., Lopez, J.: A Model for Trust Metrics Analysis. In: Furnell, S.M., Katsikas, S.K., Lioy, A. (eds.) TrustBus 2008. LNCS, vol. 5185, pp. 28–37. Springer, Heidelberg (2008)
19. Li, L., Wang, Y.: Trust Evaluation in Composite Services Selection and Discovery. In: IEEE International Conference on Services Computing, pp. 482–485 (2009)

# Defining a Trust Framework Design Process

Mark Vinkovits and Andreas Zimmermann

Fraunhofer FIT
User Centered Ubiquitous Computing
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
{mark.vinkovits,andreas.zimmermann}@fit.fraunhofer.de

**Abstract.** Trust Management researchers have created several sound frameworks for well-defined problems based on their deep understanding of the field and the current state of the art. However, Trust Management experts are rarely contacted for the design of distributed business applications. Developers of these kinds of business applications are not familiar with latest results and are only aware of a very limited set of applicable solutions. This hinders the adaption of these novel security solutions into provided future services. To support the integration of Trust Management into these areas we defined a design process, which can be applied systematically by developers. Based on the design process they can use their scenario knowledge to narrow down their design space and finally select from a limited set of applicable implementations the best fit. We extended the static TrustFraMM meta-description of Trust Management in a way to enable it to support the exploration and exclusion of existing trust functionality implementations. We built our process on a number of requirements collected through our user-centered design approach. We also provide a possible visualization of the process which we evaluated using a paper prototype. Our process had a positive user acceptance among the questioned users.

**Keywords:** Trust management, development process, user-centered design, meta-model.

## 1 Introduction

With the recent acknowledgment of soft-security and Trust Management [1] different domains seek robust procedures protecting nodes in open and distributed environments. Trust Frameworks (TF) evolved with specific tasks for well-defined domains. They estimate the probability of correctness or the related risk of consuming a service of one node or another. Such a service might be the distribution of a file in a P2P network [2], forwarding of a data packet in an ad-hoc network [3] or simply a statement about the authenticity of a person [4].

As shown by multiple surveys, security expertise and especially Trust Management (TM) expertise is not available at the implementation of new business services and

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 37–47, 2013.

applications[1]. Following the current trend of growth in the Information and Communication Technologies sector and the number of businesses in this area the situation of security expertise involvement is not likely to change. As a result these business solutions do not benefit from the novel TM solutions specifically designed for their use-cases. To overcome this obstacle our suggestion is to create systematic procedures for integrating TM that can be directly applied by non-security experts to their applications.

We created a TF design process applicable by system designers and developers building on their domain knowledge and its relevant requirements. Our process is based on the Trust Framework Meta-Model [5], which modularizes available TFs into a well-defined structure. It provided a good framework to access state of the art knowledge and divide the problem space into smaller units. We added attributes to the different static elements of the meta-model that focus on typical aspects of the element's functionality. These attributes are comprehensible to the users and enable them to describe their use-case. With the process users can explore and narrow down the design space by selecting and excluding some of the attributes. This way they can select the most compelling implementation of the TF for their application at the end of the process.

As a next step we created visualization for the process based on a user-centered design approach [6, 7]. We held multiple workshops and interviews to understand users' needs and their understanding of TM. We incorporated their requirements into a paper prototype and evaluated it using interviews. Our interviewed users appreciated the operation of the process and found the visualization appropriate to support them. They also found that by selecting attributes and seeing how this influences their remaining choices they learn possible combinations and gain additional comprehension of TM.

In this paper we first look at the state of the art. Section 3 describes the user-centered design approach we followed and the user needs we collected throughout our iterations. Based on these needs we created a design process which is presented in section 4. The prototyped visualization is presented in section 5, followed by the conclusions in section 6.

## 2 State of the Art

There is a large number of TFs available designed for different domains [2, 3, 4]. Similarly there exist many surveys collecting and categorizing TFs and identifying common aspects [8, 9]. As evident from these examples TM has exhaustively been researched and there are many findings available. It is easier to access these findings with the Trust Framework Meta-Model [5]. TrustFraMM was born from the idea to identify identical functionalities in different available TFs. These functionalities have been formalized into elements with dependencies and interfaces. Using TrustFraMM it is possible to describe any TF as a set of standard functionalities. The authors provided typical implementations for their elements based on existing solutions. However, as TrustFraMM

[1] http://www.cio.com/article/710218/Ease_the_Need_for_IT_Security_Pros_by_Writing_More_Secure_Code [accessed: 16th March 2013].

is only a static model describing frameworks it is per se not applicable for the design of a TF. We amended TrustFraMM and used it as a basis for a design process as described in the later sections.

When trying to find any defined development process for TM there is little work available to build on. Ref. [10] introduces an UML-based method, which by comparing alternative outcomes defines trust policies which ensure optimal outcome for the specific interactions. While the method is systematic its scope is more limited then we aim for. A more general approach can be seen in [11], which provides a number of questions that help to find the appropriate trust metrics for a use-case based on the assessed risks. While this process is very helpful for developers with experience in the field the questions are too unspecific to be used by a non-security expert.

Applying user-centered design [6] to security is a very novel approach. User-centered security is a specialization of UCD and is defined as "security models, mechanisms, systems and software that have usability as a primary motivation or goal" [7]. Ref. [12] uses a user-centered design approach to collect user needs of TM. We used their identified needs for the first iteration of our design process and collected additional ones.

## 3 User-Centered Design Steps and Identified Needs

Our main requirement for the design process was to support non-security expert system designers and developers in developing secure distributed applications based on TM functionality. To achieve this we decided to follow a user-centered approach [6, 7] involving such users. This ensured that we have a proper understanding of the target users of our design process and their requirements.

### 3.1 Involving Users

As seen from the workshops of [12] "… designers and developers are confident in the use of the term trust…" and "they do not fear to make decisions based on it in existing systems." So as long as there are TFs available to calculate trust values and non-security experts do not need to implement the procedures themselves they believe in their effectiveness. Building on these results but also to have additional and more specific requirements for our purpose we held a workshop with six users. Our users worked all as researchers in an IT institute but had different specializations. Three worked mainly in the HCI field, one in business informatics, one in context-aware systems and one in web development. Each of our users had significant experience with the development of distributed systems and had also industry experience but were no security experts. We presented two different scenarios to them in which they had to design different applications. They had an open discussion about how they would proceed with the development of the application, assuming they had tool

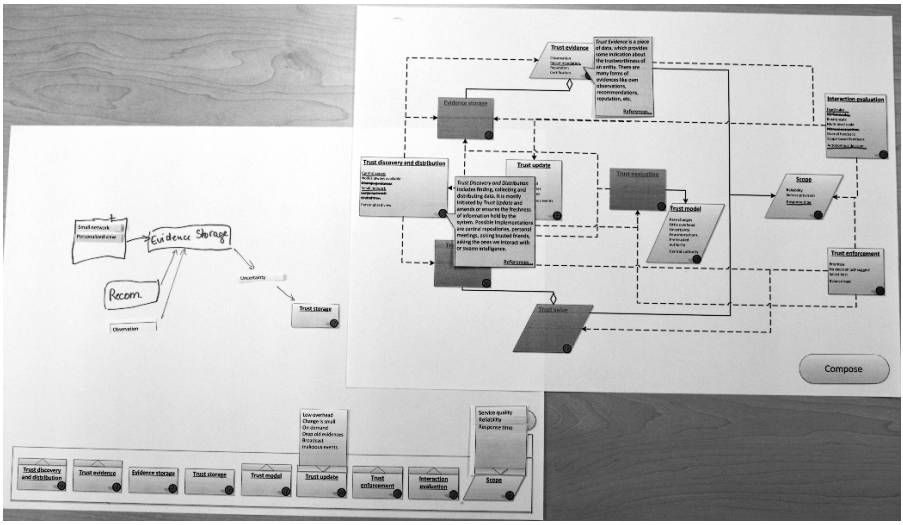



**Fig. 1.** The two paper prototypes

support for the TM part. The two scenarios were designed in such a way that they would focus on different aspects and enable us to collect several different needs.

The first scenario foresaw a simpler TF making them feel more comfortable and enabling them to focus on the development process. It was about developing a phone app in which users in real-time reported and acknowledged the existence of traffic jams. The trustworthiness of these reports was to be assessed by the system.

The second scenario was more challenging as it required the participants to design a trustworthy routing system for ad-hoc networks. As this scenario was difficult to handle for the users it made them envision how they would like to receive support for the design of the TF. Aim of the workshop was also to see whether our imagined process would fit users' expectation.

Based on the input from the workshops we finalized the design process and decided to use paper prototypes as evaluation tool [13]. The created paper prototypes can be seen in Fig. 1. We had prepared two conceptually different prototypes to simultaneously test a larger set of requirements and be able to have a comparison during our test. The prototypes consisted of large sheets of paper with the default workspace printed on them. We prepared additional whole paper sheets for larger changes in the interface. We placed smaller paper labels onto the workspace as pop-up when needed. The first prototype (left in Fig.1) was more focused on the information flow and gave users a free hand in composing their architecture. The workspace had at the bottom the TrustFraMM elements, which when clicked on listed the assigned attributes. Attributes could be dragged and dropped onto the workspace and then grouped or connected using a pen. The second prototype (right in Fig. 1) showed the TrustFraMM elements connected according to functional dependencies. The attribute selection was simulated using a pen. After each interaction of the user we

either placed pre-prepared paper labels onto the workspace or drew the exclusions with the pen.

These prototypes were separately shown to the same six users who participated in the workshop. We decided to use the same users because we wanted to iterate their formulated needs. The users were asked to use the prototype in the traffic jam scenario from the workshop and comment on the usability. After collecting comments about specific usability issues we always had a short open discussion with them. In this we asked what their general opinion was, what they thought as strengths and weaknesses and whether they would use such process. These comments were used to finalize the process and our proposed visualization.

### 3.2 Identified Needs

After the workshops and interviews we evaluated the notes we made during the discussions and identified main needs concerning the design process. These needs can be divided into two main categories. One category contains more general needs concerning the main properties a design process should have, the way it should work and how it should be used. These needs are collected in Table 1. The second category, shown in Table 2, is more related to our created design process and mainly contains the feedback gained from the paper prototypes. Due to space limitations we did not include those needs which were already listed in [12].

**Table 1.** General needs

| | |
|---|---|
| Visualization | • Visualization should serve as catalogue of options<br>• Visualization should be similar to UML<br>• Information flow should be traceable |
| Suggestions | • Support should suggest Trust Model<br>• Fine tuning should be possible based on defaults<br>• Higher level categories should be provided<br>• Colliding requirements should be traceable |
| Implementation | • Code generation is inconvenient<br>• API should provide base classes to inherit from<br>• Implementation should be event based |

**Table 2.** Specific needs

| | |
|---|---|
| Attributes | • Attribute wording should be self-explanatory<br>• Additional descriptions about the meaning of attributes should be provided<br>• Similar attributes should be grouped |
| Exclusion | • See which attribute caused what exclusions<br>• Description should be provided why two attributes collide |
| Selection | • References for further reading should be provided for implementations<br>• Implementations should be ordered based on relevance |

## 4 Defined Trust Framework Design Process

Based on the requirements presented in section 3 we aimed at creating an easy to follow TF design process, which can provide a sound solution building on the scenario knowledge of the user. We found that the creation of a TF for a new scenario is not the result of a systematic design process with sequential steps. If using approved state of the art procedures, the design rather consists of selecting suitable solutions for the scenario at hand. There are typical aspects of a problem, like the availability of a trusted entity or the need to balance load, which narrow the number of applicable solutions. Thus we conclude that selecting the appropriate TF is an explorative process. Therefore what we wished to provide the users with was a clear method to explore possible solutions guided through their scenario. For this purpose we suggest to describe the approved TF procedures using attributes, which target the problem aspects we just mentioned.

In our foreseen design process the user selects, based on the characteristics of the scenario, attributes that apply to her use-case. These attributes are assigned to the different elements of TrustFraMM to distribute the problem into smaller sets. When an attribute is selected other attributes that collide with the selected one get unavailable, thus step by step narrowing the design space. As an example, when we decided to use exclusively global reputation values (like eBay [13]) for participants, it is not consistent to use recommendations anymore as these would produce a personalized view of the participants. After the user has selected all the attributes she sees appropriate the process provides the set of remaining implementations possible. The following sections provide a more detailed description of the individual steps and related design decisions of our process.

### 4.1 Attributing the Elements

To group the attributes into smaller units as the users requested we decided to use TrustFraMM as a basis. As shown in [5] the different elements of TrustFraMM have different typical implementations. Depending on the application scenario some implementations are more appropriate than others, e.g. it is inconvenient to use a central server if there is no continuous Internet connectivity. Consequently, we propose to assign the planned attributes to each implementation to characterize the functionality and applicability of it. We also suggest the use of standardized attributes to ensure that different implementations select from the same set to describe themselves.

This approach also enables the process to be extendable and not only work on the number of implementations available when realizing it. Every time a new implementation is designed for one of the elements the author can describe it using the available set of attributes or extend the standard set, if the available ones are not sufficient for the exact description. Without any change to the process the users can be presented with and also use the new attributes together with the already available ones.

### 4.2 Exclusion of Colliding Attributes

When the process starts inside each TrustFraMM element the user is presented with the collection of all the attributes the different respective implementations have. As attributes are standardized and they describe different sub-aspects many listed attributes are going to belong simultaneously to multiple implementations. This limits the total number of individual attributes that are listed.

The user now starts selecting attributes, which seem appropriate for the specific use-case. There is no limitation to which element should be handled initially as depending from the use-case different aspects may be known in advance. For example, if in the beginning the network structure is known it is advisable to start with *Trust discovery and distribution*. On the other hand, if rather the interaction with the system is given *Interaction Evaluation* and *Trust enforcement* may seem as good starting points.

As the user selects attributes the process checks which available implementations contain it. The attributes of the remaining implementations stay selectable, while the other attributes get struck through. This process continues as long as the user selects attributes, leaving at the end only a limited number of attributes and thus limited number of possible implementations remaining.

Additional issue to solve is the relation between different TrustFraMM elements and the exclusion of attributes based on these dependencies. Similarly to attributing an implementation in section 4.1 the creator of an implementation has to look at TrustFraMM elements her element depends on. From these elements the attributes have to be chosen which when selected the implementation is not applicable anymore. Consequently in the process, when an attribute is selected all depending TrustFraMM elements have to be checked. The attributes of not colliding implementations remain selectable, while the remaining attributes are struck through.

As a dependency gets resolved, resulting in excluding attributes, this may have an effect on further dependencies. This recursive operation will stop after a limited number of steps as TrustFraMM does not contain any dependency cycles. As soon as the user is finished with selecting attributes she can proceed to the next step in the process.

### 4.3 Selecting the Specific Implementation

When the user decides that she is finished with selecting attributes the final step of the process may start. Since during every step the remaining possible implementations were checked it is possible without any further computation to list them. It is advisable however, to present one implementation for each element as default based on the largest set of attributes covered. Then, in sorted order, further implementations based on the number of attributes covered should be listed.

During this final step the user can view the implementations, which have been suggested for the individual elements. After reviewing the implementations the user may decide to change the implementation from the default one to another or go back to selecting another set of attributes. If the user has selected the implementations she would like to use for the elements the process can be closed. The user has now gained from requirements and attributes about her application a suggestion on a TF and the implementation of different elements without the need to investigate TM state of the art.

## 5 Interaction and Graphical User Interface

Section 4 introduced the theoretic basics of our proposed design process. As one of the main requirements of the users was to have a visual aid for the process we also built a possible visualization. In this section we introduce the structure of our visualization and emphasize on the features we included for better usability. We also present here fragments of the finalized paper prototype created throughout our iterations.

At the beginning the user is presented with an overview of TrustFraMM with the relevant attributes within the elements. This helps the users to know the different aspects of a TF and avoids them feeling lost in this unknown field. When a user goes over an attribute with the mouse a hint is displayed to avoid problems regarding wording. An illustration for this can be seen in Fig. 2. There is a button in each element through which a more detailed description can be reached about the responsibility of the element and the cause for dependencies to other elements.

The selection of an attribute is marked through a dot appearing at the end of the line. As attributes are selected others are struck through. A cross icon behind excluded attributes enables the users to see the list of attributes - both from this and from other elements - colliding with this one. A button in the corner of the workspace allows the user to proceed to the next step of the process.

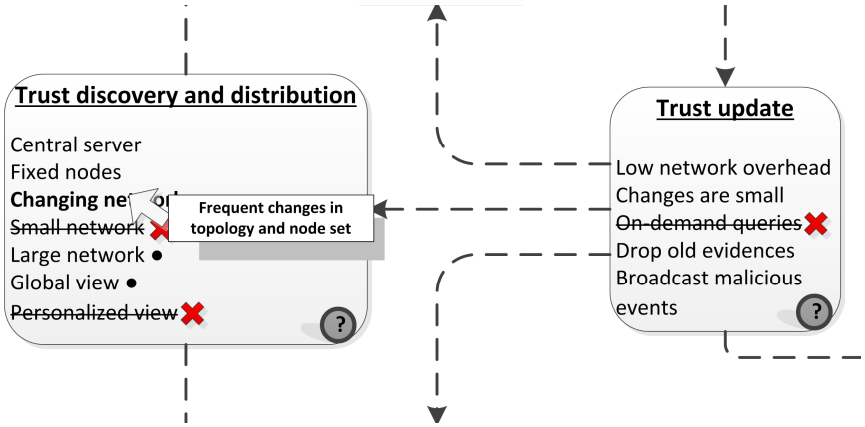



**Fig. 2.** Selecting attributes of TrustFraMM element

In the next step the user again sees an overview of TrustFraMM. This time, however, each element is filled with a specific implementation based on the selection of attributes from the previous step. An illustration of this can be seen if Fig. 3. Below the implementation name is a short text as a reminder of the functionality of the specific implementation. The box can be expanded by the use of the triangle button in

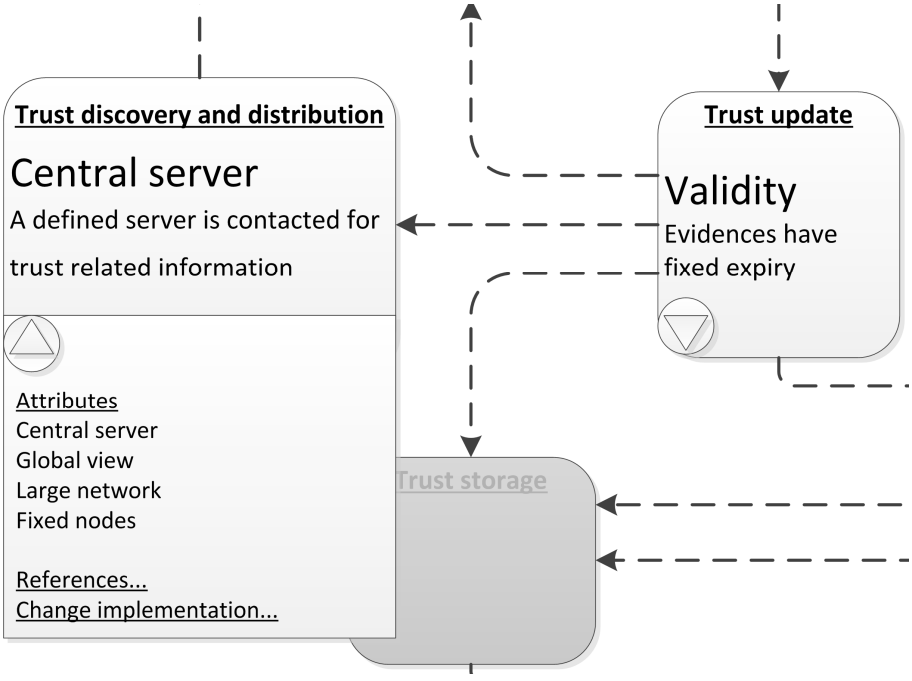



**Fig. 3.** Showing implementations for elements

the left bottom corner. When expanded the user is presented with the set of attributes the implementation fulfills. There are two additional options. With the "References…" button additional information about the implementation can be requested in form of an extended description or a number of papers describing it. To ensure that the replacement of the implementation is a conscious decision the list of further implementations can only be reached over an additional dialog. There the user can see what attributes other implementations fulfill and can collect additional details, which support the decision.

The main conclusion we gained from the GUI prototype was that a visualization of the design process also serves an e-learning purpose. This means that the user learns about TM and the possibilities while going through the design. We decided to realize this via a great number of tooltips, help dialogs and further visual clues. This additional information also serves a better understanding of TM. This is necessary as users are not going to integrate components they do not completely understand or they do not know how those work.

## 6 Conclusions

In this paper we presented a systematic approach for the design of Trust Frameworks. Throughout the creation of the design process our main requirement was that it should be applicable by security-experts but also by non-security experts. To achieve this we used a user-centered design approach and held multiple interviews with distributed system designers and evaluated our product with a paper prototype.

Our process is built on attributing the elements of the Trust Framework Meta-Model. The creators of specific implementations describe the implementation's properties with a default set of attributes and define colliding attributes for the dependent TrustFraMM elements. The user of the process can then select from these attributes and this way they exclude not fitting implementations. At the end of the process the user receives a suggestion for her use-case without having investigated into Trust Management.

We also built a possible visualization for the process based on the user needs we collected. With our visualization the user is guided through the process and based on the tooltips and further visual clues provided her knowledge of Trust Management is extended. This enables the user at the end of the process to make a more conscious decision regarding the implementation to be applied.

The result of our interviews was that the proposed design process with the suggested visualization was comprehensible and usable for the target end-users. They understood the concept, appreciated to learn more about Trust Management and felt capable of designing an appropriate Trust Framework.

As next step of our research we plan on implementing a tool, which uses the presented process and is also able to generate executable program code. This can then be integrated by the user into her application. This tool will build on the well-defined interfaces of TrustFraMM and contain several implementations from state of the art Trust Frameworks. We will also provide a number of pre-defined classes to be

implemented by the user to customize the framework to their target use-case. Using this tool we will then be able to further evaluate and improve our process.

**Acknowledgements.** This work has been performed within the ebbits project, co-funded by the EC within the FP7, theme ICT-2009.1.3 Internet of Things and Enterprise environments, grant agreement No. 257852. Possible inaccuracies of information are under the responsibility of the project. This report reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein.

## References

1. Jøsang, A., Keser, C., Dimitrakos, T.: Can We Manage Trust? In: Herrmann, P., Issarny, V., Shiu, S.C.K. (eds.) iTrust 2005. LNCS, vol. 3477, pp. 93–107. Springer, Heidelberg (2005)
2. Marti, S., Garcia-Molina, H.: Limited reputation sharing in P2P systems. In: Proceedings of the 5th ACM Conference on Electronic Commerce, EC 2004, pp. 91–101. ACM Press, New York (2004)
3. Zouridaki, C., Mark, B.L., Hejmo, M.: Byzantine robust trust establishment for mobile ad hoc networks. Telecommunication Systems 35, 189–206 (2007)
4. Blaze, M., Feigenbaum, J., Lacy, J.: Decentralized trust management. In: Proceedings of the 1996 IEEE Symposium on Security and Privacy, pp. 164–173. IEEE Comput. Soc. Press (1996)
5. Vinkovits, M., Zimmermann, A.: TrustFraMM: Meta Description for Trust Frameworks. In: ASE/IEEE International Conference on Privacy, Security, Risk and Trust, Amsterdam, Netherlands, pp. 772–778 (2012)
6. Gould, J., Lewis, C.: Designing for usability: key principles and what designers think. Communications of the ACM 28, 300–311 (1985)
7. Zurko, M.E., Simon, R.T.: User-centered security. In: Proceedings of the 1996 Workshop on New Security Paradigms, NSPW 1996, pp. 27–33. ACM Press, New York (1996)
8. Josang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision Support Systems 43, 618–644 (2007)
9. Artz, D., Gil, Y.: A survey of trust in computer science and the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web 5, 58–71 (2007)
10. Refsdal, A., Solhaug, B., Stølen, K.: A UML-based Method for the Development of Policies to Support Trust Management. In: Karabulut, Y., Mitchell, J., Herrmann, P., Jensen, C.D. (eds.) Trust Management II. IFIP, vol. 263, pp. 33–49. Springer, Boston (2008)
11. Povey, D.: Developing electronic trust policies using a risk management model. In: Baumgart, R. (ed.) CQRE 1999. LNCS, vol. 1740, pp. 1–16. Springer, Heidelberg (1999)
12. Vinkovits, M.: Towards requirements for trust management. In: Privacy, Security and Trust (PST) 2012, pp. 159–160. IEEE Comput. Soc., Paris (2012)
13. eBay - The World's Online Marketplace (2012), `http://www.ebay.com`
14. Greenberg, S., Carpendale, S., Marquardt, N., Buxton, B.: Sketching User Experiences: The Workbook. Morgan Kaufmann Publishers Inc., San Francisco (2011)

# Executable Model-Based Risk Analysis Method for Identity Management Systems: Using Hierarchical Colored Petri Nets

Ebenezer Paintsil and Lothar Fritsch

Norwegian Computing Center Oslo, Norway
{paintsil,lothar.fritsch}@nr.no

**Abstract.** Model-based risk analysis methods use graphical models to facilitate participation, risk communication and documentation and thereby improve the risk analysis process. Currently, risk analysis methods for identity management systems (IDMSs) mainly rely on time consuming and expensive manual inspections and lack graphical models. This article introduces the executable model-based risk analysis method (EM-BRAM) with the aim of addressing these challenges. The EM-BRAM employs graphical models to enhance risk analysis in IDMSs. It identifies risk contributing factors for IDMSs and uses them as inputs to a colored petri nets (CPNs) model of a targeted IDMS. It then verifies the system's risk using CPNs' state space analysis and queries.

## 1 Introduction

Identity management systems (IDMSs) create and manage identities of end-users [1]. They have three main stakeholders - the system end-users, who create or obtain and show credentials; the identity provider (IdP), the organization that issues the credentials to end-users; and the service provider (SP); the organization that provides services or resources to end-users after verifying their identities. SPs may be referred to as relying parties (RPs).

Model-based risk analysis methods use graphical models to facilitate participation, risk communication and documentation [2] and thereby enhance the risk analysis process. The extent to which model-based risk analysis methods can improve privacy and security risks analysis in IDMSs have not been the main focus of current research. Furthermore, current risk analysis methods for IDMSs mainly rely on manual inspections [3]. Manual inspections are time consuming and expensive.

This article contributes by introducing the executable model-based risk analysis method (EM-BRAM) for IDMSs. The EM-BRAM identifies risk factors inherent in IDMSs and uses them as inputs for a Colored Petri Nets (CPNs) [4] model of an IDMS system to analyze the system's privacy and security risks.

The rest of the article is organized as follows: Section 2 discusses the related work. Section 3 presents part of the privacy and security risks model for IDMSs. Section 4 is a case study on how the risk analysis method works. Finally, Section 5 concludes the article.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 48–61, 2013.

## 2 Related Work

Gajek et al. [5] analyze the Microsoft CardSpace IDMS or identity metasystem. The analysis focuses on how the vulnerabilities of a browser can threaten the security of the Microsoft CardSpace. They describe an attack where an adversary extracts and replays a security token from the protocol execution and thereby enables possible impersonation of an end-user. Gajek et al. observed that the CardSpace tokens contain end-users' claim but not their identity (ID). This contributes to identification risk in the CardSpace IDMS. In addition, end-users are not involved in the protocol execution. Thus, end-users tokens or credentials are encrypted with the relying party's public key and signed by the identity provider without their involvement. Furthermore, an attacker can subvert the same-origin policy (SOP) checks in order to acquire the privilege to access the CardSpace token. Similar manual risk analysis of the security assertion markup language (SAML) single sign-on IDMS was done by [6]. However, manual inspections are time consuming and therefore expansive. Rather than manual analysis, this article attempts to automate privacy and security risks in IDMSs. In addition, the above approaches are not model-based and therefore lack the benefits of model-based risk analysis.

Current risk analysis methods for IDMSs are mainly qualitative, rely on manual inspections and incomplete because the stakeholders' interests are ignored [3]. However, the metric-based framework proposed by Cabarcos et al. [3] for IDMSs has no intuitive risk or system model that can help stakeholders to understand the risk analysis process. The EM-BRAM is intuitive, partially automated and can reduce subjectivity in risk analysis.

Suriadi et. al. [7] formally evaluated two security and privacy goals of IDMSs. They showed that end-users could maintain anonymity throughout multiple single sign-on sessions and minimize the ability of IdPs and SPs linking their activities in their proposed user-centric federated SSO system. However, their technique is not comprehensive because it focuses on only two out of many privacy and security goals.

## 3 Risk Analysis Model

This section presents part of a risk analysis model for privacy and security risks analysis in IDMSs. The full risk model can be found in [9]. The risk model is developed from a Delphi study on characteristics of tokens or information that flow in IDMSs [8]. We studied tokens because they are personal data sources and gateway to personal data [10]. A token can be an identifier such as username, a claim such as a password, an assertion such as SAML tokens, a credential such as a X.509 certificate or combinations of these.

The Figure 1 represents the partial risk model for IDMSs. It focuses on the characteristics of tokens that can threaten privacy and security in IDMSs. The external factors are threats or vulnerabilities that may be outside the control of IDMSs. On the other hand, the internal factors are threats or vulnerabilities that

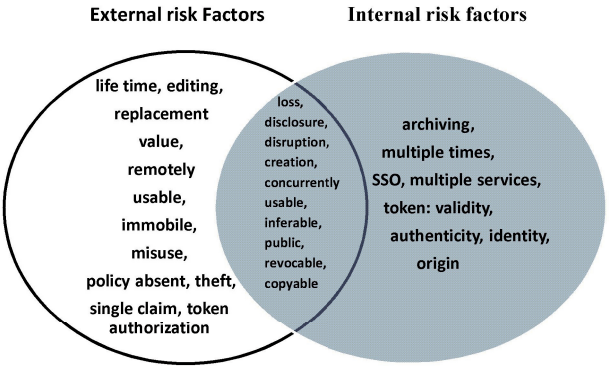



**Fig. 1.** Internal and External Risk Factors [8]

may be under the control of IDMSs and could be verified by technical means. The intersection represents both internal and external factors.

The risk analysis focuses on the internal factors and we discuss them in Table 1.

## 4 Case Study and Application

As a case study, we apply the risk model in Figure 1 to analyze privacy and security risks of SAML SSO service (SAML-SSOS) for Google Apps [16]. The attack scenario for the SAML-SSOS IDMS is shown in Figure 2 and explained in the list below:

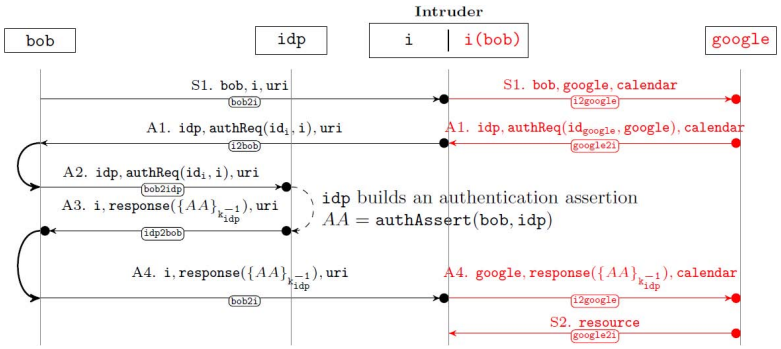



**Fig. 2.** SAML SSO service for Google Apps and an Intruder [17]

**Table 1.** Risk Factors for IDMSs [9]

| **Factors** | **Explanation** |
|---|---|
| multiple times | In an IDMS, the activities of an end-user may be linked or profiled when she uses a token multiple times. Hence multiple uses of tokens create linkability or confidentiality risk |
| single sign-on,multiple services | A token used for multiple purposes or services may be subjected to illegal processing or abuse. IDMSs that support single sign-on (SSO) allow tokens to be used for multiple services sometimes in multiple domains upon a single authentication [11]. Although SSO reduces human error, it leads to sharing of valuable information across services or domains [12]. |
| creation, archiving | The creation risk factor verifies if a token is created with sensitive personal data and its number of attributes is sufficient to protect the security and privacy of an end-user. A token created with limited or less sensitive attribute may enhance privacy because personal attributes are minimized [12]. Similarly, archiving a sensitive or excessive collection of personal attributes may lead to privacy risk |
| public, inferable, revocable | A token's secret is public if it can be found in an unauthorized or public database. Revealing a token secret to an unauthorized entity creates risk in the IDMS. A token's secret is inferable if it can be guessed or deduced. We can determine if a token's secret is inferable by computing its entropy [13], [14]. The entropy of a token is given by $H_t = -\sum_{i=1}^{N} p_i log(p_i)$ where $p(i)$ are the probabilities of individual characters in the token's secret string and $N$ is the characters space. The entropy of a password secret is given by $H = nlog_2 b$ where $b$ is the character space and $n$ is the password length [13]. For example, the character space for English keyboard is 94. The entropy of a biometric template can be found in [14]. When a token's secret is revoked the user of the token could be identified or confidential information may be made available to unauthorized persons |
| copyable, concurrently usable | If the content of a token is not protected from adversaries then it can be copied. For example, the content of a RFID tag with no additional security could easily be read by anyone with an appropriate reader but a RFID tag that comes with additional security may ensure that only authorized readers have access to its content. A token is "copyable" if its content can be read by an unauthorized agent. This risk can occur externally or internally. Concurrent use of a token may contribute to privacy and security risks if the token is stolen or disclosed without the knowledge of the token owner. On the other hand, concurrent use of token can enhance availability since the token can be used concurrently in many parallel sessions |
| loss, disclosure, disruption | The value at risk when a token is lost, disclosed or disrupted is determined by these factors. Sharing a token in an IDMS can lead to a conflict situation where a token can be lost. A token can be disclosed inside or outside an IDMS. For example, if a token is not encrypted in an IDMS its content can be disclosed. The cost of disclosure may depend on the application using the IDMS. A token can be disrupted in an IDMS if there is a deadlock in the system. This risk can occur externally if the token fails to function. |
| origin, authenticity, identity, validity | To enhance security, an IDMS should have a mechanism for checking the authority who issues a token if the token. In addition, there should be a means of ensuring the validity, identity and authenticity of the token [15]. The authority who issued the token should be clearly identified (token origin). Token authenticity determines if a token belongs to the entity presenting it to the IDMS. Token identity determines if a token identifies the subject or the entity possessing the token. Token validity determines if a token has not expired, its lifespan is within the validity period or has passed the validity test |

1. S1: A user (bob) or browser agent attempts to reach a service hosted by the Intruder SP (IntruderSP).
2. S1: The Intruder SP being aware that bob has a possible subscription with Google requests for Google calendar service from GoogleSP with bob's identity.
3. A1: The Intruder SP waits for an authentication request (authReq) from the Google calendar application or cloud service.
4. A1: Upon receiving the authentication request, the Intruder SP requests for authentication assertion from bob's IdP.
5. A2 and A3: Authentication request is forwarded to the IdP, the IdP builds authentication assertion and sends response to the Intruder SP.
6. A4 and S2: The Intruder SP sends a response to the Google cloud service and receives bob's resources.

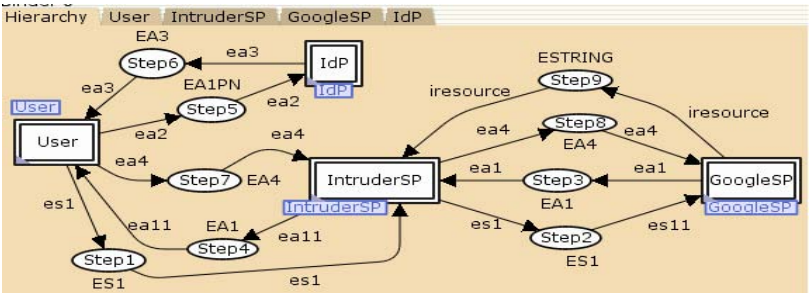



**Fig. 3.** Hierarchical CPNs Model for SAML-SSO Service for Google Apps

The CPNs model in Figure 3 consists of **places, transitions (events), input and output** arcs. We represent the places by ellipses, transitions by rectangles, input/output arcs by directed arcs [4]. A place may hold a collection of tokens and may represent system conditions. A CPNs token is a variable with data type and a value. It is not the same as security tokens discussed above. We refer to the data type as color set and the values as token colors. The set of tokens on all the places at a given moment represents the system state or marking. The transition represents the events or actions that can cause a system to change state. An arc serves as data input and output for a transition. It enables a transition to remove one or more tokens from an input place to an output place. When this happens, we say that the transition is fired.

Figure 3 is the hierarchical CPNs model for the SAML-SSOS attack scenario described in Figure 2. We use the hierarchical CPNs to make a large model manageable and compositional. Figure 3 has five substitution transitions or sub-models. Substitution transitions are the rectangles with double lines while normal transitions are marked with single lines. We use the substitution transitions to represent the system agents – User, IdP, IntruderSP and GoogleSP. The

scenario for each system agent is modeled in the respective sub-model. The IntruderSP represents the intruder (i) in Figure 2. The numbered places represent the possible sequence of information flow in the top-level model. For example, Step1 in Figure 3 represents the first place to receive token in the hierarchical model.

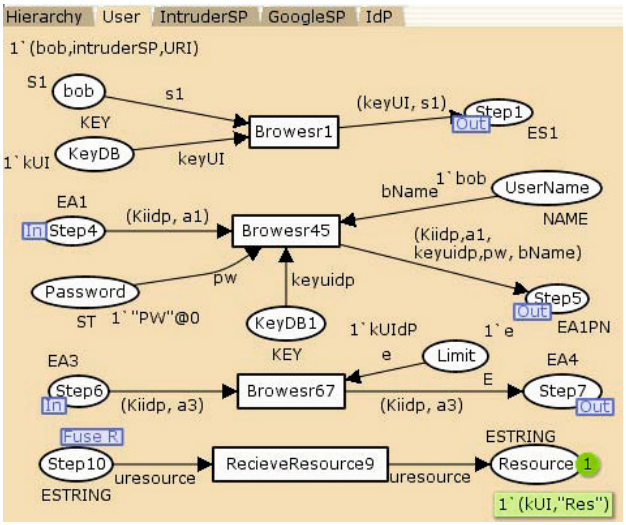



**Fig. 4.** User

The system is initiated by the User sub-model. The User sub-model is in Figure 4. In Figure 4, the place "bob" starts the process by submitting a token (bob,intruderSP,URI) to the IntruderSP. The Browser1 transition moves the token to the output port Step1. The output port Step1 is connected to the Step1 place in Figure 3, the top-level model. The token is sent through Step1 in Figure 3 to the IntruderSP sub-model in Figure 5. The IntruderSP sub-model receives the token through the input port Step1. The transition IntruderSP12 verifies if the IntruderSP has the required session key to decrypt the token. The IntruderSP then creates a new request token for the GoogleSP and sends it through the output port Step2.

In the GoogleSP sub-model in Figure 6, the input port Step2 receives the request. A new authentication request is created and sent via the output port Step3 to the IntruderSP. In the IntruderSP in Figure 5, the IntruderSP modifies the token and sends it to the IdP sub-model via the output port Step4. The token is received by the input port Step4 in the User sub-model. The end-user's authentication information is then added to the token and sent via the Step5 output port to the IdP sub-model.

The input transition Step5 receives the token in the IdP sub-model in Figure 7. The IdP transition authenticates the token and issues an assertion via Step6

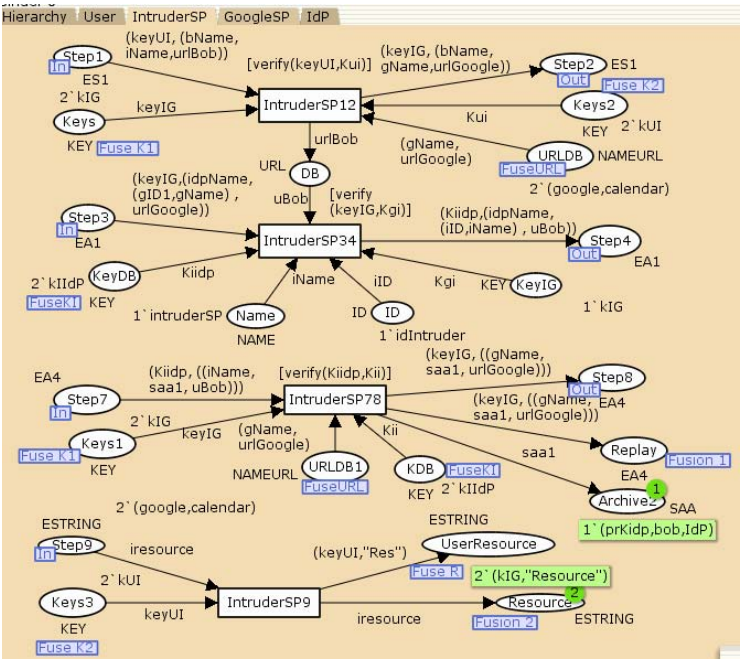



**Fig. 5.** Intruder SP

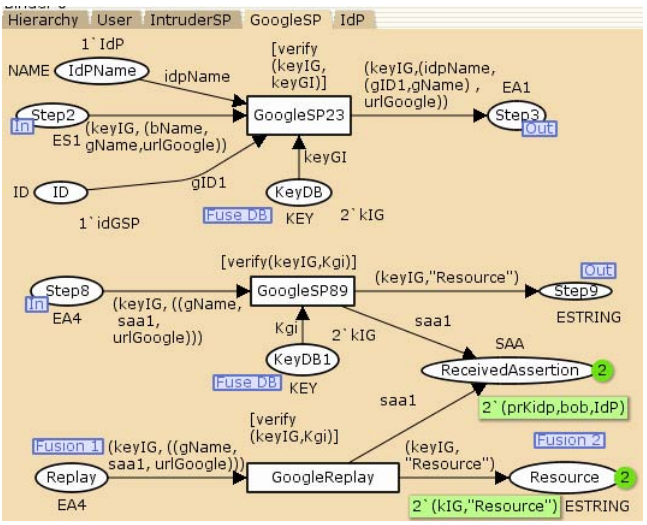



**Fig. 6.** Google SP

output port. The User sub-model in Figure 4 receives the assertion through the input port Step6 and forwards it to the IntruderSP. The Step7 input port of the IntruderSP receives the token, modifies it and sends to the GoogleSP via the Step8 output port. In addition, it uses the fusion set Replay to resend the token to the GoogleSP. Furthermore, the IntruderSP archives the token on the place Archive2.

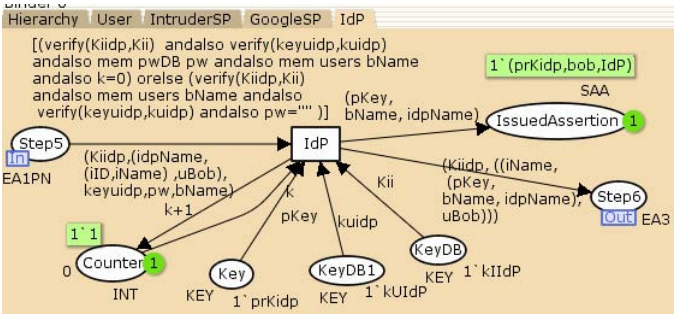



**Fig. 7.** IdP

The GoogleSP sub-model receives the token on the input place Step8. It verifies the session key and releases the resource via the output port Step9 if the verification is successful. In addition, it stores the assertions on the ReceivedAssertion place. The GoogleReplay transition releases a resource whenever a new token is received on the fusion set Replay. Finally, the IntruderSP receives the unauthorized resources.

The tokens of the IDMS are composed of fields. Some of the fields are atomic others are structured. The atomic fields or data types include keys and IDs. Structured fields are constructed from the atomic ones. For example, a cipher or an assertion is given by the order pair (K,I) where K is an encryption key and I is an identity.

The data types or colors are declared as follows:

```
colset KEY = with  prKidp|kGP|kPI|kUI|kUIdP|kIG|kIIdP; colset ST
= string timed;colset URL= with URI|calendar;colset ID = with
idIntruder|idISP|gsp|idGSP;colset EA2=EA1; colset NAME = with
intruderSP|google|IdP|bob|alice;colset A2=A1;colset ESTRING =
product KEY *STRING;colset EA1=product KEY*A1; colset S1 =
product NAME*NAME*URL;colset ES1=product KEY*S1;colset NAMEURL=
product NAME*URL;colset SAA=product KEY*NAME*NAME; colset AA =
product ID * NAME;colset A1 =product NAME* AA*URL;colset A3=
product NAME*SAA*URL;colset EA3=product KEY*A3;colset A4=A3;
colset EA4= EA3;colset A1PN=product AA*URL*NAME;colset EA1PN=
product KEY*A1*KEY*STRING*NAME; fun verify(k1:KEY,k2:KEY)=k1=k2;
 val users =[bob,alice];val pwDB=["PW", "PW1"];
```

The color sets with "E" in front of their names represent the encrypted version of the corresponding color sets without an E. E.g. ESTRING is the encrypted version of STRING. The color set of the variable s1 is S1, es1 is ES1, a1 is A1, and ea1 is EA1 etc. The function verify(k1:KEY, k2:KEY) checks if an agent has the required session key to decrypt a token. It returns true if the agent has the required session key.

### 4.1 Privacy and Security Risks Analysis

This section shows how privacy and security risks of the SAML-SSOS model in Figure 3 can be analyzed. The objective of risk analysis is to identify and assess all risks and suggest a set of controls that will reduce these risks to an acceptable level [18]. Hence, we analyze if the characteristics of tokens in the SAML-SSOS IDMS threaten system privacy and security and suggest suitable controls based on the risk model in [9].

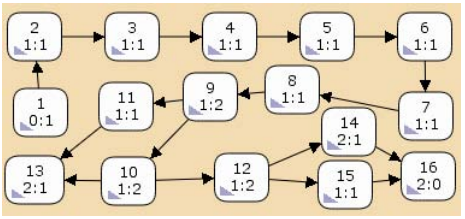



**Fig. 8.** State Space Graph

We use CPNs simulation tools to check the correctness of the model. We then use the CPNTools [4] to generate the state space graph of the system model and search through the graph for privacy and security risks. Figure 8 is the state space graph automatically generated by the CPNTools. It has the 16 nodes and 18 arcs. The nodes represent the system states and the arcs are the transitions from one system state to another. In other words, nodes correspond to the set of reachable markings and the arcs correspond to occurring binding elements. A marking is the number of tokens and the token colors on the individual places which together represent the state of a system.

We analyze the privacy and security risks as follows:

**Multiple times**: The place GoogleSP'ReceivedAssertion in Figure 6 stores all the tokens or assertions received by the GoogleSP. To verify multiple uses of tokens, we use the query "PredAllNodes(multipleUse())" to find the upper integer bound of all the nodes where tokens on the GoogleSP'ReceivedAssertion place is greater than 1. The result shows that multiple uses of tokens occurred at nodes 13,14 and 16. This means the end-user can be profiled by the GoogleSP, hence we have profiling or linkability risk.

```
fun multipleUse()=fn n=>size(Mark.GoogleSP'ReceivedAssertion 1
n)>1;
```

**SSO/Multiple Services:** The alias or the identity "bob" is not supposed to be seen by the GoogleSP because "bob" requested a resource from the IntruderSP. However, because the end-user used the same identity "bob" for the two services, it was easy for the IntruderSP to mount an attack using the identity of "bob". Hence, the end-user's token can access the calendar service without her consent or knowledge. We verify this risk using the query "PredAllNodes(isMultipleService())". The function "isMultipleService()" is defined below. The results show that the identity "bob" appears in GoogleSP's domain at nodes 10,11,12,13,14,15 and 16.

```
fun isMultipleServices()=fn n=> isSubstring "bob"
(st_Mark.GoogleSP'ReceivedAssertion 1 n);
```

**Creation/Archiving:** We verify whether the token has the recommended number of attributes in every state of execution and contains no sensitive attributes. Sensitive attributes include criminal record, BankID, health status etc. [19]. The SAML-SSOS IDMS requires four attributes for assertions i.e. AuthAssert(ID,User, IdP, SP) [17], and at least two attributes (Entity's name and URL) to be secure. The query (a) retrieves all the attributes or the binding elements of the model. The binding element can be analyzed to verify whether the number of the attributes is six as recommended or more than sufficient.

The query (b) verifies whether the token attributes contain sensitive data. For example, query (b) searches through all the states of execution to find if the sensitive data "bankID" is one of the attributes. The query returns an empty list which indicates that "bankID" is not an attribute in any binding elements. We assume that the attributes have standard names.

```
SearchNodes (EntireGraph, fn n => (length(OutArcs(n))>=0),
NoLimit, fn n => ArcDescriptor n,[],op ::) --(a)
SearchNodes(EntireGraph,fn n =>isSubstring "bankID"
(ArcDescriptor n),NoLimit,fn n=>ArcDescriptor n,[],op ::) --(b)
```

The *SearchNodes* function in the queries (a) and (b) traverses the nodes of the state space. It has six arguments. The first argument specifies the part of the state space to be searched. E.g the "EntireGraph" argument means search the entire state space graph. The second argument maps each node into a boolean and uses the nodes that are true for the analysis. The third argument specifies the number of times the predicate function (e.g. $fn\ n => (length(OutArcs(n)) >= 0$) can evaluate to true before it terminate. "NoLimit" means unlimited times. The fourth argument is the evaluation function. It analyzes the nodes selected by the second arguments. The fifth argument specifies a constant. The constant enables the last argument to combine the results obtained from the fourth and the fifth arguments [20].

To verify if the tokens of the initial marking can be archived, we use the reachability analysis. We verify if a token can reach any of the Archive places with the function "Reachable(1,16)". The function returns true, which means that, the tokens of the initial markings can reach the last node which has non-empty colors in their archive places. Hence, the tokens of the initial marking can be archived.

**Public:** We verify if the end-user's password can be found outside the User and the IdP domain or unauthorized domain using the query below:

```
SearchNodes (EntireGraph, fn n=>(isSubstring "PW"(ArcDescriptor n)
andalso isSubstring "IntruderSP" (ArcDescriptor n)) orelse
(isSubstring "PW" (ArcDescriptor n) andalso isSubstring "GoogleSP"
(ArcDescriptor n)), NoLimit, fn n => n, [],op ::)
```

The result shows that the password "PW" is not found outside the User and IdP domains but occurs in state 5 and 6 in the IdP and User domains. Hence, the password between the User and the IdP is not public or it is in an unauthorized domain. The inferable password can be computed outside the system model.

**Copyable/Concurrently Usable:** We use the following query to determine if the token is copyable.

```
SearchNodes (EntireGraph,fn n => (isSubstring "empty" (st_Mark.
Hierarchy'Step1 1 n))=false andalso (isSubstring "k" (st_Mark.
Hierarchy'Step1 1 n))=false,NoLimit, fn n => st_Mark.Hierarchy'
Step1 1 n, [],op ::)
```

The query is repeated for all intermediate nodes. The query retrieves all tokens that pass through the intermediate places (Step1,..Step9) and verifies whether they are encrypted. The encryption keys begin with the letter "k". If a token is not encrypted on an intermediate place then such token is copyable because it flows in plain text. The results of the query is empty list which indicates that no unencrypted token passes through "Step1".

Secondly, we verify if a session key of a token can be found in an unauthorized database using the query below. The secret keys used in the sessions are kUI, kIG, kUIdP and kIIdP. kUI is the secret key for User and IntruderSP while kIG is that of the GoogleSP and IntruderSP. The session key kIIdP is for the IntruderSP and IdP. kUIdP is for the User and the IdP.

```
SearchNodes (EntireGraph, fn n => (isSubstring "kUIdP"
(ArcDescriptor n) andalso isSubstring "IntruderSP"
(ArcDescriptor n) ) orelse (isSubstring "kUIdP"
(ArcDescriptor n) andalso isSubstring "GoogleSP"
(ArcDescriptor n)),NoLimit, fn n => n, [], op ::)
```

The query above returns empty list indicating that the secret between the end-user and the IdP ("kUIdP") is not found outside the two domains. Hence, the tokens are not copyable. We can repeat the query for other entities in the model.

For concurrently usable of tokens, we have found that the model can hold more than one token using the PredAllNodes(multipleUse()) query. Upper integer bound also determines system concurrency [4]. Hence, the model supports concurrent use of tokens.

**Loss, Disclosure/Disruption:** The function "ListDeadMarkings()" returns the node 16 as the only dead node. In addition, the function "Terminal(16)" returns true which indicates that node 16 is the terminal node. Furthermore, the output from the "print(NodeDescriptor 16)" function shows that all the places in the system are empty except the archive places. This means that all the tokens have been received at node 16, hence, there was no conflict, deadlock or an attack in the system that could lead to the loss of tokens.

We have already shown that no token is copyable; this means that no token can be disclosed in the model. Moreover, no token is disrupted in the model because the only deadlock in the model occurred at the last node where the model terminates.

**Token's Origin, Authenticity, Identity/Validity:** Token's origin can be found in the states by examining the markings of the intermediate places Step1 ..Step9, using the query below.

```
SearchNodes (EntireGraph, fn n =>  (isSubstring ‘‘empty’’
(st_Mark.Hierarchy’Step1 1 n))=false, NoLimit, fn n =>
st_Mark. Hierarchy’Step1 1 n, [], op ::)
```

The query is repeated for all the intermediate places. The query retrieves all tokens that pass through the intermediate places. This can be examined for originators of the tokens. For example, a token from the IdP to User must contain the originator of the token which is the IdP.

Token authenticity requires that tokens are not forged or belong to the entity presenting it. We verify this factor by comparing the tokens (assertions) issued by the IdP to the tokens received by the GoogleSP using the following query:

```
fun auth()=(Mark.IdP’IssuedAssertion 1 16)=
(Mark.GoogleSP’ReceivedAssertion 1 16);
```

The query returns false which indicates that some of the tokens (assertions) belong to a different entity or were forged. This risk occurs because the model does not verify the authenticity of the assertions.

We use the query below to verify whether identification of the end-user was successful. The query verifies if the authentication was successful and the assertions contain the identity of the end-user "bob". The query returns true which indicates that the assertions contain the identity of the end-user. A false result will require further examination of the tokens on the two places IdP'IssuedAssertion and IntruderSP'Archive2 to ascertain whether the inconsistency was not caused by the identities.

```
fun isIdentity()=(Mark.IdP’IssuedAssertion 1 16)=(Mark.
IntruderSP’Archive2 1 16) andalso isSubstring "bob"
```

```
(st_Mark.IdP'IssuedAssertion  1 16) andalso isSubstring
"bob" (st_Mark.IntruderSP'Archive2 1 16);
```

We can model the token validation by introducing additional time field in the assertion issued by the IdP. This will then be validated by the GoogleSP. To simplify the model, the time field is not considered. This enabled the IntruderSP to use replay attack (IntruderSP'Replay) to successfully access a resource. We conclude that token validation failed in the model.

**Table 2.** Risk Analysis Report

| **Factors** | **Risk Value** | Meaning |
|---|---|---|
| multiple times | Yes | Tokens can be linked or profiled by a SP |
| single sign-on/ multiple services | Yes | Tokens can be linked or profiled by different SPs |
| creation | Yes<br>No | Token has insufficient number of attributes<br>Token has no sensitive attributes |
| archiving | Yes/No | Token can be archived by SPs but token is not sensitive |
| public | No | Token secret is kept private between end-users and IdP |
| inferable | Yes | Tokens' secret can be guessed by SPs |
| copyable | No | Tokens cannot be copied outside their requested domain |
| concurrently usable | Yes/No | Token can be used concurrently |
| loss | No | Tokens cannot be lost in the IDMS |
| disclosure | No | Tokens cannot be disclosed in the IDMS |
| disruption | No | Tokens are not disrupted by conflict or deadlock in the IDMS |
| origination | No | Tokens' originators are included in the information flow |
| authentication | Yes | Tokens' authentication failed |
| identification | No | Assertions include the identity of the end-user |
| validation | Yes | Tokens' validation failed |

Table 2 is the summary report of the risk identified in the vulnerable SAML-SSOS for Google Apps IDMS. The "Yes" in column two of the table indicates possible privacy or security risk in the IDMS. The "Yes/No" means the privacy or the security risk depends on the security goal being protected. The risks are caused by system vulnerabilities in the IDMS. The possible controls (use cases) for risk mitigation are proposed in [9].

## 5 Conclusion

This article introduces the executable model-based risk analysis method (EM-BRAM) for identity management systems (IDMSs). The method identifies risk factors inherent in IDMSs and uses them as inputs to a Colored Petri Nets (CPNs) model of a targeted IDMS to analyze the system's risk. The method is applied to analyze privacy and security risks of the SAML single sign-on service (SAML-SSOS) for Google Apps. The EM-BRAM provides an initial step towards a comprehensive model-based risk analysis method for IDMSs. The method is partially automated and has the potential of reducing subjectivity in risk analysis.

## References


[1] Audun, J., Simon, P.: User centric identity management. In: AusCERT Conference (2005)

[2] Lund, M.S., Bjørnar Solhaug, K.S.: Model-Driven Risk Analysis, The CORAS Approach, 1st edn. Springer (2011) 978-3-642-12322-1

[3] Cabarcos, P.: Risk assessment for better identity management in pervasive environments. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 389–390 (2011)

[4] Kurt, J., Lars, K.M.: Colored Petri Nets: Modelling and Validation of Concurrent Systems: Modeling and Validation of Concurrent Systems. Springer, Heidelberg (2009) ISBN:978-3-642-00283-0

[5] Gajek, S., Schwenk, J., Steiner, M., Xuan, C.: Risks of the cardspace protocol. In: Samarati, P., Yung, M., Martinelli, F., Ardagna, C.A. (eds.) ISC 2009. LNCS, vol. 5735, pp. 278–293. Springer, Heidelberg (2009)

[6] Gross, T.: Security analysis of the saml single sign-on browser/artifact profile. In: Proceedings of the 19th Annual Computer Security Applications Conference, pp. 298–307 (2003)

[7] Suriadi, S., Foo, E., Jøsang, A.: A user-centric federated single sign-on system. J. Netw. Comput. Appl. 32, 388–401 (2009)

[8] Paintsil, E.: Evaluation of privacy and security risks analysis construct for identity management systems. IEEE Systems Journal PP(99), 1 (2012)

[9] Paintsil, E.: A model for privacy and security risks analysis. In: 2012 5th International Conference New Technologies, Mobility and Security (NTMS), pp. 1–8 (2012)

[10] Naumann, I., Hogben, G.: Privacy features of european eid card specifications. Technical Report 1.0.1, ENISA (2009)

[11] WP3: D3.1: Structured overview on prototypes and concepts of identity management systems. Deliverable 1.1, Future of Identity in the Information Society (2005)

[12] Maler, E., Reed, D.: The venn of identity: Options and issues in federated identity management. IEEE Security and Privacy 6, 16–23 (2008)

[13] NIST: Electronic authentication guideline. Technical Report 1.0.2, NIST Special Publication 800-63 (2006)

[14] Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. IBM Systems Journal 40(3), 614–634 (2001)

[15] Mac Gregor, W., Dutcher, W., Khan, J.: An Ontology of Identity Credentials - Part 1: Background and Formulation. Technical report, National Institute of Standard and Technology, Gaitersburg, MD, USA (2006)

[16] Google: SAML Single Sign-On Service for Google Apps (2012), `https://developers.google.com/google-apps/sso/saml_reference_implementation`

[17] Armando, A., Carbone, R., et al.: Formal analysis of saml 2.0 web browser single sign-on: breaking the saml-based single sign-on for google apps. In: Proceedings of the 6th ACM Workshop on Formal Methods in Security Engineering, FMSE 2008, pp. 1–10. ACM, New York (2008)

[18] Gerber, M., von Solms, R.: From risk analysis to security requirements. Computers and Security 20(7), 577–584 (2001)

[19] Yamada, et al.: Information security incident survey report. Technical report, NPO Japan Network Security Association, JNSA (2006)

[20] Kurt Jensen, S.C., Kristensen, L.M.: Cpn tools state space manual. Technical report, University of Aarhus (2006)

# Preserving the User's Privacy in Social Networking Sites

Alexandre Viejo*, Jordi Castellà-Roca, and Guillem Rufián

Departament d'Enginyeria Informàtica i Matemàtiques,
UNESCO Chair in Data Privacy, Universitat Rovira i Virgili
Av. Països Catalans 26, E-43007 Tarragona, Spain
alexandre.viejo@urv.cat

**Abstract.** In the last years, social networking sites (SNSs) have enjoyed an undeniable success. Those web platforms have huge quantities of active users sharing lots of information everyday. Usually, user-generated content may be almost innocuous, however, some studies have shown that it may also contain very sensitive personal data. This situation may pose a serious privacy threat to the users due to the fact that third parties can gather and exploit that knowledge for their own benefit. There are some proposals in the literature that try to address this situation. Nevertheless, they fail to provide a practical solution capable of working with well-known SNSs. In this paper, we propose a new scheme that fills this gap. More specifically, we present a privacy-preserving system that enables the users to decide which individuals (e.g., other users, third parties or even the SNS itself) can access to their user profiles. We have implemented our scheme to be used by Facebook users. We have run some tests with our prototype and the results show that the added overhead is affordable.

**Keywords:** Privacy, Confidentiality, Social Networks, Access Control, Facebook.

## 1 Introduction

Social networking sites (SNSs) are the most representative result of the rise of the Web 2.0 and its related technologies. In these environments, users publish and share information and services that can be easily accessed by a global audience.

The success of these platforms can be effectively measured in terms of number of users, and the results are really stunning. Specifically, main players like Facebook or Twitter claim to have more than 800 and 100 million active users respectively [1]. More impressive is the fact that those numbers grow each day and their limit cannot be still envisaged.

With such a huge quantity of users and so many different activities available on SNSs, the amount of user data which can be gathered from those places is especially large and heterogeneous. Particularly, user-generated content may

* Corresponding author.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 62–73, 2013.

reflect general opinions and information which can be considered innocuous but it also might contain very sensitive personal data. In this way, the *Consumer Reports'2010 State of the Net analysis* [2] states that more than half of users of social networks share private information about themselves online.

The existence of sensitive information among the data publicly shared by the users may represent a relevant privacy threat due to the fact that third parties can gather and exploit that knowledge for their own benefit. More specifically, leakage of personal data, especially one's identity, may invite malicious attacks from the cyberspace (*e.g.*; personalized spamming, phishing, etc) and even from the real world (*e.g.,* stalking) [3].

Recently, these privacy concerns have been reported to negatively affect the way the users use SNSs. In this way, a survey presented in [4] shows a strong association between low engagement and privacy concern. Specifically, users who report concerns around sharing control, comprehension of sharing practices or general SNS privacy concern, also report consistently less time spent as well as less posting, commenting and "Like"ing of content. This situation can be harmful for the SNSs since their business model requires large quantities of users generating new content without limit.

Therefore, in the last years, the SNSs themselves have provided some privacy settings for their users that allow them to set the privacy level of their online profiles and to disclose either some or none of the attributes in their profiles [5]. However, this privacy-preserving approach suffers from two main problems: (i) these privacy settings are generally not sufficiently understood by the average users who seldom change the default configuration [6](according to [7], this configuration generally makes most of the user information public [7]; and (ii) this method does not prevent the SNS itself from gathering the sensitive user data, in fact, a relevant percentage of the users are worried about how SNSs protect their privacy [8] due to the fact that they are aware of their data being exploited by advertisers [9].

Due to the fact that the companies that support SNSs are not fully reliable in terms of protecting the user's privacy, in order to limit the privacy problems that have been stressed above, it is necessary to design new privacy-preserving mechanisms intended to be deployed and managed by the users themselves.

### 1.1 Contribution and Plan of This Paper

In this paper we propose a new scheme that enables the users of SNSs to decide exactly which individuals can access to their published sensitive data. This implies that other users, third parties or even the SNS itself cannot obtain any protected information if this is not explicitly allowed by the owner. Obviously, this approach does not rely on the active collaboration of the SNS.

The target platform of this proposal is a typical SNS where the user has a profile, a list of friends and a place to publish photographs or images (e.g., Photo Album or similar). Even though the proposed mechanism can be deployed in any SNS that fulfills those requirements, in this work, we have implemented it to be used with Facebook.

Section 2 introduces the state of the art related to the privacy-preserving approaches which can be found in this field of research. Section 3 introduces the system model. Section 4 details our new proposal. Section 5 evaluates the runtime cost of the proposed scheme. Finally, Section 6 reports some concluding remarks.

## 2 Previous Work

The use of user privacy policies (in the form of a contract) to ensure the proper protection of private data is one of the main approaches to provide privacy-preserving SNSs. For example, the works presented in [10,11,12] follow this idea. The main shortcoming of these proposals is that SNSs are supposed to implement such policies to improve the privacy of their users and this is currently unrealistic. Under this research line, it is worth to mention the existence of Persona [13], a social network integrated with Facebook as an application to which users log in through a Firefox extension. In Persona, users define a privacy policy that manages access to their information. As a result, only users of Persona with the necessary access rights can get the protected data. Nevertheless, this tool is only a Facebook application that can be easily removed by Facebook from the applications directory.

Other researchers have focused on designing new SNSs that effectively address the privacy concerns of the users. These platforms generally trust on a completely distributed architecture. Diaspora [14] is a clear example of that. This SNS is a privacy aware, personally controlled, do-it-all distributed open source social network. This project is described as a network that allows everyone to install their own "seed" (*i.e.*, a personal web server used to store photos, videos and everything else) within the larger network. That seed is fully owned and controlled by the user, so the user can share anything and still maintain ownership over it. In this way, the social network gives individuals control over their personal information without being subjected to changing privacy policies and sell-outs to third parties [15]. Diaspora is not the unique system that follows this approach. Other privacy-preserving SNSs based on p2p architectures have been proposed in [16,17,18]. However, the main drawback of all these systems is that they will be hardly adopted by the mainstream audience. Note that centralized SNSs like Facebook and Twitter are very well established and it is quite unrealistic to assume that a new competitor without a very strong company behind will get enough users to represent a proper alternative.

Focusing on privacy-preserving approaches that can be integrated with traditional SNSs, a straightforward solution to prevent any unauthorized entity from accessing the protected user data is using cryptography primitives to cipher any text or attribute before publishing it. Applying this method, only the individuals with the correct cryptographic keys will be able to access the protected content. Nevertheless, this solution is quite problematic because, usually, registering on well-know social networks under a pseudonym, or obfuscating personal information in any way is forbidden by the terms of service. More specifically,

Facebook (the target platform of this work) has banned users who have violated those terms [19]. According to that, the ideal privacy-preserving method should generate protected data that does not look suspicious to the SNS, this is, the information to be published must look real while being incorrect (or, at least, partially incorrect).

Following this idea, the authors in [20] present a scheme that, first, divides the private data into atoms and, then, replaces each atom with a corresponding atom from another randomly selected user who uses the same application. Two significant shortcomings of this proposal are: (i) it requires a certain number of users to provide anonymity; and (ii) it requires some external infrastructure that keeps the relations between the users and their atoms.

A similar proposal is introduced in [21]. This is a Firefox extension that allows users to specify which data or activity need to be kept private. The sensitive data is substituted with fake one, while the real data is stored in a third party server that can be only accessed by the allowed users. Like in the former proposal, one of the main shortcomings of this scheme is that it relies on a centralized infrastructure that must be honest and always available.

Finally, [22] addresses the problem of the centralized infrastructure by locally storing the real data on the allowed friends' machines. In this way, only fake information is stored on Facebook. When a user using this scheme browses a profile of another user who also uses this system, a software component is in charge of transparently showing the real information stored locally, instead of the one actually published on the SNS. This solution requires the users to always connect to the SNS using the computer that locally stores the real data. This may be a main problem for certain users. Moreover, whenever a certain user modifies her protected information, it has to be individually sent to all authorized friends. This issue is not quite efficient in terms of bandwidth usage and it might generate some unstable situation where not all the authorized recipients would have access to the newest information. Also, this solution requires the users to store in their own computers unspecified quantities of information related to others. Some users may feel uncomfortable with this situation, while others might not be willing to spend their storing resources on this task.

In order to solve all these issues, the authors in [22] propose to store all the protected information steganographed within images published in the SNS. Even though this idea is quite promising, the authors do not develop it in their work and it is even not considered for future work.

## 3 System Model

As explained previously, we propose a new privacy-preserving scheme that enables the users of SNSs to decide exactly which individuals can access to their published sensitive data.

We next detail the kind of SNSs which can be the target of our proposal. Then, the requirements of the designed system are provided. Finally, we briefly describe how our system works and its architecture.

### 3.1 Target SNS

Our work has been designed at high level to be integrated with any SNS that offers the following assets: (i) a user profile; (ii) list of friends; and (iii) place to publish photographs or images (e.g., Photo Album).

Due to the fact that Facebook is a really well-known SNS that properly fulfills all those requirements, we have chosen this platform to implement our proposal and retrieve some empirical results. Accordingly, Facebook is considered the target SNS in the rest of this document.

### 3.2 System Requirements

At the current development stage, the main target of the proposed scheme is to enable users to only protect their personal data which appears in their "User Profile" section of the SNS. This implies that hiding other sources of information such as the list of friends or the timeline/wall (i.e., a section of the SNS where users and friends publish text and images) is left for future work.

A complete user profile in a SNS such as Facebook reveals a lot of sensitive information from the owner: gender, date of birth, current location, religious or political views, current and past jobs, interests, education, marital status, etc. This fact clearly stresses the relevance of preventing any unauthorized entity from freely gathering information from this source of data.

Personal data must be protected but this must be done in a transparent way from both the point of view of the users and the SNS. In the case of the users, nowadays, a huge quantity of them are already used to interact with classic SNS (like Facebook) in a determinate way. Therefore, in order to be fully adopted, any privacy-preserving solution should not interfere (or interfere the least possible) in the fixed routine of the users. Regarding the point of view of the SNS, we have explained previously that Facebook (or other similar platforms) does not allow its users to publish fake information in their accounts. Therefore, in order to reach its target, the privacy-preserving mechanism must publish fake data that looks real in front of the SNS.

### 3.3 Our Scheme in a Nutshell

The main idea behind the proposed system is to replace the sensitive data that can be found in the "User Profile" section of a SNS with fake information introduced by the user herself. The proposed scheme first uses cryptography to protect the original sensitive information and, then, it hides the ciphered data in a certain image by means of steganography. Access-control techniques are applied to allow only certain users to retrieve the original information. The resulting image is finally published in the place reserved by the SNS to publish images (e.g., Photo Album).

When any entity (e.g., users, external third-party, the SNS itself) tries to read the "User Profile" of a protected user, two main situations may apply depending on whether this entity is aware of the privacy-preserving system used or not:

- *The reader is not aware.* In this case, this entity only obtains the fake information introduced by the user who runs the privacy-preserving method. If the target SNS does not allow users to obfuscate their personal information, the introduced fake data must look real in order to fool it.
- *The reader is aware.* In this case, the reader looks in the Photo Album for the image that contains the real information (i.e., the stego-object), obtains the ciphered data and applies its cryptographic material to retrieve the authentic user profile. At this point, the access-control method grants or revokes the reader depending on whether it has been authorized by the user running the privacy-preserving system or not.

### 3.4 Proposed Architecture

The general structure of the proposed solution is depicted in Figure 1. Next, the main parts of the proposed architecture are briefly described.

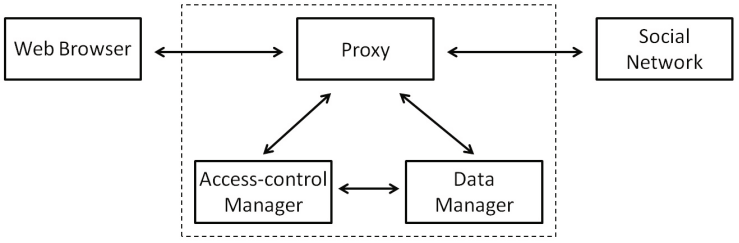



**Fig. 1.** Structure of the application

- *Proxy.* This module is the core of the application. Its target is to capture the HTML requests and responses that are transmitted between the *Web Browser* and the *Social Network* server (i.e., Facebook) and modify them in order to show the real data to the user in a transparent way. This implies that the user is not aware whether the real data is directly obtained from the user profile stored in the SNS or from the stego-object published in the Photo Album.
- *Data Manager.* The objective of this module is to manage which data is published in the SNS and which one is really shown when browsing *Facebook*. When a user wants to protect her profile, this module generates a stego-object and uses the *Proxy* module to publish it. On the other hand, when a user is browsing the protected profile of another individual, this module obtains the real information from the stego-object published in the Photo Album and submits this data to the *Proxy* module which is in charge of showing it to the user.
- *Access-control Manager.* This module manages the cryptographic material required to allow users to retrieve protected information. It also encrypts or decrypts information under request of the *Data Manager* module.

# 4 Our Scheme in Detail

In this section, we detail the two main algorithms that are used to protect personal data and retrieve it. After that, we focus on the steganographic technique used to hide the protected information in the SNS and also the method used to perform the cryptographic key management which is essential to perform a proper access control on the protected data. Finally, some deployability issues are discussed.

## 4.1 Proposed Algorithms

The proposed system is formed by two main algorithms that focus on protecting the sensitive data of the user profile and retrieving it. The first procedure is executed by the user who wants to protect her privacy from any other entity of the system (e.g., other users, the SNS, external third parties, etc). The second procedure is run by any authorized user who wants to retrieve the protected information. Both are detailed in the following two subsections.

**Algorithm-1: Protecting Personal Data.** First, let us consider that a user profile $P$ in a SNS is mainly a finite set of items $I$ that provide some kind of information. This is $P = \{I_1, \ldots, I_n\}$. Now, let us assume a user $U_i$ who wants to protect some items of her user profile $P_{U_i}$. In order to do that, $U_i$ executes the following protocol:

1. $U_i$ requests to the SNS the web page that contains her user profile.
2. For each item $I_w$ that $U_i$ desires to keep private, she replaces the existent data with fake information. Note that, if the user is not comfortable with introducing fake data and the item is not mandatory for the SNS (e.g., birthday and gender are mandatory fields in Facebook), it is also possible to leave it blank.
3. $U_i$ selects which users from her list of friends will be authorized to access the protected data.
4. The proposed system builds a XML file $M$ that contains all the real data that must be protected. This file is then encrypted using the AES cryptosystem and the corresponding cryptographic key $K_{U_i}$. This is $C = E_{K_{U_i}}(M)$.
5. The system uses *broadcast encryption techniques* to perform the access control to the protected content according to the selection done by $U_i$ in the step-3. In this way, key $K_{U_i}$, which is required to decrypt $M$, is ciphered according to the selected broadcast encryption technique (see Section 4.3 for more details about this). Let us denote the resulting element as $\lambda$. Note that the use of broadcasting encryption requires $U_i$ to share a set of secret keys with each one of her friends in the SNS.
6. The system uses *steganographic techniques* (see Section 4.2 for more details about this) to hide $C$ and $\lambda$ in a *cover image* $\delta$ provided by $U_i$.
7. Finally, the system publishes the stego-object $\delta$ in the place reserved by the SNS to store images uploaded by users (i.e., Photo Album).

**Algorithm-2: Retrieving Protected Data.** Let us consider a user $U_j$ who is also using the proposed privacy-preserving scheme. This user is browsing the Facebook profile of $U_i$ and wants to retrieve a certain item $I_w$ from $P_{U_i}$. In order to achieve that, $U_j$ executes the following protocol:

1. $U_j$ requests to the SNS the web page that contains the user profile of $U_i$.
2. The privacy-preserving system tries to find a valid stego-object $\delta$ in the place reserved by the SNS to store images uploaded by $U_i$ (i.e., Photo Album). If $\delta$ is not found, it means that all the information published in the user profile is real and, hence, no further works is required and the protocol ends at this step. On the other hand, if $\delta$ is found, the proposed system continues the protocol.
3. The system uses a *steganographic method* (see Section 4.2 for more details about this) to obtain two items from the $\delta$: (i) ciphertext $C$; and (ii) the access control element $\lambda$ that was generated using a *broadcast encryption* method (see Section 4.3 for more details about this) and that contains a ciphered version of $K_{U_i}$.
4. The system uses the set of secret keys shared between $U_j$ and $U_i$ to retrieve $K_{U_i}$ from $\lambda$. If $U_j$ has not been authorized by $U_i$, $U_j$ will retrieve an invalid key and she will be unable to get the real user profile of $U_i$. In other case, $U_j$ obtains $K_{U_i}$, she is able to decrypt $C$ and, hence, she gets the real content $M$ (i.e., $D_{K_{U_i}}(C) = D_{K_{U_i}}(E_{K_{U_i}}(M)) = M$).
5. The system shows the real content to $U_j$ instead of the fake information that is stored in the SNS. All the information is transparently shown to $U_j$ using her own browser.

### 4.2 Hiding Information from the SNS

As explained previously, SNSs generally do not allow their users to publish fake information in their accounts. Therefore, published fake data must look real in front of the SNS and the protected information must be hidden somewhere. In this way, the authors in [22] proposed to store all the protected data steganographed within images published in the SNS itself.

Using certain steganographic methods, a lot of data can be hidden inside standard images. Unfortunately, in this scenario, achieving a good *information rate* is not enough. More specifically, we require a steganographic scheme that also provides imperceptibility and robustness. Moreover, it should be oblivious (the recovery algorithm should not require the original unmarked image).

The well-known *F5* algorithm [23] fulfills the aforementioned requirements, hence, we first used it to hide information in the images uploaded to Facebook. Nevertheless, Facebook applies a heavy compression on the uploaded images, modifies the points-per-inch (ppi) to 72 ppi and changes any embedded profile to sRGB. As a result of all these transformations, no embedded data can be recovered from uploaded images marked with F5.

In order to overcome these difficulties, we developed a new stenographic algorithm robust enough to resist all the modifications currently applied by

Facebook. This algorithm is not the main contribution of this paper and, for this reason, we only give a brief description:

- *Embedding process.* First, the cover image is divided in cells of 8x8 pixels. Then, each cell is analyzed. If a cell is homogeneous (all pixels are similar), one bit of information is embedded, otherwise it is discarded. Finally, for each selected cell, we do the following: if we want to embed a "1", the less significant bits of each pixel are replaced with a certain fixed pattern $a$; otherwise, if we want to embed a "0", these bits are replaced with a certain fixed pattern $b$. Additionally, Reed-Solomon correcting codes [24] are used to improve the robustness.
- *Recovering process.* First, cells containing embedded information are identified. Then, for each one we get "0" or "1" depending on the number of pixels which are closer to pattern $b$ or to pattern $a$. Finally, the correcting-codes retrieve the hidden information.

We have tested this method and it has been able to recover the embedded information from images uploaded to Facebook. Note that it is not the purpose of this paper to study the suitability of other stenographic algorithms present in the literature.

### 4.3 Access Control and Key Management

A practical privacy-preserving scheme should not rely on a central server or require the users to be always on-line. In order to fulfill those requirements, we propose the use of *broadcast encryption* because it allows the owner of the protected data to grant or revoke access to one or several users in an easily way. Additionally, the owner can be off-line (i.e., users who try to get the protected information do not need to establish a direct connection with the owner). Instead of that, all the required access control data can be found embedded in the stego-object, together with the protected information.

In our implementation, we have used the well-known *Subset Difference (SD)* broadcast encryption scheme [25]. The reason is that it is is a particularly efficient scheme that generally requires a small amount of access control data even if there are several revoked users. Note that studying the deployability of other broadcast encryption schemes is out of the scope of this paper.

Finally, it is worth to mention that every user in our system uses a back-office application to interact with the "Access-control Manager". This application allows them to generate cryptographic keys for their friends and deny/grant access to their protected information. Then, these keys can be sent/received by e-mail. The list of friends and their email addresses can be directly found in the SNS.

### 4.4 Deployability Issues

Even though the general idea of the proposed system can be applied to any SNS that offers classic functionalities (such as image uploading support, user profiles,

etc), the implementation is completely platform-dependent due to the specific particularities of the HTML traffic generated by each SNS. This issue is not limited to the deployment of the proposed mechanism in different SNS, in fact, if the Facebook implementation changes, the *Proxy* module already implemented should be adapted to deal with the changes. This implies that a realistic privacy-preserving scheme based in our proposal should be continuously supported (e.g. by the open source community) in order to work properly. This shortcoming is shared with the scheme presented in [22].

## 5 Evaluation

In order to evaluate the performance of the proposed system we have measured the runtime cost of the following tests:

- *Test-1*. Retrieve the "User Profile" web page of a certain Facebook user $U_i$ using a clean Firefox browser.
- *Test-2*. Retrieve the "User Profile" web page of $U_i$ using a Firefox browser that is connected to Facebook through the *Proxy module* proposed in this paper. Note that, in this test, only the *Proxy* module is used.
- *Test-3*. Protect the "User Profile" of $U_i$ using the proposed system.
- *Test-4*. An authorized user retrieves the *protected* "User Profile" web page of $U_i$.
- *Test-5*. A revoked user tries to get the *protected* "User Profile" web page of $U_i$.

All these tests have been run using a computer equipped with an Intel Core i7 at 2.7 Ghz, 8GByte of RAM, Windows 7 and DSL connection 10Mbit/1Mbit. Table 1 shows the different runtimes (in seconds) achieved by each test. The results provided are the average of 100 executions.

**Table 1.** Runtime cost (in seconds) for each test

| Test | Runtime cost |
|---|---|
| Test-1 | 4.886 |
| Test-2 | 5.495 |
| Test-3 | 4.137 |
| Test-4 | 6.903 |
| Test-5 | 5.638 |

It is worth to mention that these results represent the time required to fully download a *complete* "User Profile" web page. This point is relevant because, in addition to the requested profile, a "User Profile" web page also contains additional data such as advertisements, Facebook chat, etc. This fact justifies the 4.886 seconds required by Test-1.

Focusing on the time cost needed to obtain a protected user profile (Test-4), the overhead introduced by the proposed privacy-preserving scheme is around 2.017 seconds. We believe that this cost can be affordable for those users interested in explicitly controlling who can retrieve their personal data. Also, it is worth to mention that this is a first prototype and, probably, there is room for improvement.

## 6 Concluding Remarks

In this paper, we have proposed a new system that enables the users of SNSs to protect their personal data. More specifically, by means of our proposal, they can exactly decide which individuals can access to their published information. As a result, even the SNS that hosts the user data cannot obtain any protected information if this is not explicitly allowed by the user. In addition to that, the new scheme has been designed to work properly with well-known SNSs such as Facebook.

Our scheme has been implemented and tested. We believe that the runtime costs obtained are quite competitive when compared with a direct connection to Facebook. More specifically, the proposed system introduces an approximate overhead of 2 seconds.

Regarding future work, it would be interesting to try to protect other sensitive elements which are present in SNSs such as user publications in the timeline/wall, the list of friends, etc.

**Disclaimer and Acknowledgments.** This work was partly supported by the European Commission under FP7 project Inter-Trust, by the Spanish Ministry of Science and Innovation (through projects eAEGIS TSI2007-65406-C03-01, CO-PRIVACY TIN2011-27076-C03-01, ARES-CONSOLIDER INGENIO 2010 CSD2007-00004, Audit Transparency Voting Process IPT-430000-2010-31, ICWT TIN2012-32757 and BallotNext IPT-2012-0603-430000) and by the Government of Catalonia (under grant 2009 SGR 1135).

## References

1. McMillan, G.: Twitter reveals active user number, how many actually say something. In: Time - Techland (September 2011)
2. Consumer Reports National Research Center: Annual state of the net survey 2010. Consumer Reports 75(6) (2010)
3. Zhang, C., Sun, J., Zhu, X., Fang, Y.: Privacy and security for online social networks: Challenges and opportunities. IEEE Network 24(4), 13–18 (2010)
4. Staddon, J., Huffaker, D., Larking, B., Sedley, A.: Are privacy concerns a turn-off? engagement and privacy in social networks. In: Proc. of the Eighth Symposium on Usable Privacy and Security, SOUPS 2012 (2012)
5. Zheleva, E., Getoor, L.: To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In: Proc. of the 18th International Conference on World Wide Web, WWW 2009, pp. 531–540 (2009)

6. Van Eecke, P., Truyens, M.: Privacy and social networks. Computer Law & Security Review 26(5), 535–546 (2010)
7. Bilton, B.: Price of facebook privacy? start clicking. In: The New York Times (May 2010)
8. Wilson, D.: Users are worried about social network security and privacy. The Inquirer (October 2011)
9. Crimes, S.: Twitter sells old tweets to marketers - should users be worried? The Inquirer (March 2012)
10. Dhia, I., Abdessalem, T., Sozio, M.: Primates: a privacy management system for social networks. In: Proc. of the 21st ACM International Conference on Information and Knowledge Management, CIKM 2012, pp. 2746–2748 (2012)
11. Cheek, G., Shehab, M.: Privacy management for online social networks. In: Proc. of the 21st International Conference Companion on World Wide Web, WWW 2012, pp. 475–476 (2012)
12. Aimeur, E.: Privacy management for online social networks. In: Proc. of the International Conference on Availability, Reliability, and Security, ARES 2010, pp. 172–179 (2010)
13. Baden, R., Bender, A., Spring, N., Bhattacharjee, B.: Persona: an online social network with user-defined privacy. In: Proc. of the ACM SIGCOMM 2009 Conference on Data Communication, SIGCOMM 2009, pp. 135–146 (2009)
14. Diaspora, `http://joindiaspora.com` (last accessed: February 12, 2013)
15. Vaughan-Nichols, S.: Diaspora: It's no facebook ... yet. Computerworld (September 2010)
16. Cutillo, L., Molva, R., Strufe, T.: Safebook: A privacy-preserving online social network leveraging on real-life trust. IEEE Communications Magazine 47(12), 94–101 (2009)
17. Vu, L., Aberer, K., Buchegger, S., Datta, A.: Enabling secure secret sharing in distributed online social networks. In: Proc. of the Annual Computer Security Applications Conference, ACSAC 2009, pp. 419–428 (2009)
18. Nilizadeh, S., Jahid, S., Mittal, P., Borisov, N., Kapadia, A.: Cachet: a decentralized architecture for privacy preserving social networking with caching. In: Proc. of the 8th International Conference on Emerging Networking Experiments and Technologies – CoNEXT 2012, pp. 337–348 (2012)
19. Scoble, R.: Facebook disabled my account. In: Scobleizer (January 2008)
20. Guha, S., Tang, K., Francis, P.: NOYB: Privacy in online social networks. In: Proc. of the First Workshop on Online Social Networks (2008)
21. Luo, W., Xie, Q., Hengartner, U.: Facecloak: an architecture for user privacy on social networking sites. In: Proc. of the 2009 International Conference on Computational Science and Engineering, pp. 26–33 (2009)
22. Conti, M., Hasani, A., Crispo, B.: Virtual private social networks. In: Proc. of the First ACM Conference on Data and Application Security and Privacy, CODASPY 2011, pp. 39–50 (2011)
23. Westfeld, A.: F5 - a steganographic algorithm. In: Proc. of the 4th International Workshop on Information Hiding, IHW 2001, pp. 289–302 (2001)
24. Reed, I., Solomon, G.: Polynomial codes over certain finite fields. Journal of the Society for Industrial and Applied Mathematics (SIAM) 8(2), 300–304 (1960)
25. Naor, D., Naor, M., Lotspiech, J.: Revocation and tracing schemes for stateless receivers. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 41–62. Springer, Heidelberg (2001)

# A Classification of Factors Influencing Low Adoption of PETs Among SNS Users

Konstantina Vemou and Maria Karyda

Department of Information and Communication Systems Engineering,
University of the Aegean, Samos, GR-83200, Greece
{kvemou,mka}@aegean.gr

**Abstract.** Privacy concerns among Social Networking Services (SNS) users are increasing. However, Privacy-Enhancing Technologies (PETs) are not, yet, widely deployed and their deployment rate is marginally growing. This is surprising given the fact that PETs are widely recognized as effective at reducing privacy risks. This paper explores this paradox, by presenting a classification of the key factors influencing the adoption of PETs. The conclusions of our analysis suggest that, certain factors are overemphasized, while the importance of others has been overlooked. Our classification is based on relevant literature and experimental analysis of PETs, and can inform both practitioners for designing and enhancing PETs, as well as researchers, as we identify several open issues.

**Keywords:** Social Network Services, Privacy-Enhancing Technologies, adoption.

## 1 Introduction

Privacy concerns among Social Networking Services (SNS) users are increasing [1], [2], and there is even a small proportion of users who are willing to pay for privacy-friendly services [3]. Privacy research in SNS focuses on developing and applying Privacy-Enhancing technologies (PETs) to support users participating in social networks, while maintaining their privacy. PETs used in the context of SNS include, mainly, attribute-based controls, such as Facecloak, decentralized SNS, such as Diaspora and privacy management applications, such as MyPermissions Cleaner.

However, privacy enhancing technologies are not, yet, widely deployed [3], [4]; moreover the rate at which their deployment has grown over the last few years has not been substantial. This is surprising given the fact that PETs are widely recognized as effective at reducing privacy risks [4], [5]. This paper discusses this paradox and addresses the question why PETs adoption by social network users is so far limited. Understanding this issue and analyzing the underlying causes can serve as a guide for future research and practice, to provide users with more effective and attractive PETs.

To analyze the problem of low adoption of PETs, we have followed a multifaceted approach: First we identified all relevant factors associated with the low adoption of PETs from the extant literature. Then we conducted experiments with several PETs,

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 74–84, 2013.

and evaluated them against these factors in order to derive more insights with regard to their use. This exploration resulted to a classification of key factors influencing low adoption of PETs in SNS, based on literature research and experimental use of PETs by the authors.

The literature analysis allowed us derive important conclusions and identify contradicting findings. For instance, while several papers argue on the importance of users being aware of PETs [6], [7], others suggest that awareness is not associated with their increased deployment. This paper provides a deeper understanding of the issues pertaining to the use of privacy enhancing technologies, whereas current approaches tend to shed light to specific aspects of the issue, while neglecting others.

The contribution of this paper is both theoretic and practical: On a theoretical level, we identify and provide a classification of the key factors influencing the limited use of PETs in the context of SNS. We discuss these factors and show that some may have been overestimated while the importance of others seems to evade researchers' attention. From a practitioner's viewpoint, we illustrate aspects of privacy protection that commonly used PETs fail to meet, thus contributing to users' abstention from the use of privacy preserving technologies.

The paper is structured as follows: in the following chapter, we describe PETs used in SNS. In chapter 3 the classification of key factors affecting PETs adoption by SNS users is presented. The last chapter contains conclusions deriving from our work, as well as highlights of areas for further research.

## 2 Background: Privacy Enhancing Technologies and Social Networks

Privacy concerns related to the use of Social Network Services are increasing [2]. To address these concerns several technological measures have been developed, aiming to protect published information from unauthorized audiences and raise the users' awareness when it comes to sharing personally identifiable information (PII). Such technologies, commonly known as Privacy Enhancing Technologies, or PETs, include a wide range of applications including access control, privacy signaling tools, third party tracking tools, social identity management systems and decentralization of Social Network Services.

PETs used in SNS include attribute-based controls which are based on encryption (e.g. Lockr [8], Persona [9] and EASiER [10]), role-based access controls, based on encryption and/or obfuscation or perturbation (e.g. BlogCrypt [11], FlyByNight [12], Facecloak [13], FaceVPSN [14] and NOYB [15]), and audience segregation (e.g. the Clique Prototype [7]). Another approach aiming at protecting PII via avoiding central repositories has been implemented either as web-based decentralized SNS (Diaspora [16], Vis-à-vis [17], Frenzy [18]) or as Peer-To-Peer SNS (Safebook [19], PeerSoN [20], Life Social [17], Likir [17]).

Privacy signaling technologies such as RMP-Respect My privacy [21] and P3P [22] can also be applied in SNS, while other tools include privacy wizards that help users set their privacy settings (e.g. Collaborative policy analysis [23], PriMa [24],

MyPermissions Cleaner [25], privacyfix [26], Priveazy LOCKDOWN [27]). Privacy Mirrors help users understand which of their personal information is visible to other users (such as Facebook's ViewAs and Search engine profile preview [28], Privacy Mirror [29], Privacy Check [30], PrivAware [31], make myself clear [32]).

There are also Social Network Visualization Tools (such as Vizster [33], Friendwheel [34]) and Personal Containers, which register which information about the user has been published and where they were published (Privacy Delegate [35], Privacy Butler [36]). Last but not least, there are tools that reveal which social networking services track users while surfing the internet (Disconnect [37]).

Despite this plethora of privacy tools, some of which are independent applications while others are embedded into SNS platforms, users still don't seem to be taking advantage of them, despite rising privacy concerns and thriving use of SNS. For instance, relevant literature reports on the limited use of access controls and privacy settings that are provided within the SNS platforms [38], [1], [39], [40]. It should be noted though that perception of low adoption of PETs is based mainly on literature and there is lack of published research and statistics about the use of specific standalone PETs in practice.

But why does this phenomenon happen? Why do so few users employ privacy enhancing technologies? Several reasons have been proposed, including lack of knowledge, lack of skills [5], the time needed to learn a new technology, the complexity of existing technologies [52] the multiplicity of approaches to privacy protection [41], cost [42], usability issues [41], lack of support by the platform [43], users cognitive and behavioral biases [42]. Another aspect of users' paradoxical behavior has been traced in their unawareness of some of privacy threatening e-service aspects [44]. Most relative studies try to answer this question by focusing on a specific aspect of the problem, especially to why some users change their privacy settings within SNS, when this is provided as an option, to limit the audience of what they share, while others don't [45], [5].

Up to now, however, no relevant study has attempted a thorough discussion of all factors contributing to the low adoption of PETs by SNS users. In the following we provide an in-depth discussion of the key factors we have identified through literature review and deployment of a large set of available PETs that are applied by SNS users.

## 3 Key Factors Affecting PETs Adoption by SNS Users

### 3.1 Awareness of Privacy Risks and PETs

It has been suggested that many SNS users are unaware of the existence of some PETs [46]. Moreover, it is often the case that users are not aware of certain privacy-threatening aspects of the services they are using, such as, for instance, privacy dangers deriving from third-party applications [44]. Therefore, they cannot benefit from special purpose PETs, such as those aiming at limiting the access of third party applications to personal information (e.g. MyPermissions Cleaner [25]). Relevant literature, however, also reports findings where individuals despite being aware of PETs, did not use them [4], [47].

Generally, privacy concerns and awareness of privacy risks are considered to contribute to a user's informed decision to reveal PII [6], as well as to take measures against its misuse. This is also true the other way around, as Xu and al. (2009) found the level of privacy concern to be inversely linked to perceptions of control on the flow of information disclosure, including PETs use [48]. The level of privacy concern acts as a motive for PETs use, however it is a weak predictor to the users' decision, as the user faces cognitive and behavioral biases [41].

Conclusively, being aware of privacy tools is an essential prerequisite for their use, awareness is only weakly linked to their deployment. It this thus important, to include other individual as well as technical related factors in order to gain a deeper understanding of the problem.

### 3.2 Requirements for Special IT Skills

Lack of technical skills and the time needed to learn a technology have also been identified as possible inhibitors for the use of PETs [5], [4]. As Yao [49] states, many online privacy protection strategies require technical skills beyond that of an average user, and this is true even for young adolescents [50].

Our experience from testing relevant tools, indicates that SNS users need to be familiar with the not so trivial use of browser extensions in order to use applications such as Priveasy Lockdown [27] and FaceVPSN [14]. Moreover, users need at least basic knowledge of concepts as encryption is applied in most access control solutions. For example, to use Blogcrypt [11] the user has to manually import and export encryption keys. Even worse, to use PETs deployed in distributed social networks, as in the case of Vis-à-Vis [17], users must be able to create and publish their own profile, and maintain them in their personal computer resources.

### 3.3 Complexity and Diversity

The need for privacy protection, including adoption of PETs, stems from a set of multiple and different risks, resulting of different aspects of SNS use, e.g. posting of photographs, chatting, sharing friends list. As a result, different practices are applied to protect a user's privacy, some aiming at awareness and some aiming at information concealment. Moreover, researchers provide different solutions to the same aspects of privacy risks. An example is the implementation of access controls by obfuscation, as in NOYB [15] or encryption, as in StegoWeb [51]. This leaves the user with multiple and diverse tools or technologies to evaluate, in order to choose which one to use, a process that requires a significant amount of time, effort and knowledge. In addition to the diversity of PETs, users encounter complex and unusable interfaces [38] that make the tools difficult to configure [52], thus adding to the difficulty of PETs adoption.

### 3.4 Direct and Indirect Cost

As with other software products, the use of PETs may entail direct cost for acquiring the tool, as well as intangible costs, related to time for learning [53], limited functionality

and usability issues, such as how seamless is the authentication to third-party websites [41] etc.

Most users report they are not always willing to pay for acquiring a privacy tool, despite having privacy concerns [41]. Rose (2005) found that, although most participants in a survey reported being very sensitive to privacy issues, less than half of them would be willing to pay roughly $29 to have their privacy protected by means of property rights on personal information [41]. However, many PETs can be acquired and used with no direct cost.

Moreover, SNS users are not keen of experiencing delays, changing their habits of interacting with an e-service or discounting usability, due to PETs use. For example, a typical Facebook user with an average number of 130 friends [54], who wants to encrypt her posted data using FlyByNight, needs to encrypt messages with each of her friends public keys, thus experiencing significant overhead and delay [12].

Switching costs, usually described as platform lock-in, can also affect the intention of SNS users to deploy PETs, if this requires switching to a new SNS platform [53]. For instance, to use Scrambls, all recipients need to use the required platform plugin [55] to decrypt a message. Ajami and Ramadan [43] argue that if an SNS provider identifies the use of Facecloak, a PET that replaces selected information with other meaningful values when these are posted to the SNS, then they may suspend the user account. This also adds to the switching costs for the use of PETs, since a privacy sensitive user needs to switch to another social network platform in order to apply privacy preserving tools.

Generally, PETs are technologies that make processing of personal data more costly or may prevent it altogether. Only a subset of PETs can claim to be 'positive-sum' in the sense that they allow the delivery of services as well as or better than would be the case without them. [4]

### 3.5 Low Visibility of Effectiveness and Inadequate Feedback

Users' awareness of the benefits derived from preserving their privacy is also a critical factor with regard to their decision to use privacy enhancing applications. There are users who report that they do not believe in the effectiveness of PETs [56]. This can be attributed to the way PETs communicate, or rather fail to do so, their results and to the way they give feedback for actions they have performed to protect the user [57] or to the way privacy related dangers are presented by the technology used [58].

For instance, Disconnect, a block-tracking tool that filters traffic to third-party sites to prevent tracking, does not provide any feedback on the privacy risks deriving from the third-party websites that are blocked [37]. The same problem exists with the use of encryption enabling PETs that do not inform users who or what were prevented from accessing their personal information.

### 3.6 Privacy Requirements are Partially Addressed

Most privacy enhancing technologies meet specific, only, privacy requirements. While privacy protection generally entails protecting PII from unauthorized information

collection, processing and dissemination, informing users and providing them with control over their personal data [59], [60], each privacy tool typically meets only a small fraction of these requirements.

For instance, both FlybyNight [12] and NOYB [15] use encryption and obfuscation in order to conceal user's information from the SNS platform and unauthorized users, but fail to protect the future inappropriate use of this information by users that may be authorized to access it [43]. At the same time, other types of PETs, such as privacyfix [26] and MyPermissions Cleaner [25], aim at raising user's privacy awareness, by visualizing the entities that may access their information or highlight issues deriving from the privacy policy, but offer no actual data shield, unless the user actively changes her privacy settings.

### 3.7 The Role of the SNS Platform

Some PETs, such as P3P [22], need to be supported by the SNS provider in order for users to employ them. However, providers are not always happy to support PETs if they are not obliged to, as there is no evidence that they will gain competitive advantage by establishing the use PETs [61] and at the same time they need to abandon personal information collection and pay the cost of acquiring a technology, as well as changing their technical infrastructure [62].

A typical example is Facebook's complex access control mechanisms, offered in Privacy Settings. While privacy breaches due to this type of access control have reached spotlight and Google+, a competing SNS provider is built on the idea of personal circles [63], Facebook has not redesigned social networks organization on the principles of audience segregation to support PETs such as Clique Prototype [7]. Finally, the application of basic access controls in some SNS was a late response to privacy advocate requests and not an initiative of SNS providers to protect personal information [64].

### 3.8 Responsibility Misconceptions

When it comes to privacy protection, many users have the belief that providers and government are applying necessary measures to ensure it, and are not aware that privacy protection is partly their responsibility as well. In fact, a Location Based Services Privacy survey, conducted in 2009, showed that PETs were perceived to be a relatively weaker mechanism for enhancing control and reducing privacy risk because they shift the responsibility of privacy protection on the individual users [48]. What is more, most existing PETs for SNS are based on the user's choice to use, such as browser add-ons that encrypt posted messages or highlight potential privacy issues, deriving from default privacy settings. Studies have shown that belief of low effectiveness of privacy regulation or company privacy policies is an incentive for protection technology adoption by the user [65], so low adoption of PETs appears as a result of this belief.

Complex privacy policies published in most SNS contribute to this finding because many users misinterpret their presence as enabled privacy protection, while if the

presentation was simple and direct, they could understand the privacy issues and would be willing to pay for PETs [41][66]. On the other hand, SNS compliance to privacy regulations is difficult to audit, due to lack to accountability mechanisms. For example, there is the discussion of whether self-regulation, co-regulation or direct regulation should be used to enforce respect to users' stating their preference by employing the Do Not Track (DNT) mechanism [3].

### 3.9 Culture

Privacy concerns and privacy behavior are culture dependent. It has been found, for instance, that in Eastern culture, excessive self-disclosure is considered inappropriate, so privacy concerns are increased [67]. In 2009, a study by Hichang Cho et al. found that internet users' privacy concerns and behavioral responses such as opt-out and avoidance, varied significantly across nationalities, and they can be partially explained by national culture values [68]. However, multinational studies do not focus on how effectiveness of individual privacy protection mechanisms and strategies, including PETs, is perceived by individuals of different cultural background [48][68].

## 4 Conclusions and Further Research

The protection of PII in SNS is a complex issue involving several stakeholders, such as the users, PETs industry and developers, SNS providers, governments and regulatory bodies and third parties (e.g. advertisers). It thus calls for combined solutions, in which economic forces, cryptographic technologies, and targeted regulatory guidelines conspire to create a system with adequate enforcement and control powers (see also OECD (1997)) [41].

This paper presents an in-depth analysis of the key factors contributing to the limited adoption of privacy supporting technologies among SNS users. To the best of our knowledge, this is the first attempt to provide a unified view of the problem. Extant literature provides partial explanations derived from specific viewpoints: e.g. some researchers draw on social theory and employ diffusion of innovation models, others employ behavioral theories and technology acceptance models [41] or even economic theories [41]. This paper presents a critical discussion of all factors that have been identified and provides an integrated approach to the problem.

Our analysis has showed that the importance of awareness is rather overestimated, since many users are aware of different PETs but still refrain from their use. Cost, both direct and indirect, also contributes to low PETs adoption, but it is also the issues of the diversity and multiplicity of tools and applications that needs to be considered. Moreover, complexity and usability issues are also important determinants of PETs deployment, while the fact that users tend to underestimate their effectiveness due to low visibility of their results, seems to be ignored by vendors and developers.

It is also important to note that PETs currently offer very specific and limited functions with regard to privacy requirements in the context of SNS and that researchers and providers need to provide more integrated privacy solutions. Finally, the role of

culture seems to play an important role with regard to users' inclination against the use of PETs, and should be further explored.

Our effort to identify and evaluate the adoption of specific PETs by SNS users, was limited by the complete lack of relevant statistics and studies on the actual use of privacy tools by SNS users. Future research includes measuring the importance of each of the factors we have identified through a qualitative analysis, using actual user data.

## References


1. Acquisti, A., Gross, R.: Imagined communities: awareness, information sharing, and privacy on Facebook. In: Danezis, G., Golle, P. (eds.) PET 2006. LNCS, vol. 4258, pp. 36–58. Springer, Heidelberg (2006)
2. Boyd, D., Hargittai, E.: Facebook privacy settings: Who cares? First Monday 15(8) (2010)
3. ENISA: Privacy considerations of online behavioural tracking, report (2012)
4. London Economics: Study on the economic benefits of privacy-enhancing technologies (PETs). Final Report to The European Commission DG Justice, Freedom and Security (2010)
5. Compañó, R., Lusoli, W.: The Policy Maker's Anguish: Regulating Personal Data Behavior Between Paradoxes and Dilemmas. In: Moore, T., et al. (eds.) Economics of Information Security and Privacy, pp. 169–185. Springer Science+Business Media, LLC (2010)
6. Pötzsch, S.: Privacy Awareness: A Means to Solve the Privacy Paradox? In: Matyáš, V., Fischer-Hübner, S., Cvrček, D., Švenda, P. (eds.) The Future of Identity. IFIP AICT, vol. 298, pp. 226–236. Springer, Heidelberg (2009)
7. Van den Berg, B., Leenes, R.E.: Audience Segregation in Social Network Sites. In: Proceedings for SocialCom 2010/PASSAT 2010 (Second IEEE International Conference on Social Computing/Second IEEE International Conference on Privacy, Security, Risk and Trust), pp. 1111–1117. IEEE (2010)
8. Tootoonchian, Y.G.A., Saroiu, S., Wolman, A.: Lockr: Better privacy for social networks. In: Proceedings of the 5th ACM International Conference on emerging Networking EXperiments and Technologies (CoNEXT), pp. 169–180. ACM, New York (2009)
9. Baden, R., Bender, A., Spring, N., Bhattacharjee, B., Starin, D.: Persona: An online social network with user-defined privacy. In: Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication, pp. 135–146. ACM, New York (2009)
10. Jahid, S., Mittal, P., Borisov, N.: EASiER: Encryption-based Access Control in Social Networks with Efficient Revocation. In: Proceedings of 6th ACM Symposium on Information, Computer and Communications Security, pp. 411–415. ACM, New York (2011)
11. Paulik, T., Földes, Á.M., Gulyás, G.: BlogCrypt: Private content publishing on the Web. In: Proceedings of the Fourth International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2010, pp. 123–128. IEEE (2010)
12. Lucas, M.M., Borisov, N.: Flybynight: mitigating the privacy risks of social networking. In: Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society, pp. 1–8. ACM, New York (2008)
13. Luo, W., Xie, Q., Hengartner: Facecloak: An architecture for user privacy on social networking sites. In: 2009 International Conference on Computational Science and Engineering, pp. 26–33. IEEE (2009)

14. Conti, M., Hasani, A., Crispo, B.: Virtual private social networks. In: Proceedings of the First ACM Conference on Data and Application Security and Privacy, pp. 39–50. ACM, New York (2011)
15. Guha, S., Tang, K., Francis, P.: NOYB: privacy in online social networks. In: Proceedings of the First Workshop on Online Social Networks, pp. 49–54. ACM, New York (2011)
16. Diaspora*, https://joindiaspora.com/
17. Shakimov, A., Lim, H., Caceres, R., Cox, L.P., Li, K., Liu, D., Varshavsky A.: Vis-à-Vis: Privacy-preserving online social networking via Virtual Individual Servers. In: Third International Conference on Communication Systems and Networks, COMSNETS 2011, pp. 1–10. IEEE (2011)
18. Frenzy – The Dropbox powered social network, http://frenzyapp.com/
19. Cutillo, L.A., Molva, R., Strufe, T.: Safebook: a privacy preserving online social network leveraging on real-life trust. IEEE Communications Magazine 47(12), 94–101 (2009)
20. Buchegger, S., Schiöberg, D., Vu, L.H., Datta, A.: PeerSoN: P2P social networking: early experiences and insights. In: Proceedings of the Second ACM EuroSys Workshop on Social Network Systems, pp. 46–52. ACM, New York (2009)
21. Kang, T., Kagal, L.: Enabling Privacy-awareness in Social Networks. In: Intelligent Information Privacy Management Symposium at the AAAI Spring Symposium 2010 (2010)
22. Cranor, L.: P3P: Making privacy policies more useful. IEEE Security and Privacy 1(6), 50–55 (2003)
23. Toch, E., Sadeh, N.M., Hong, J.: Generating default privacy policies for online social networks. In: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 4243–4248. ACM, New York (2009)
24. Squicciarini, A., Paci, F., Sundareswaran, S.: PriMa: an effective privacy protection mechanism for social networks. In: Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, pp. 320–323. ACM, New York (2010)
25. My permissions.org – Scan your permissions. Find out who gained access to your personal info, http://mypermissions.org/
26. Privacyfix – Lock down your privacy, https://privacyfix.com/start
27. Priveazy – The 'eazy' way to protect your privacy and stay safe online, https://www.priveazy.com/
28. Data use Policy | Facebook, Interactive Tools, https://www.facebook.com/about/privacy/tools
29. Privacy Mirror on Facebook, http://apps.facebook.com/privacy_mirror/
30. Privacy Check, http://www.rabidgremlin.com/fbprivacy/
31. Becker, J., Chen, H.: Measuring Privacy Risk in Online Social Networks (2009), https://www.cs.pitt.edu/~chang/265/proj10/zim/measureprivacyrisk.pdf
32. Make myself clear, protecting you from you since (2012), http://makemyselfclear.com/
33. Heer, J., Boyd, D.: Vizster: Visualizing Online Social Networks. In: Proc. IEEE Symp. Information Visualization, pp. 32–39 (2005)
34. Friend Wheel, https://friend-wheel.com/
35. Monjas, M.A., Del Alamo, J.M., Yelmo, J.C., Hogberg, J.: Privacy Delegate: a browser-based tool for privacy self-management in social networks, Ericsson Position paper: W3C Workshop on identity in the browser (2010)

36. Wishart, R., Corapi, D., Madhavapeddy, A., Sloman, M.: Privacy Butler: A Personal Privacy Rights Manager for Online Presence. In: Proceedings of the 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 672–677. IEEE (2010)
37. In private browsing & search | Stop online tracking & malware | Disconnect, https://disconnect.me
38. Strater, K., Lipford, H.: Strategies and struggles with privacy in an online social networking community. In: Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, vol. 1, pp. 111–119. British Computer Society, Swinton (2008)
39. Madejskiy, M., Johnson, M., Bellovin, S.M.: The Failure of Online Social Network Privacy Settings. In: CUCS-010-11 (2011), http://academiccommons.columbia.edu/catalog/ac:135406
40. Hallinana, D., Friedewalda, M., McCarthyb, P.: Citizens' perceptions of data protection and privacy in Europe. Computer Law & Security Review 28, 263–272 (2012)
41. Acquisti, A.: The Economics of Personal Data and the Economics of Privacy, DRAFT (November 24, 2010), http://www.heinz.cmu.edu/~acquisti/papers/acquisti-privacy-oecd-22-11-10.pdf
42. Acquisti, A., Grossklags, J.: Privacy Attitudes and Privacy Behavior: Losses, Gains, and Hyperbolic Discounting (2004), Preliminary draft http://www.heinz.cmu.edu/~acquisti/papers/acquisti_grossklags_eis_refs.pdf; Final version in Camp, J., Lewis, R. (eds.): The Economics of Information Security. Kluwer (2004)
43. Ajami, R., Ramadan, N., Mohamed, N., Al-Jaroodi, J.: Security Challenges and Approaches in Online Social Networks: A Survey. IJCSNS International Journal of Computer Science and Network Security 11(8) (August 2011)
44. Wang, N., Grossklags, J., Xu, H.: An Online Experiment of Privacy Authorization Dialogues for Social Applications. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW 2013, pp. 261–272. ACM, New York (2013)
45. Lewis, K., Kaufman, J., Christakis, N.: The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. Journal of Computer-Mediated Communication 14, 79–100 (2008)
46. Flash Eurobarometer 225: Data Protection in EU: Citizens' Perception. European Commission (2008)
47. Oomen, I., Leenes, R.: Privacy Risk Perceptions and Privacy Protection Strategies. In: de Leeuw, E., Fischer-Hübner, S., Tseng, J., Borking, J. (eds.) IFIP International Federation for Information Processing, Policies and Research in Identity Management, vol. 261, pp. 121–138. Springer, Boston (2008)
48. Xu, H., Gupta, S., Rosson, M.B., Carroll, J.M.: Effectiveness of Privacy Assurance Approaches in Location-Based Services: A Study of India and the United States. In: Proceedings of the Eighth International Conference on Mobile Business, pp. 278–283. IEEE (2009)
49. Yao, M.Z.: Self-Protection of Online Privacy: A Behavioral Approach. In: Trepte, S., Reinecke, L. (eds.) Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web. Springer, Heidelberg (2011)
50. Youn, S.: Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors Among Young Adolescents. Journal of Consumer Affairs 43, 389–418 (2009)

51. Besenyei, T., Földes, A.M., Gulyás, G.G., Imre, S.: StegoWeb: Towards the Ideal Private Web Content Publishing Tool. In: Proceedings of SECURWARE 2011, The Fifth International Conference on Emerging Security Information, Systems and Technologies, pp. 109–114. IARIA (2011)
52. Leon, P.G., Ur, B., Balebako, R., Cranor, L.F., Shay, R., Wang, Y.: Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems, pp. 589–598. ACM, New York (2012)
53. Edlin, S.A., Harris, R.G.: The Role of Switching Costs in Antitrust Analysis: A Comparison of Microsoft and Google. Forthcoming in Yale Journal of Law and Technology 15 (February 7, 2013)
54. Facebook Statistics | Statistic Brain, http://www.statisticbrain.com/facebook-statistics/
55. Scrambls, https://scrambls.com
56. Hallinana, D., Friedewalda, M., McCarthyb, P.: Citizens' perceptions of data protection and privacy in Europe. Computer Law & Security Review 28, 263–272 (2012)
57. Shin, D.H.: The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption. Interacting with Computers 22, 428–438 (2010)
58. Kobsa, A., Teltzrow, M.: Contextualized Communication of Privacy Practices and Personalization Benefits: Impacts on Users' Data Sharing and Purchase Behavior. In: Martin, D., Serjantov, A. (eds.) PET 2004. LNCS, vol. 3424, pp. 329–343. Springer, Heidelberg (2005)
59. Solove, D.J.: A taxonomy of privacy. University of Pennsylvania Law Review 154(3) (January 2006)
60. Schwaig, K.S., Kane, G.C., Storey, V.C.: Compliance to the fair information practices: How are the Fortune 500 handling online privacy disclosures? Information and Management 43(7), 805–820 (2006)
61. Fairchild, A., Ribbers, P.: Privacy-Enhancing Identity Management in Business. In: Camenisch, J., Leenes, R., Sommer, D. (eds.) Digital Privacy. LNCS, vol. 6545, pp. 107–129. Springer, Heidelberg (2011)
62. Feigenbaum, J., Freedman, M.J., Sander, T., Shostack, A.: Economic barriers to the deployment of existing privacy technologies (position paper). In: Proceedings of the Workshop on Economics of Information Security (2002)
63. McNulty, S.: The Google+ Guide: Circles, Photos, and Hangouts. Peachpit Press (2012)
64. Beato, F., Kohlweiss, M., Wouters, K.: Enforcing Access Control in Social Network Sites. In: Procedings of Hot Topics in Privacy Enhancing Technologies, HotPETS (2009)
65. Lwin, M., Wirtz, J., Williams, J.D.: Consumer online privacy concerns and responses: a power–responsibility equilibrium perspective. Journal of the Academy of Marketing Science 35, 572–585 (2007)
66. Bonneau, J., Preibusch, S.: The Privacy Jungle: On the Market for Data Protection in Social Networks. In: Economics of Information Security and Privacy, pp. 121–167. Springer, US (2010)
67. Zhao, H., He, M.: Study on Social Culturology of the 'Internet Sharer'. In: Proceedings of the 1st IEEE Symposium on Web Society, pp. 219–224. IEEE (2009)
68. Cho, H., Rivera-Sánchez, M., Lim, S.S.: A multinational study on online privacy: global concerns and local responses. New Media & Society 11(3), 395–416 (2009)

# Towards Privacy-by-Design Peer-to-Peer Cloud Computing

Leucio Antonio Cutillo and Antonio Lioy

Politecnico di Torino
Dip. Automatica e Informatica
Corso Duca degli Abruzzi, 24, 10129 Torino, Italy
{leucio-antonio.cutillo,lioy}@polito.it

**Abstract.** Current Cloud services raise serious security and privacy concerns due to the potential misuse of user data by the omniscient Cloud service provider. Solutions proposing the "Cloud-of-clouds" paradigm just mitigate service availability threats, and additional encryption operations do not prevent users from being identified and traced. Moreover, these solutions still fail to address a main orthogonal problem, i.e. the intrinsic contrast between the provider's business model and the user's privacy. In this paper, we propose a new architecture for Cloud computing which addresses the protection of the user's privacy from the outset. Cloud services are provided by a number of cooperating independent parties consisting in the user nodes themselves. Unlike current Cloud services, the proposed solution provides user anonymity and untraceability. Such architecture can still take part in the "Cloud-of-clouds", allowing users to select service providers on the basis of the expected privacy protection.

**Keywords:** Peer-to-Peer, Cloud, privacy-by-design.

## 1 Introduction

The *Everything as a Service* paradigm proposed by Cloud computing is changing *de facto* the way Internet users, such as individuals, institutions and companies, deal with data storage and computation. Globally deployed cost-efficient and scalable resources are made available on demand, allowing users to access them via lightweight devices and reliable Internet connection. In recent reports, Comscore pointed out that 9.4 million new smartphones were acquired in EU5[1] in December 2012, and 136 million people now have a smartphone in this area [4]. Moreover, 92% of the world's data has been created in just the last two years, and right now popular Cloud platforms such as YouTube store 72 hours of new videos every minute [5].

Evidence shows that the benefits from apparently unlimited resources come at extremely high security and privacy costs [13]. User identity, location and activity information is constantly uploaded and synchronized at Cloud providers'

[1] UK, Germany, France, Italy and Spain.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 85–96, 2013.

facilities through mobile services, online social networks, search engines and collaborative productivity tools. Such data can be misused both by the provider itself and by attackers taking control of it. Additionally, due to the huge user base, Denial of Service (DoS) attacks reveal to be more effective.

While dependability can be addressed by running several service instances on different Clouds at increased costs, therefore moving to the so-called ***Cloud-of-clouds***, security and privacy vulnerabilities still remain open issues and have a severe impact on user trust in the Cloud [1].

Adoption of an appropriate encryption mechanism may appear a viable solution to protect user privacy. Unfortunately, securing outsourced data and computation against untrusted Clouds through encryption is cost-unfeasible [3], being outsourcing mechanisms up to several orders of magnitude costlier than their non-outsourced, locally run, alternatives. Moreover, the simple use of encryption to provide data confidentiality and integrity fails to hide sensitive information such as user identity and location, session time and communication traces.

However, even at the presence of fine-grained, cost-effective security and privacy protection tools based on encryption, current Cloud solutions would still suffer from a main orthogonal problem: the intrinsic contrast between their business model and user privacy. As a matter of fact, all current Cloud services are run by companies with a direct interest in increasing their user-base and user demand; service level agreements are often stringent to the user, and countermeasures against privacy violations are usually taken a-posteriori, once the violation has been detected. Given the promising public Cloud service market size, which is estimated to grow from $129.9 billion in 2013 to $206.6 billion in 2016 [7], Cloud providers are not likely to address this problem in the near future.

In this work, we assume the protection of user privacy against the omniscient Cloud provider to be the main objective for Clouds, and we present a sketch for our novel approach to Cloud-of-clouds services that helps to better protect the security of users while allowing for the full scale of operations they are used to from existing Clouds.

The main contributions of our work are two: (i) to facilitate confidentiality and privacy by avoiding potential control from any central omniscient entity such as the Cloud provider through a distributed architecture for Cloud-of-clouds, where each Cloud user is a Cloud service provider too; (ii) to leverage on the real life trust relationships among Cloud users to lower the necessity for cooperation enforcement with respect to Cloud service availability. As an additional objective, the protection of the user's privacy against malicious users is also addressed.

The proposed architecture aims at preserving the user's privacy from the outset, and targets ***privacy-by-design***.

This paper is organized as follows: section 2 introduces the main security objectives we expect to meet with our novel approach, which is presented in section 3 and detailed in section 4; section 5 provides a preliminary evaluation

of the approach against such objectives, while section 6 presents the related work. Finally, section 7 concludes this paper and outlines future work.

## 2 Security Objectives

We assume the protection of the user's privacy against the omniscient Cloud service provider to be the main objective for Cloud services.

**Privacy.** Privacy is a relatively new concept, born and evolving together with the capability of new technologies to share information. Conceived as *"the right to be left alone"* [14] during the period of newspapers and photographs growth, privacy now refers to the ability of an individual to control and selectively disclose information about him.

The problem of users' data privacy can be defined as the problem of *usage control* [11], which ensures access control together with additional control on the later usage of the data, even once information has already been accessed. Access to the content of user-generated data should only be granted by the user directly, and this access control has to be as fine-grained as specified by the user.

In addition, communication privacy calls for inference techniques aiming at deriving any type of information with regard to: (1) *anonymity*, meaning that users should access resources or services without disclosing their own identities; (2) *unobservability*, i.e. the requirement that no third party should gather any information about the communicating parties and the content of their communication; (3) *unlinkability*, which requires that obtaining two messages, no third party should be able to determine whether both messages were sent by the same sender, or to the same receiver; (4) *untraceability*, which demands that no third party can build a history of actions performed by arbitrary users within the system; in other words, it demands both anonymity and unlinkability.

In summary, the objective of privacy is to hide any information about any user at any time, even to the extent of hiding their participation and activities within the Cloud service in the first place. Moreover, privacy has to be met by default, i.e. all information on all users and their actions has to be hidden from any other party internal or external to the system, unless explicitly disclosed by the users themselves.

**Integrity.** In Cloud services, any unauthorized modification or tampering of user-generated information has to be prevented. This encompasses the protection of real identity of users within the Cloud platforms. In this sense, the definition of integrity is extended in comparison with the conventional detection of modification attempts on data. Moreover, problems with integrity of user profiles and their contents may have devastating impact on the objectives put forth with respect to the privacy of Cloud users. Since the creation of profiles in popular Cloud services is easy, protection of real identities is insufficient in today's platforms. In particular, providers offering Cloud services for free are often unable (and perhaps even not interested in) to ensure that a profile is associated to the corresponding individual from the real world.

**Availability.** The objective of availability for Clouds aims at assuring the robustness of the services in the face of attacks and faults. Due to their exposure as single points of failure, centralized Cloud services are exposed to *denial-of-service* attacks, which directly impact the availability of user's data.

Also distributed services, which are implemented in a decentralized way, possibly via peer-to-peer systems, or which follow other types of service delegation, may be vulnerable to a series of attacks against availability as well. These attacks include *black holes*, aiming at collecting and discarding a huge amount of messages; *selective forwarding*, where some traffic is forwarded to the destination, but the majority is discarded; and *misrouting*, which aims to increase the latency of the system or to collect statistics on the network behavior. In any case, attacks on distributed Cloud systems are more effective in case of *collusion* amongst malicious users or in the presence of Sybil nodes controlled by the attacker, which is not the case for the centralized Cloud providers.

## 3 A New Approach

Our system provides Cloud services based on a peer-to-peer architecture. The peer-to-peer architecture meets the privacy concerns by avoiding potential control and misuse of user's data from the omniscient Cloud service provider or attackers taking control of it. Furthermore, cooperation among peers is enforced by leveraging on the real life trust relationships among the user themselves. Each participant is associated to a ***User Identifier*** (UId) and joins the network from multiple devices associated to different ***Node Identifiers*** (NIds). Resources of the participant's devices are available to the participant himself, and to the participant's trusted contacts and contacts-of-contacts with the participant's consent.

### 3.1 System Overview

Our system consists of three main components (Fig. 1): a ***Web of Trust*** (WoT), a ***Distributed Hash Table*** (DHT), and a series of ***Trusted Identification Services*** (TISs).

The WoT provides the basic distributed structure used to supply Cloud services, the DHT provides a basic dictionary service to perform lookups, finally each TIS serves the purpose of user authentication.

**Web of Trust.** The WoT (Fig. 2) is a digital mapping of the trust relationships users entertain in their real life, and serves the purpose of Cloud service provisioning. In a user's WoT view, each user's trusted contact acts as a ***Trusted Cloud Service Provider*** (TCSP), and provides the user with storage and computing resouces. Such resources can be allocated both on the TCSP hardware and in that one of its respective TCSPs ones. However, since we don't assume transitivity of trust, a TCSP of a user's TCSP is considered an ***Untrusted Cloud Service Provider*** (UCSP). To preserve both the consumer's privacy and the trust edges in the WoT, UCSP resources are transparently accessed through

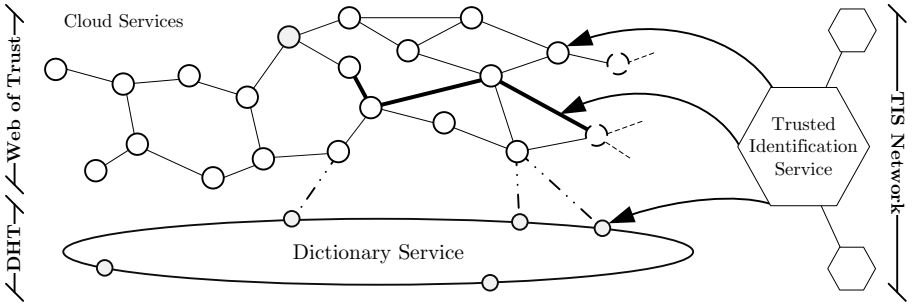



**Fig. 1.** Main components of the system: Distributed Hash Table, Web of Trust, and Trusted Identification Service network

TCSP only. Finally, in case a TCSP is offline, the set of UCSP resources accessible through that TCSP is still reachable through a set of ***Auxiliary Access Points*** (AAPs), which lead to the TCSP contacts through ***random walks*** on the WoT graph.

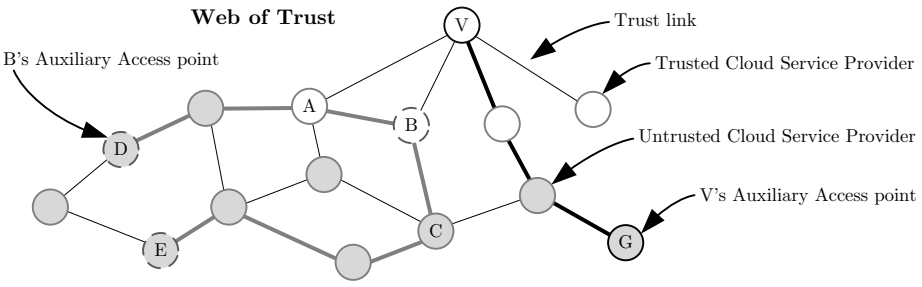



**Fig. 2.** The Web of Trust Component: in white, Trusted Cloud Service Providers (trusted contacts) for $\mathcal{V}$; in light gray, Untrusted Cloud Service Providers for $\mathcal{V}$. Node $\mathcal{B}$ is offline, part of the services $\mathcal{B}$ provides to $\mathcal{V}$ are still accessible from $\mathcal{B}$'s Auxiliary Access Points $\mathcal{D}$ and $\mathcal{E}$. Random walks in light gray forward $\mathcal{V}$'s request for $\mathcal{B}$'s services to $\mathcal{B}$'s direct contacts $\mathcal{A}$ and $\mathcal{C}$ without revealing the real requester $\mathcal{V}$'s identity.

**DHT.** The DHT is implemented by an overlay on top of the internet where peers are arranged thanks to their NId and serves the purpose of TCSP lookup. The DHT is maintained by the users' nodes and provides three distinct lookup services: it returns IP addresses associated to a target NId; it returns a list of AAP for a given UId; it returns the UId associated to a target TCSP identity, such as the user's full name.

Therefore, the DHT allows for building the WoT overlay and addressing the TCSP services.

**TIS Network.** TISs are independent trusted parties serving the purpose of user and device authentication. A TIS provides each user with a certified UId and a set of certified NIds, one for each user device. Any TIS computes identifiers starting from the user's real identity by running the same algorithm.

TISs are offline entities contacted at the first join and do not play any role neither in the communication between users nor in the Cloud service provisioning. Consequently, they do not break the main purpose of decentralization.

### 3.2 Orchestration

A newcomer generates a series of public-private key pairs, contacts a TIS and obtains as an answer his User- and Node- Identifiers, together with certificates associating each identifier with a public key. The newcomer device joins the DHT thanks to the Node Identifier, and the newcomer starts looking for trusted contacts from a series of properties such as the contact name. As an answer, the newcomer receives a set of AAPs for different User Identifiers, and starts retrieving publicly available profile data associated to each identifier through the respective AAPs. Once identified the correct trusted contact, the newcomer sends a contact request to the AAPs which is forwarded along a random walk on the WoT graph. Replies are forwarded back along the same path. A new trusted link is therefore established in the WoT graph. Contact requests contain available devices Node Identifiers and a series of secrets to access their running services. The newcomer queries the P2P system for the IP addresses of each device, and establishes one-hop connections with them. The newcomer then sends to the trusted contact a random walk request, which will create a random walk ending to an AAP for the newcomer. By repeating the abovementioned process, the newcomer adds further real-life trusted contacts and creates a random walk for each of them. The newcomer's trusted contacts act as TCSPs and provide services encompassing Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS).

## 4 Operations

In the following, we sketch the operations implementing the distributed Cloud-of-Clouds, which consist of: (i) account creation, (ii) trusted contact establishment, and (iii) Cloud service access.

Each operation calls for the execution of a series of secure protocols aiming at obtaining credentials, building and keeping the consistency of the WoT and DHT overlays and establishing secure communication channels.

Throughout the description of these protocols, $\{M\}_{S_{\mathcal{X}_\mathcal{X}}}$ denotes a message $M$ being signed by user $\mathcal{X}$'s private key $\mathcal{K}_\mathcal{X}^-$, and $E_{\mathcal{K}_\mathcal{Y}}\{M\}$ denotes the message $M$ being encrypted with the user $\mathcal{Y}$'s public key $\mathcal{K}_\mathcal{Y}^+$[2]. The distinct identifiers of

[2] More precisely, session keys are used to encrypt the payload. Such keys are advertised at the beginning of the message encrypted with the target Node Id public key.

users are associated with keypairs: while $\mathcal{N}_\mathcal{X} = \{\mathcal{N}_\mathcal{X}^-, \mathcal{N}_\mathcal{X}^+\}$ denotes the keypair for the Node Id, $\mathcal{U}_\mathcal{X} = \{\mathcal{U}_\mathcal{X}^-, \mathcal{U}_\mathcal{X}^+\}$ denotes the keypair for the User Id, and $\mathcal{W}_\mathcal{X} = \{\mathcal{W}_\mathcal{X}^-, \mathcal{W}_\mathcal{W}^+\}$ denotes the keypair for the random walk Id of node $\mathcal{X}$.

### 4.1 Account Creation

In order to create an account, a newcomer $\mathcal{V}$ generates a series of keypairs $\mathcal{U}_\mathcal{V}$, $\{\mathcal{N}_{\mathcal{V},0\leq i\leq n}\}$, $\{\mathcal{W}_{\mathcal{V},0\leq j\leq m}\}$ to be associated, respectively, to his personal identity, his $n$ devices and $m$ random walks. All public keys are sent to the TIS, together with $\mathcal{V}$'s identity record $name_\mathcal{V} =< firstName, \ldots, nationality >$ and a proof of identity ownership[3].

Starting from a series of secrets $MK_\mathcal{U}$, $MK_\mathcal{N}$, $MK_\mathcal{W}$, the TIS computes $\mathcal{V}$'s User Identifier $UId_\mathcal{V}$ as a keyed hash function $h_{MK}$ applied to $name_\mathcal{V}$. Similarly, the TIS also computes a set of $n$ node and $m$ random walk identifiers by using $MK_\mathcal{N}$, $MK_\mathcal{W}$ respectively, and applying the keyed hash function on a concatenation of $name_\mathcal{V}$ and an integer $0 \leq p \leq n$.

Finally, each identifier is sent back to $\mathcal{V}$ together with a certificate associating such identifier to a different user-generated public key. The certificate also contains further information on the TIS, a timestamp and an expiration time. Additional meta identifiers are sent back to $\mathcal{V}$. Each identifier is computed as a well known hash function of a possible combination of the values in $name_\mathcal{V}$ such as $h\,(firstName, nationality)$.

Once received the certified identifiers, $\mathcal{V}$ can join the P2P network.

**Trusted Contact Establishment.** The newcomer $\mathcal{V}$ needs to build his web of trust to access (and provide) Cloud services from his node pool. To find a trusted contact $\mathcal{U}$, $\mathcal{V}$ hashes a subset of properties of $name_\mathcal{U}$ and looks for this meta identifier on the DHT. As an answer, all User Identifiers $UId_i$ associated to such meta identifier are provided to $\mathcal{V}$, which triggers another request for each of them. A list of random walk identifiers $WId_{ij}$ and corresponding auxiliary access points is retrieved for each $UId_i$. User $\mathcal{V}$ then triggers a profile request to the AAPs. Requests are routed along the random walks thanks to $WId_{ij}$; publicly available profile data is served by each $UId_i$ (or one of his trusted contacts) and is forwarded back along the same path. At the reception of profile data, $\mathcal{V}$ selects the correct target $UId_\mathcal{U}$ and sends him a contact request containing $\mathcal{V}$'s User- and Nodes- Identifiers together with the TIS certificates and a list of available Cloud services running at $\mathcal{V}$'s node pool. Again, the contact request is sent to $\mathcal{U}$'s AAPs and is routed along the random walks. User $\mathcal{U}$ can accept the request and reply directly to $\mathcal{V}$.

Once a bidirectional trust link has been built, $\mathcal{V}$ can access Cloud services offered by $\mathcal{U}$, and vice-versa.

**Cloud Service Access.** The first Cloud service $\mathcal{V}$ accesses is the ***Communication Obfuscation as a Service*** (COaaS), where $\mathcal{V}$ creates a random walk of $q$ hops starting from $\mathcal{U}$. A random walk request RWR message is sent to $\mathcal{U}$

[3] Such proof can consist of a secret shared OOB after face-to-face identity verification.

and forwarded along the WoT graph. Such RWR contains a walk token $WTok$, a recursively signed Time To Live message and a signed random number $rnd_{S_{\mathcal{W}^-}}$. The $WTok$ contains the jth random walk Id certificate $Cert\left(WId_{\mathcal{V},j}\right)$ and an expiration time signed with $\mathcal{W}^-$. At each hop, a user in the random walk decreases the $TTL$ and selects a random contact to forward the request. When $TTL = 0$, the current node $\mathcal{G}$ verifies the signature on the random number is associated to $Cert\left(WId_{\mathcal{V},j}\right)$, and registers the pair $\langle DHTkey, DHTvalue\rangle$ on the DHT, where $DHTkey = WId_{\mathcal{V},j}$ and $DHTvalue = [WTok, Cert\left(NId_{\mathcal{G}}\right)]\, S_{\mathcal{N}_{\mathcal{G}}^-}$. The presence of $rnd_{S_{\mathcal{W}^-}}$ in the DHT storage request for $\langle DHTkey, DHTvalue\rangle$ triggered by $\mathcal{G}$ poses as an authorization.

Once such association has been registered, a confirmation is routed back according to $WId_{\mathcal{V},j}$ along the random walk. At the same time, $\mathcal{V}$ stores a new association $\left\langle UId_{\mathcal{V}}, [Cert\left(UId_{\mathcal{V}}\right), Cert\left(WId_{\mathcal{V},j}\right), exptime]\, S_{\mathcal{U}_{\mathcal{V}}^-}\right\rangle$ in the DHT.

Storage of $\left\langle metaId_{\mathcal{V}}, [Cert\left(metaId_{\mathcal{V}}\right), Cert\left(UId_{\mathcal{V}}\right), exptime]\, S_{\mathcal{U}_{\mathcal{V}}^-}\right\rangle$ is optional and may happen at any time.

A series of IaaS, SaaS, PaaS services can be provided, with the user consent, to the user's contacts in addition to the user himself. User $\mathcal{U}$ has $n$ real/virtual nodes in the DHT which form $\mathcal{U}$'s node pool and provide basic services like COaaS and storage. Among such nodes, those with higher resources run an hypervisor and instantiate virtual machines, which may be connected to form more complex virtual data centers.

Within the MapReduce framework, trust-based parallel processing is achieved by splitting problems in sub-problems and distributing them to trusted user maintained nodes. Sub-problems may further be divided and dispatched along the WoT.

As shown in Fig. 3, among the series of services $\mathcal{U}$ provides to $\mathcal{V}$, part of them may not be run directly on $\mathcal{U}$'s nodes. $\mathcal{U}$ may in fact advertise services provided to him by his trusted contact $\mathcal{Z}$. In this case, $\mathcal{U}$ acts as a proxy for $\mathcal{V}$. When all $\mathcal{U}$'s nodes are disconnected, $\mathcal{V}$ can still access services running at $\mathcal{Z}$ nodes by contacting $\mathcal{U}$'s AAPs.

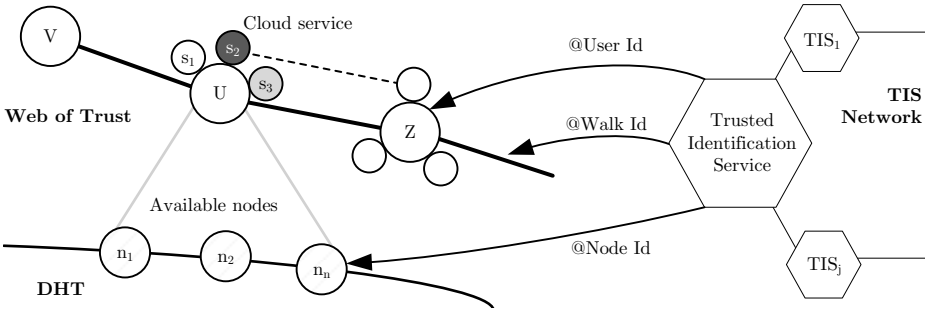



**Fig. 3.** Availability of Cloud services provided by user $\mathcal{U}$ to $\mathcal{V}$: in white, available services running at $\mathcal{U}$'s nodes; in black, unavailable ones; in gray, services running at $\mathcal{Z}$ nodes available for $\mathcal{U}$ which $\mathcal{U}$ respectively provides, as a proxy, to $\mathcal{V}$

## 5 Preliminary Evaluation

A complete performance evaluation of the proposed system is not available at this point, as we are describing work in progress. In the following we however give an overview to evaluate the compliance with respect to the security objectives we required in section 2.

**Integrity.** The one-to-one correspondance between a user's identity and his User Identifier is ensured by the TIS. This prevents malicious users from creating ***fake identities***, or mounting ***impersonation attacks*** and disrupt the WoT. Even when a user starts the account creation process with two different TISs, since the different $MK$ used for identifier computation are shared within the TIS network, such user will receive the same identifiers.

A newcomer $\mathcal{V}$ is prevented from sending a trusted contact request to AAPs which are not those associated to the legitimate target $\mathcal{U}$. The association between a certified meta identifier and a certified UId cannot be stored in the DHT in case of mismatch between the public keys contained in both certificates. The same applies to the association between a UId and a WId.

Finally, the association between a certified WId and a certified AAP Nid cannot be stored without the proof such AAP is at the end of a random walk originated at the legitimate $\mathcal{V}$. Such proof is represented by $rnd_{S_{w^-}}$.

Since the association between a walk identifier and the AAP Node Identifier is publicly available, ***collusion*** between DHT nodes and unlegitimate AAPs can be detected and malicious association removed from the DHT.

**Privacy.** The proposed solution guarantees anonymity, since a user can choose to skip the registration between any of his meta identifiers and his UId. Computation of the victim's UId cannot be performed with ***dictionary attacks*** due to the adoption of a keyed hash function to compute identifiers. Even ***brute force attacks*** to guess a victim's UId fail in case such victim does not provide publicly available information on his identity.

An anonymous user $\mathcal{V}$ may still find and establish a trusted contact with a non anonymous user $\mathcal{U}$. However, when both $\mathcal{V}$ and $\mathcal{U}$ are anonymous users, trusted contact establishment succeeds only after UIds have been exchanged out-of-band. Currently, a TIS can detect if a user is participating in the system or not. We are investigating further distributed authentication mechanisms to overcome this limitation.

Hop-by-hop communication integrity and confidentiality is ensured by signing each message with the sender's node private key and encrypting the concatenation of message and signature with the receiver's node public key. Eavesdroppers are therefore unable to detect any information on parties' identifiers and communication content. Additionaly, in case of trusted contact establishment or asynchronous communication, end-to-end communication integrity and confidentiality is ensured by signature and encryption operations based on keypairs associated to User Identifiers.

In case TCSPs provide storage space, a user $\mathcal{V}$'s data is stored at the TCSP $\mathcal{U}$ in an encrypted and partitioned form. In fact, each user is considered ***honest but curious***.

User identity cannot be linked to IP addresses since the relationship between Node Identifiers and User Identifier is not stored in the DHT and is disclosed to trusted contacts only at the act of contact establishment.

Nodes participating in serving random walks cannot derive any information on the originator of the walk: given a WId, it is impossible to retrieve any information on any other WId related to the same user, consequently, if present, any publicly available information on such user.

Finally, no entity in the system can derive the composition of the WoT, since a user's knowledge of the WoT is limited to his trusted contacts only.

**Availability.** Adoption of certified identifiers prevents sybil and denial of service attacks. Malicious nodes fail in perpetrating ***DHT poisoning*** due to the presence of signatures in each record they store. Furthermore, due to the parallelism of DHT lookup and storage operations, a malicious peer dropping requests does not impact provisioning of results.

## 6 Related Work

To the best of our knowledge, no work in literature proposes the adoption of decentralized architectures to preserve Cloud users' privacy. On the contrary, a series of work aims at moving towards P2P to increase Cloud service availability.

The work in [15] proposes a P2P Cloud solution to provide a distributed data storage environment to address the problem of bottlenecks due to central indexes in Cloud storage systems. In Cloud-to-Cloud [8], Cloud service providers participate in a pool of computer resources. Resource requests which cannot be provisioned by a particular Cloud service provider can be met from such a shared ecosystem of resources in a seamless manner. In [10], authors present the design of P2P MapReduce, an adaptive framework which exploits a P2P model to allow for MapReduce in Cloud-of-clouds environment. Compared to centralized implementations of MapReduce, such solution provides resiliency against node churn and failures.

Researchers also got inspired from P2P systems to propose new way of looking up for resources in Cloud environments. ***Cloud peer*** [12] creates an overlay network of virtual machines to address service discovery and load-balancing. VM instances update their software and hardware configuration to a DHT supporting indexing and matching of multidimensional range, so that provisioning software can search for and discover them. Authors in [2] propose a distributed IaaS Cloud system using gossip based protocols to manage a pool of peers without coordinators. Users can request a fraction of the available resources matching a given query. Authors in [9] propose an hybrid overlay composed by a structured P2P system with an unstructured one to support multi-attribute range query.

The overlay is organized in clusters, each cluster being composed by resource groups, in turn composed by peers with the same resources.

The work which is closest to our approach has been presented in Safebook [6], a decentralized architecture for privacy preserving online social networks. As in our solution, Safebook relies on the trust relationships among users to provide P2P services. As opposed to Safebook, our solution does not depend on a single TIS, allows multiple user's nodes to join the network at the same time, and provides higher privacy protection in terms of communication untraceability thanks to the adoption of random walk based routing. Moreover, while Safebook exploits direct contacts resources to provide a limited range of social network services, our solution exploits the entire WoT and is conceived to provide a wide range of Cloud services.

## 7 Conclusion and Future Work

In this paper, we proposed a decentralized approach to protect critically sensitive user data against the potential control of malicious omniscient Cloud service providers and attackers taking control of them. The approach leverages on existing trust among users to provide Cloud services and anonymize traffic. Our solution consists on a Web of Trust among users, who also play the role of Cloud service providers. User nodes take part in a DHT and run Cloud services, which may be accessed directly or indirectly, as a function of the WoT shape and with the user consent. Resilience against offline Cloud providers is achieved through Auxiliary Access Points for those providers' services. The sensitive information describing WoT users and their relationships is protected through encryption and anonymization techniques similar to onion routing, applied to random walks on the WoT graph.

As future work, we plan to develop a full model of the system in order to study the service availability. We have already started to develop a prototype based on the new WebRTC[4] technology, which allows to build a P2P network from popular web browsers. We plan to complete our prototype and integrate it with popular IaaS solutions such as OpenStack and SaaS such as Hadoop. Since trust between users does not correspond to trust on their hardware, we plan to investigate new mechanisms to guarantee the execution and correctness of a sequence of operations at the Cloud side. Finally, we also plan to evaluate the tradeoff between usage control and anonymity through new distributed collaborative privacy policy enforcement schemes.

**Acknowledgments.** This research has been partly funded by the European Commission within its seventh Framework Programme (FP7/2007-2013) under grant agreement n. 257243 (TClouds[5] project).

[4] `http://www.webrtc.org`
[5] `http://www.tclouds-project.eu`

## References


1. Cloud Security Alliance, ISACA, Cloud computing market maturity (September 2012), http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/2012-Cloud-Computing-Market-Maturity-Study-Results.aspx
2. Babaoglu, O., Marzolla, M., Tamburini, M.: Design and implementation of a P2P cloud system. In: SAC 2012: 27th Annual ACM Symposium on Applied Computing, Trento, Italy, March 26-30, pp. 412–417 (2012), doi:10.1145/2245276.2245357
3. Chen, Y., Sion, R.: On securing untrusted clouds with cryptography. In: WPES 2010: 9th Annual ACM Workshop on Privacy in the Electronic Society, Chicago, IL, USA, pp. 109–114 (October 4, 2010), doi:10.1145/1866919.1866935
4. Comscore, How technology and analytics drive the mobile market (February 2013), http://www.comscore.com/Insights/Presentations_and_Whitepapers/2013/How_Technology_and_Analytics_Drive_the_Mobile_Market/
5. Comscore, The rise of big data on the Internet (January 2013), http://www.comscore.com/Insights/Presentations_and_Whitepapers/2013/The_Rise_of_Big_Data_on_the_Internet/
6. Cutillo, L., Molva, R., Strufe, T.: Safebook: A privacy-preserving online social network leveraging on real-life trust. IEEE Communications Magazine 47(12), 94–101 (2009), doi:10.1109/MCOM.2009.5350374
7. Gartner, Forecast overview: Public cloud services, worldwide, 2011-2016, 2q12 update (August 2012), http://www.gartner.com/id=2126916
8. Gupta, A., Kapoor, L., Wattal, M.: C2C (Cloud-to-cloud): An ecosystem of cloud service providers for dynamic resource provisioning. In: Abraham, A., Lloret Mauri, J., Buford, J.F., Suzuki, J., Thampi, S.M. (eds.) ACC 2011, Part I. CCIS, vol. 190, pp. 501–510. Springer, Heidelberg (2011)
9. Lai, K., Yu, Y.: A scalable multi-attribute hybrid overlay for range queries on the cloud. Information Systems Frontiers 14(4), 895–908 (2012), doi:10.1007/s10796-011-9328-7
10. Marozzo, F., Talia, D., Trunfio, P.: P2P-MapReduce: Parallel data processing in dynamic cloud environments. Journal of Computer and System Sciences 78(5), 1382–1402 (2012), doi:10.1016/j.jcss.2011.12.021
11. Park, J., Sandhu, R.: Towards usage control models: beyond traditional access control. In: SACMAT 2002: 7th ACM Symposium on Access Control Models and Technologies, Monterey, CA, USA, June 3-4, pp. 57–64 (2002), doi:10.1145/507711.507722
12. Ranjan, R., Zhao, L., Wu, X., Liu, A., Quiroz, A., Parashar, M.: Peer-to-peer cloud provisioning: Service discovery and load-balancing. In: Antonopoulos, N., Gillam, L. (eds.) Cloud Computing, pp. 195–217. Springer, London (2010), doi:10.1007/978-1-84996-241-4_12
13. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications 34(1), 1–11 (2011), doi:10.1016/j.jnca.2010.07.006
14. Warren, S.D., Brandeis, L.D.: The right to privacy. Harward Law Review 4(5), 193–220 (1890)
15. Xu, K., Song, M., Zhang, X., Song, J.: A cloud computing platform based on P2P. In: ITIME 2009: IEEE International Symposium on IT in Medicine & Education, Jinan, Shandong, China, August 14-16, pp. 427–432 (2009), doi:10.1109/ITIME.2009.5236386

# Preservation of Utility through Hybrid $k$-Anonymization*

Mehmet Ercan Nergiz, Muhammed Zahit Gök, and Ufuk Özkanlı

Department of Computer Engineering, Zirve University, Gaziantep, Turkey

**Abstract.** Anonymization-based privacy protection ensures that published data cannot be linked back to an individual. The most common approach in this domain is to apply generalizations on the private data in order to maintain a privacy standard such as $k$-anonymity. While generalization-based techniques preserve truthfulness, relatively small output space of such techniques often results in unacceptable utility loss especially when privacy requirements are strict. In this paper, we introduce the *hybrid generalizations* which are formed by not only generalizations but also the *data relocation* mechanism. Data relocation involves changing certain data cells to further populate small groups of tuples that are indistinguishable with each other. This allows us to create anonymizations of finer granularity confirming to the underlying privacy standards. Data relocation serves as a tradeoff between utility and truthfulness and we provide an input parameter to control this tradeoff. Experiments on real data show that allowing a relatively small number of relocations increases utility with respect to heuristic metrics and query answering accuracy.

**Keywords:** Privacy, Anonymization, Privacy-preserving databases.

## 1 Introduction

The advance of technology along with the low cost of handling data have led service providers to collect personal information with the hope of turning this data into profit. In some cases, the potential value of such data is so great, it needs to be outsourced for analysis or it has to be published for research purposes as is the case with health related data in medical research. However, such data often contain sensitive information that needs to be kept private such as diagnosis and treatments. Thus sharing it raises every privacy concern [10]. In order to preserve the privacy of individuals, data needs to be properly anonymized before publishing meaning the link between sensitive information and individual identity should be removed. Such an anonymization must not only satisfy the

* This work was funded by The Scientific and Technological Research Council of Turkey (TUBITAK) Young Researchers Career Development Program under grant 111E047.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 97–111, 2013.

underlying privacy requirements but also preserve the utility of the data. Otherwise, it would be difficult to extract useful information from the anonymized data.

Unfortunately, just removing uniquely identifying information (e.g., SSN) from the released data is not enough to protect privacy. Works in [23] and [24] show that using publicly available sources of partially identifying information (*quasi-identifiers*) such as age, gender and zip-code, data records can be re-identified accurately even if there is no direct identifying information in the dataset. For example, in Table 1, suppose we release $T$ as a private table. Even if $T$ does not contain unique identifiers, an adversary that knows that her 41 years old friend Obi from USA with zip 49001 is in the dataset will be able to identify him as tuple q7.

To prevent identification, many different privacy metrics [23,24,18,16,27,21] have been introduced for various adversary models. As an example, *k-anonymity* requires that for each tuple $t$ in the anonymization, there should be at least $k-1$ other tuples indistinguishable with $t$. Two individuals are said to be indistinguishable if their records agree on the set of quasi-identifier attributes. To achieve the underlying privacy standard, many algorithms have been proposed. A common feature of these algorithms is that they manipulate the data by using *generalizations* which involves replacing data values with more general values (values that include the meaning of the original value and that may also imply other atomic values, e.g., 'Italy' is changed to 'Europe') so that more tuples will express similar meanings. As an example, suppose the desired privacy standard is 3-anonymity. In Table 1, $T^*_{\mu_1}$ is a 3-anonymous generalization of $T$. Note that generalizations applied to $T$ create two *equality groups* that contain similar tuples with respect to QI attributes. From the adversary's point of view, tuples with each equality group are indistinguishable from each other. If the data owner releases $T^*_{\mu_1}$ instead of $T$, Obi can at best be mapped to the white equality group of size 5 and to a set of salaries {18K, 35K, 14K, 25K, 29K}.

A nice feature of generalizations is that unlike perturbation techniques (that apply noise to data cells independently before publishing), generalizations preserve the truthfulness of data. However, generalizations result in information loss, thus over-generalization should be avoided as long as the privacy requirements are satisfied. To solve this problem, many heuristics have been designed, however relatively small output space of such techniques often results in huge utility loss especially when privacy requirements are strict [3]. Preservation of utility still stands as a major problem for generalization-based techniques. One of the main reasons for over-generalization is the existence of outliers in private datasets. As the neighborhood of the outliers is not heavily populated in the high dimensional domain, it becomes difficult for an anonymization algorithm to generate an equality group of sufficient size. For those algorithms that are vulnerable to outliers, a relatively large group can degrade the overall utility of the whole dataset [20].

To address the negative effects of outliers and over-generalization, in this paper, we propose the *hybrid generalization* technique which combines the

generalization technique with a *data relocation* mechanism in order to achieve more utilized anonymizations. Data relocation involves changing certain data cells (that act as outliers) to further populate small equality groups of tuples. Over relocation harms truthfulness and localized utility, thus over-relocation should be avoided as well. This can be achieved by bounding the number of relocations that the algorithm can apply, thus controlling the trade-off between truthfulness and utility. Even a small number of relocations can prevent over-generalization. As an example, in Table 1, Table $\widehat{T}$ is a relocation of Table $T$ in which less than 10% of the data cells are relocated (see tuple q4). Table $\widehat{T}^*_{\mu_1}$ shows a 3-anonymization of $\widehat{T}$ (which we will also name as a 10%-hybrid 3-anonymization of $T$). $\widehat{T}^*_{\mu_1}$ is more specific than $T^*_{\mu_1}$ [1] and possibly more utilized. Our contributions in this paper are as follows:

- We introduce the hybrid $k$-generalization concept that allows relocation of tuples between groups to increase the overall utility at the cost of truthfulness.
- We show how one can use hybrid generalizations to achieve $k$-anonymity.
- We present hybrid anonymization algorithms that address three classes of adversaries.
- We empirically compare the hybrid algorithms with previously proposed algorithms and show that hybrid generalizations create better utilized anonymizations.

## 2 Background and Related Work

Given a dataset (table) $T$, $T[c][r]$ refers to the value of column $c$, row $r$ of $T$. $T[c]$ refers to the projection of column $c$ on $T$ and $T[.][r]$ refers to selection of row $r$ on $T$. We write $|t \in T|$ for the cardinality of tuple $t \in T$ (the number of times $t$ occurs in $T$).

Although there are many ways to generalize a given value, we stick to generalizations according to domain generalization hierarchies (DGH) given in Figure 1(a).

**Definition 1 (i-Gen Function).** *For two data values $v^*$ and $v$ from some attribute $A$, we write $v^* = \Delta_i(v)$ if and only if $v^*$ is the $i$th (grand) parent of $v$ in the DGH for $A$. Similarly for tuples $t, t^*$; $t^* = \Delta_{i_1 \cdots i_n}(t)$ iff $t^*[c] = \Delta_{i_c} t[c]$ for all columns $c$. Function $\Delta(v)$ without a subscript returns all possible generalizations of a value $v$.*

E.g., given Figure 1(a), $\Delta_1$(`USA`)=`N.AM`, $\Delta_{0,2,3}$( `<12,USA,47906>`)=`<12,AM, 47***>`, $\Delta$(`USA`)={`USA`,`N. AM`,`AM`,`*`}

**Definition 2 ($\mu$-Generalization).** *A generalization mapping $\mu$ is any surjective function that maps tuples from domain $D$ to a generalized domain $D^*$ such*

[1] 'more specific' does not necessarily mean 'more utilized'. We should take into account the cost of the relocations. We show, in Section 5, that utility gained due to lesser degrees of generalizations more than compensates the local utility loss due to relocations.

**Table. 1.** $\mathbf{T}$: private table; $\widehat{\mathbf{T}}$: a 10%-relocation of $T$; $\mathbf{T}^*_{\mu_1}$, $\widehat{\mathbf{T}}^*_{\mu_2}$: 3-anonymous single dimensional generalizations of $T$ and $\widehat{T}$ respectively; $\mathbf{T}^*_{\mu_2}$: a single dimensional generalization of $T$

| Id | Age | Nation | Zip | Sal. |
|---|---|---|---|---|
| q1 | 12 | Greece | 47906 | 13K |
| q2 | 19 | Turkey | 47907 | 15K |
| q3 | 17 | Greece | 47907 | 28K |
| q4 | 23 | Spain | 49703 | 14K |
| q5 | 38 | Brazil | 49705 | 18K |
| q6 | 33 | Peru | 49812 | 35K |
| q7 | 41 | USA | 49001 | 14K |
| q8 | 43 | Canada | 49001 | 25K |
| q9 | 48 | Canada | 49001 | 29K |

$T$

| Not. | Definition |
|---|---|
| $T$ | A private table |
| $T^*$ | A generalization of $T$ |
| $\widehat{T}$ | A relocation of $T$ |
| $\widehat{T}^*$ | A hybrid generalization of $T$ |
| $t \in T$ | A tuple in $T$ |
| $\mu$ | A generalization mapping |
| $T^*_\mu$ | The generalization of $T$ with mapping $\mu$ |

Notations

| Id | Age | Nation | Zip | Sal. |
|---|---|---|---|---|
| q1 | 12 | Greece | 47906 | 13K |
| q2 | 19 | Turkey | 47907 | 15K |
| q3 | 17 | Greece | 47907 | 28K |
| q4 | 31 | Brazil | 49703 | 14K |
| q5 | 38 | Brazil | 49705 | 18K |
| q6 | 33 | Peru | 49812 | 35K |
| q7 | 41 | USA | 49001 | 14K |
| q8 | 43 | Canada | 49001 | 25K |
| q9 | 48 | Canada | 49001 | 29K |

$\widehat{T}$

| Id | Age | Nation | Zip | Sal. |
|---|---|---|---|---|
| q1 | 11-30 | EU | 4* | 13K |
| q2 | 11-30 | EU | 4* | 15K |
| q3 | 11-30 | EU | 4* | 28K |
| q4 | 11-30 | EU | 4* | 14K |
| q5 | 31-50 | AM | 4* | 18K |
| q6 | 31-50 | AM | 4* | 35K |
| q7 | 31-50 | AM | 4* | 14K |
| q8 | 31-50 | AM | 4* | 25K |
| q9 | 31-50 | AM | 4* | 29K |

$T^*_{\mu_1}$

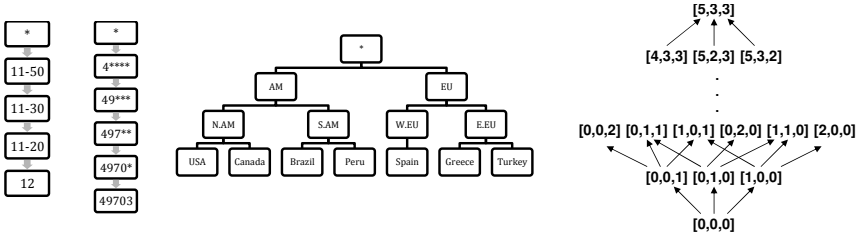
| Id | Age | Nation | Zip | Sal. |
|---|---|---|---|---|
| q1 | 11-20 | E. EU | 47* | 13K |
| q2 | 11-20 | E. EU | 47* | 15K |
| q3 | 11-20 | E. EU | 47* | 28K |
| q4 | 21-30 | W. EU | 49* | 14K |
| q5 | 31-40 | S. AM | 49* | 18K |
| q6 | 31-40 | S. AM | 49* | 35K |
| q7 | 41-50 | N. AM | 49* | 14K |
| q8 | 41-50 | N. AM | 49* | 25K |
| q9 | 41-50 | N. AM | 49* | 29K |

$T^*_{\mu_2}$

| Id | Age | Nation | Zip | Sal. |
|---|---|---|---|---|
| q1 | 11-20 | E. EU | 47* | 13K |
| q2 | 11-20 | E. EU | 47* | 15K |
| q3 | 11-20 | E. EU | 47* | 28K |
| q4 | 31-40 | S. AM | 49* | 14K |
| q5 | 31-40 | S. AM | 49* | 18K |
| q6 | 31-40 | S. AM | 49* | 35K |
| q7 | 41-50 | N. AM | 49* | 14K |
| q8 | 41-50 | N. AM | 49* | 25K |
| q9 | 41-50 | N. AM | 49* | 29K |

$\widehat{T}^*_{\mu_2}$

*that for $t \in D$ and $t^* \in D^*$; we have $\mu(t) = t^*$ (we also use notation $\Delta_\mu(t) = \mu(t)$ for consistency) only if $t^* \in \Delta(t)$. We say a table $T^*$ is a $\mu$-generalization of a table $T$ with respect to a set of attributes $QI$ and write $\Delta_\mu(T) = T^*$, if and only if records in $T^*$ can be ordered in such a way that $\Delta_\mu(T[QI][r]) = T^*[QI][r]$ for every row $r$.*




(a) DGH structures (b) Generalization Lattice

**Fig. 1.**

In Table 1, $T^*_{\mu_1}$ and $T^*_{\mu_2}$ are two generalizations of $T$ with mappings $\mu_1$ and $\mu_2$ respectively; E.g, $\Delta_{\mu_1}(T) = T^*_{\mu_1}$. $\Delta_{\mu_1}$(<41,US,49001>) = <31-50,AM,4****>

**Definition 3 (Single Dimensional Generalization).** *We say a mapping $\mu$ is $[i_1, \cdots, i_n]$ single dimensional iff given $\mu(t) = t^*$, we have $t^* = \Delta_{i_1 \cdots i_n}(t)$. We define in this case the level of $\mu$ as $i_1 + \cdots + i_n$.*

Each attribute in the output domain of a single dimensional mapping contains values from the same level of the corresponding DGH structure. In Table 1, $T^*_{\mu_1}$ and $T^*_{\mu_2}$ are [2,2,4] and [1,1,3] generalizations of $T$ respectively.

Given two single dimensional mappings $\mu_1$=$[i^1_1, \cdots, i^1_n]$ and $\mu_2 = [i^2_1, \cdots, i^2_n]$, we say $\mu_1$ is a higher mapping than $\mu_2$ and write $\mu_1 \subset \mu_2$ iff $\mu_1 \neq \mu_2$ and $i^1_j \geq i^2_j$ for all $j \in [1 - n]$.

We also cover multidimensional generalizations. Due to page limitations, we refer the reader to [22] for related definitions and discussion on the advantages of both approaches.

$k$-Anonymity privacy protection limits the linking of a record from a set of released records to a specific individual even if adversaries can link individuals to QI:

**Definition 4 ($k$-Anonymity [23,5]).** *A table $T^*$ is $k$-anonymous with respect to a set of quasi-identifier attributes $QI$ if each tuple in $T^*[QI]$ appears at least $k$ times.*

$T^*_{\mu_1}$ is a 3-anonymous generalization of $T$. Note that given $T^*_{\mu_1}$, the same adversary can at best link Bob to tuples q5, q6, q7, q8, and q9.

**Definition 5 (Equality group).** *The equality group of tuple $t$ in dataset $T^*$ is the set of all tuples in $T^*$ with identical quasi-identifiers to $t$.*

In dataset $T^*_{\mu_1}$, the equality group for tuple $q7$ is $\{q5, q6, q7, q8, q9\}$. We use colors to indicate equality groups in Table 1.

While $k$-anonymity limits identification of tuples, it fails to enforce constraints on the sensitive attributes in a given equality group, thus there is still a risk of sensitive information disclosure. We start our analysis with $k$-anonymity because it has a simple definition and $k$-anonymity and $k$-anonymization is still used in several domains as a privacy metric [9,27,25] and as a sub procedure [26].

In Section 4, we use the *anti-monotonicity* property of $k$-anonymity. Given $\mu^1 \subset \mu^2$ and a dataset $T$, if $\Delta_{\mu^1}(T)$ is not $k$-anonymous, neither is $\Delta_{\mu^2}(T)$. In Table 1, if $T^*_{\mu_2}$ is not 3-anonymous, neither is $T$.

There may be more than one $k$-anonymization of a given dataset, and the one with the most information content is desirable. Previous literature has presented many metrics to measure the utility of a given anonymization [12,20,13,6,2]. We use the LM cost metric defined in [12]. Given $a$ is the number of attributes:

$$LM(T^*) = \frac{1}{|T^*| \cdot a} \sum_{i,j} \frac{|\Delta^{-1}(T^*[i][j])| - 1}{|\Delta^{-1}(*)| - 1}$$

**Related Work.** The value of utility preservation in anonymized dataset has been widely recognized by the literature since the very first works on anonymization-based privacy protection.

The first class of works on utility introduces new heuristic algorithms that generates equality groups composed of tuples that are as close to each other as possible. Grouping of close tuples achieves better utilized generalizations. [23] observes that all possible single dimensional mappings create a lattice over the subset operation. The proposed algorithm finds an optimal $k$-anonymous generalization (optimal in minimizing a utility cost metric) by performing a binary search over the lattice. [14] improves this technique with a bottom-up pruning approach and finds all optimal $k$-anonymous generalizations. [2] introduces more flexibility by relaxing the constraint that every value in the generalization should be in the same generalized domain. Works in [18,19,16,7,21] adopt previous single dimensional algorithms for other privacy notions such as $\ell$-diversity, $t$-closeness, and $\delta$-presence. Among other works on heterogeneous generalizations, works in [20,4,1,17] use clustering techniques to provide $k$-anonymity. [15] and [11] partition the multidimensional space to form $k$-anonymous and $\ell$-diverse groups of tuples. [8] makes use of space filling curves to reduce the dimensionality of the database and provides $k$-anonymity and $\ell$-diversity algorithms that work in one dimension. All of the above works are based on pure generalizations and are orthogonal to our approach. As will be clear in later sections, the relocation technique proposed in this paper can be used to utilize most generalizations regardless of the underlying algorithm. Even though we are not proposing standalone anonymization algorithms, our approach can be considered in this category as we aim to create better equality groups out of existing groups at the cost of truthfulness.

Another way to improve utility is by releasing more information on the equality groups without changing the groupings of the tuples. Our approach differs from such an approach as we form new groupings without specifying how we release the groups. We refer the reader to [22] for a detailed discussion on these approaches.

## 3 Hybrid Anonymizations

### 3.1 Classical Adversaries

The classical adversary is the same adversary addressed in most previous literature (see Section 2). The adversary knows the QI attributes of an individual and tries to discover, from the released dataset, the sensitive value belonging to the individual.

As mentioned in Section 1, data relocations can improve utility of released datasets. One should be careful on the number of relocations applied to the dataset as each relocation makes data less truthful. We now formally define *$p$%-table relocations* in which the maximum number of cell relocations is bounded by the $p$% of the whole dataset:

**Definition 6 ($p$%-Table Relocations).** *We say a table $\widehat{T}$ is a $p$%-relocation of a table $T$ with respect to a set of attributes $QI$ and write $\widehat{T} \frown^p T$, if and only if records in $\widehat{T}$ can be ordered in such a way that*

- *$T[c][r] \neq \widehat{T}[c][r]$ for at most $p$% of all possible (attribute $c \in QI$, row $r$) pairs and*
- *$T[c][r] = \widehat{T}[c][r]$ for all (attribute $c \notin QI$, row $r$) pairs.*

In Table 1, $\widehat{T}$ is a 10%-relocation of $T$ as only two data cells (see q4) out of 27 is relocated. We now formally define hybrid anonymizations which are created by anonymizing table relocations:

**Definition 7 ($p$%-Hybrid Generalization).** *We say a table $\widehat{T}^*$ is a $p$%-hybrid generalization of a table $T$ with some mapping $\mu$ if and only if there exist a $\widehat{T} \frown^p T$ such that $\widehat{T}^* = \Delta_\mu(\widehat{T})$.*

In Table 1, $\widehat{T}^*_{\mu_2}$ is a 10%-hybrid generalization of table $T$ with mapping [1,1,3]. From now on, we assume $p = 10\%$ and do not mention $p$ in our discussions.

**Definition 8 (Hybrid $k$-anonymity).** *We say a table $\widehat{T}^*$ is a $k$-anonymous hybrid of a table $T$ if and only if $\widehat{T}^*$ is a $p$%-hybrid generalization of a table $T$ and $\widehat{T}^*$ is $k$-anonymous.*

In Table 1, $\widehat{T}^*_{\mu_2}$ is a 3-anonymous hybrid of the table $T$.

As the domain of all possible generalizations of a given table $T$ is a subset of the domain of all possible hybrid anonymizations of $T$, LM cost of an optimal hybrid anonymization will be at least as small as that of a generalization under the same privacy standard. For example, $T^*_{\mu_1}$ and $\widehat{T}^*_{\mu_2}$ both satisfy 3-anonymity, however, $\widehat{T}^*_{\mu_2}$ has a smaller LM cost as $\mu_2$ is a more specific mapping. This does not necessarily mean $\widehat{T}^*_{\mu_2}$ is more utilized as LM cost does not take into account the information loss due to relocations. However, in practice, for most applications, a small number of relocations can increase the overall utility of the released dataset at the expense of decreasing utility on relocated data cells. In order to benefit from hybrid anonymizations, we now state the problem of $k$-anonymity in the context of hybrid anonymizations. In Section 4, given a private table $T$, we propose algorithms to find a $k$-anonymous hybrid $\widehat{T}^*$ of $T$ that minimizes the LM cost metric. .

### 3.2 Statistical Adversaries

In a hybrid anonymization, the distribution of the tuples in the released data will deviate from the original distribution. If the deviation is too large, an adversary that knows about the original distribution may suspect that some groups in the released data have been artificially populated. For example, in a census dataset, if the adversary sees that there are considerably more males than females, the adversary can suspect that some females are relocated. To defend against such attack, the distance between the original distribution and relocated distribution should be bounded such that the deviation should look as if occurred by chance.

**Definition 9 ($\alpha$-Hybrid $k$-Anonymization).** *Let $T$ be a private table and $X$ be the multinomial random variable from which the tuples are drawn from. $\widehat{T}^*$ is an $\alpha$-hybrid if the hypothesis that the group sizes in $\widehat{T}^*$ are consistent with the parameters of $X$ cannot be rejected at the significance level $\alpha$.*

For significance testing, we use the Pearson's chi-squared test for multinomial distributions. Given $\widehat{T}^* = \{G_1, \cdots, G_n\}$ with mapping $\mu$ and size $N$, the $X^2$ can be approximated as follows. Let $E_i = N \cdot \sum_{t \mid \mu(t)=G_i} \mathcal{P}(X = t)$.

$$X^2 = \sum_{i}^{n} (\frac{|G_i| - E_i}{E_i})$$

Note that we assume a strong adversary that knows the exact distribution $X$ of the tuples or sensitive values. In reality, the adversaries may only know partial information about $X$, such as "the number of Italians is less than Chinese". As it is difficult to predict the true background of the adversary, we assume a worst case scenario.

In addition to the above mentioned adversaries, we will also assume adversaries might know the underlying hybrid algorithm. It has been shown in [26,29] that such adversaries can reverse-engineer the anonymization algorithm and learn information that would not be allowed by the underlying privacy metric. However, ensuring a theoretical bound on the disclosure against such adversaries is not a trivial problem and can also result in a huge decrease in utility. Instead, we will make it practically hard for such an adversary to reverse-engineer the algorithm by making random decisions during the algorithm. We will employ multiple random sources within the algorithm that will generate many possible pathways for the algorithm to follow.

**Algorithm 1.** S-Hybrid

**Require:** a private table $T$ from domain $D$, privacy parameter $k$, a utility cost metric $CM$, a user threshold $p$;

**Ensure:** return a minimum cost $k$-anonymous single dimensional hybrid generalization of $T$.

1: create lattice $lat$ for all possible generalization mappings for $D$. Let $n$ be the maximum level of mappings in $lat$.
2: **for all** level $i$ from $n$ to 0 **do**
3:  **for all** mapping $\mu$ of level $i$ in $lat$ **do**
4:  $\widehat{T}^* = createHybrid(\Delta_\mu(T), k, p)$
5:  **if** $\widehat{T}^*$ is not $k$-anonymous **then**
6:  delete node $\mu$ and all children and grandchildren of $\mu$ from $lat$.
7:  **else**
8:  calculate cost $CM(\widehat{T}^*)$ and store the cost on the lattice node.
9: **if** no node exists on the $lat$ **then**
10:  return null.
11: find the mapping $\mu$ with the minimum cost on the $lat$.
12: return $createHybrid(\Delta_\mu(T), k, p)$

## 4 Hybrid Anonymization Algorithms

In this section, we present a set of single dimensional hybrid $k$-anonymization algorithms each addressing a different adversary as mentioned in Section 3. All

**Algorithm 2.** CreateHybrid

**Require:** a generalization $T^*$, privacy parameter $k$, and a user threshold $p$;
**Ensure:** return $p\%-$hybrid generalization $\widehat{T}^*$ with the same mapping as that of $T^*$. $\widehat{T}^*$ will not contain any more non $k$-anonymous groups than $T^*$.

$\widehat{T}^* = T^*$;
let $G$ be the set of equality groups in $\widehat{T}^*$.
let $G_{sm} \subset G$ be the set of groups with less than or equal to k/2 tuples.
let $G_{big} \subset G$ be the set of groups with more than k/2 tuples, but less than k tuples.
**for all** $g \in G_{sm}$ **do**
  let $ptr$ be an empty group pointer to hold a target group.
  **if** $G_{big}$ is not empty **then**
    find the group $g' \in G_{big}$ that is closest to $g$.
    $ptr \rightarrow g'$
  **else**
    find the group $g' \in G - G_{sm}$ that is closest to $g$.
    $ptr \rightarrow g'$
  **if** $ptr$ is null **then**
    return $\widehat{T}^*$.
  change all tuples in $g$ so that the tuples are moved (relocated) into $g'|ptr \rightarrow g'$
  remove $g'$, update $G, G_{sm}, G_{big}$ accordingly.
  **if** $\widehat{T}^*$ is not a $p\%-$hybrid generalization **then**
    roll back the last change and return $\widehat{T}^*$;
**for all** $g \in G_{big}$ **do**
  find the group $g' \in G - G_{sm} - G_{big}$ that has more than $2k - |g|$ tuples and is closest to $g$.
  **if** no such $g'$ exists **then**
    return $\widehat{T}^*$;
  pick any $k - |g|$ tuples in $g'$ and change the tuples such that they are moved (relocated) into $g$.
  update $G, G_{sm}, G_{big}$ accordingly.
  **if** $T^*$ is not a $p\%-$hybrid generalization **then**
    roll back the last change and return $\widehat{T}^*$;
return $\widehat{T}^*$;

algorithms trace the whole space of single dimensional mappings and returns a mapping as an approximation to the problem of hybrid anonymity. The algorithms are based on the optimal single dimensional anonymization algorithm, Incognito [14] but improve Incognito by searching the space of hybrid generalizations (Definition 7) rather than table generalizations 3.

**Deterministic S-Hybrid.** The pseudocode for the S-Hybrid algorithm is given in Algorithm 1. The algorithm traverses the whole space of single dimensional

mappings, applies each mapping to the private dataset, produces a hybrid generalization by calling the function createHybrid. Fortunately, the possible single dimensional mappings over a table domain form a lattice on the $\subset$ relation (see Figure 1(b)). In lines 2-8, we traverse the lattice in a top-down manner. In lines 5-6, we use the anti-monotonicity property of $k$-anonymity to prune the lattice, thus reduce the search space.

The pseudocode for the createHybrid function is given in Algorithm 2. The aim of the algorithm is to convert the given generalization into a $k$-anonymous hybrid generalization with fewest relocations as possible. We start by classifying the groups as $G_{sm}$ (groups with less than or equal to $k/2$ tuples), $G_{big}$ (groups with more than $k/2$ but less than $k$ tuples) and $G$ (all groups). The reason for such a classification is that distributing the tuples in groups of few tuples while completing groups that have almost $k$ tuples potentially minimizes the required number of relocations. With such reasoning, the relocation of tuples are done in two phases:

**Distribution:** In lines 5-18, the algorithm attempts to relocate the tuples in $G_{sm}$ first into the closest group in $G_{big}$. Two tuples are closest if they agree on the most number of attributes. If $G_{big}$ is empty, the tuples are relocated into the closest $k$-anonymous group. After this phase, some groups in $G_{big}$ may become $k$-anonymous, thus may be removed from $G_{big}$.

**Completion:** In lines 19-26, the algorithm relocates tuples from closest $k$-anonymous groups that has enough number of tuples into groups in $G_{big}$.

After each relocation, the algorithm checks if the maximum number of allowed relocations (specified with the input $p$) has been exceeded. If that is the case, the algorithm roll backs the last relocation and returns the non $k$-anonymous hybrid generalization generated so far.

As an example, in Table 1, if we use $\mu_2$=[1,1,3] as the generalization mapping, $k = 3$, and $p = \%10$; $T^*_{\mu_2}$ will be the input to the createHybrid algorithm. The algorithm will set $G_{sm} = \{\{q4\}\}$ and $G_{big} = \{\{q5, q6\}\}$. The algorithm starts distributing the tuples in $G_{sm}$ into groups in $G_{big}$. The tuple $q4$ will be sent to the only (closest) group $\{q5, q6\}$ in $G_{big}$. As a result, $q4$ becomes `<31,Brazil,49703>`. Note that this change only relocates 2 out of 27 data cells thus performing the change creates a 10%-table relocation. Since the resulting hybrid generalization is $k$-anonymous, the algorithm returns $\widehat{T}^*_{\mu_2}$.

**Randomized S-Hybrid.** As mentioned in Section 3, adversaries that know the underlying algorithm can attempt to reverse-engineer the algorithm and create non $k$-anonymous subgroups [26,29]. Such attacks pose a threat to privacy especially if the underlying algorithm is deterministic. To resist reverse-engineering attacks, we create the Randomized S-Hybrid algorithm. The algorithm makes random decisions at certain points, thus can follow multiple pathways making reverse-engineering attacks difficult. The sources of randomness can be listed as follows:

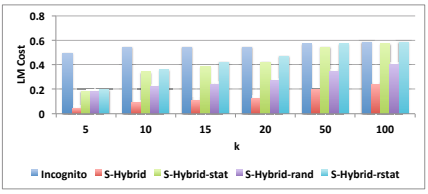
$\rightarrow$ In the distribution/completion phases, in lines 8, 11, and 20, instead of picking the closest group as the target/source group for relocation, we pick the

target group randomly from the sets $G_{big}$ or $G$. Most of the time, the sizes of these sets are large enough to create a probability space of sufficient size for the flow of the algorithm.

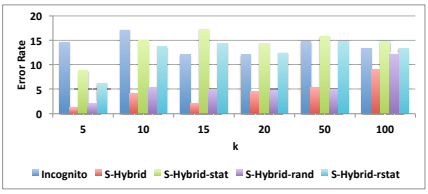
→ In the completion phase, in line 23, instead of relocating exactly $k - |g|$ tuples, we relocate a random number of tuples such that both the target and the source group remains / becomes $k$-anonymous. Relocating a random number of tuples prevents the source group to contain exactly $k$ tuples.

**Statistical S-Hybrid.** As mentioned in Section 3.2, statistical adversaries use the known distribution of the tuples to identify artificial relocations. $\alpha$-Hybrid $k$-anonymization addresses such adversaries by bounding the statistical difference between the original distribution and relocated distribution. Deterministic S-Hybrid algorithm can easily be modified so that it accepts the statistical threshold $\alpha$ as an input and returns $\alpha$-Hybrid generalizations. In Algorithm 2, in lines, 17 and 25, whenever we check if a relocation violates $p$%-Hybrid anonymity, we instead check if the relocation violates $\alpha$-Hybrid anonymity.

It should be noted that even with low $\alpha$ settings, $\alpha$-Hybrid anonymity is a strict privacy definition. That is, the definition allows fewer number relocations making it more difficult to create a $k$-anonymous hybrid from a non $k$-anonymous generalization. In Section 5, we empirically compare $\alpha$-Hybrid anonymity with $p$%-Hybrid anonymity in terms of utility and show that the former allows a lower level of utility at the benefit of stronger privacy.




(a) Single Dimensional- LM Cost



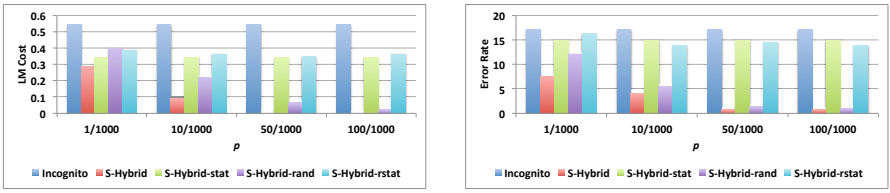

(b) Single Dimensional - Error rate

**Fig. 2.** Varying $k$ - Sal

We also designed multidimensional versions of the algorithms proposed in this section (M-Hybrid). Due to space constraints, we refer the reader to [22] for details.

## 5 Experiments

This section presents the experimental evaluation of S-Hybrid and M-Hybrid algorithms. In addition to the three algorithms mentioned in Section 4, we also evaluate algorithm *S-Hybrid-rstat* which is randomized S-Hybrid against statistical adversaries. During our experiments, we use the real datasets 'Sal' and 'Occ' that were extracted from CENSUS data and previously used by [28,8]. Both datasets contain 100.000 tuples. As the results were similar for both dataset, we




(a) Single Dimensional- LM Cost (b) Single Dimensional - Error rate

**Fig. 3.** Varying Distortion Limit $p$ - Sal

present results only on the 'Sal' dataset. Results on the 'Occ' dataset can be accessed from [22].

We used two metrics to measure utility: LM cost metric defined in Section 2 and the range query error rate used in [8,15]. Query error is calculated by issuing count queries on anonymizations and normalizing the deviation of count from the original count in the private dataset.

**Varying Privacy Parameter** $k$

We first fix the distortion limit as %1, $\alpha$ as %5, vary the value of $k$ and compare S-Hybrid algorithm with the previously proposed single dimensional Incognito algorithm with respect to the LM cost and query error metric. Note that %1 distortion limit is almost a negligible sacrifice from the truthfulness of data. Figure 2 shows the results on the 'Sal' dataset. According to utility cost experiments, in nearly all cases, all Hybrid approaches give better results than the algorithm Incognito. S-Hybrid and randomized S-Hybrid perform better than statistical Hybrid algorithms. As mentioned before, this is because, compared to hybrid $k$-anonymity, $\alpha$-hybrid $k$-anonymity is a strict privacy definition assuming a powerful adversary. In most cases, the number of allowed relocations for $\alpha$-hybrid anonymity is much smaller than that for hybrid anonymity resulting in less utilized anonymizations. For statistical S-Hybrid and Incognito, for some cases, we observed fluctuations in error rates when we increase $k$. The reason is that, in these settings, the resulting mappings cannot be ordered with respect to the $\subset$ operator (see the definition of higher mappings in Section 2) and are close to each other in the generalization lattice. Any one of the mappings can be considered better utilized than the other depending on the underlying application.

**Varying Distortion Limit** $p$

In these experiments, we fix the value of $k$ as 10, $\alpha$ as %5, vary the value of the distortion limit $p$. We present the results in Figure 3. We see that non statistical S-Hybrid algorithms increase in utility as we increase the distortion limit (e.g., as we apply more and more relocations). Statistical S-Hybrid algorithms are not very sensitive to changes in $p$. The reason is that significance test via the parameter $\alpha$ is more decisive on the number of allowed relocations than the limit via the $p$. Generally, significance test for further relocations fail even before

the number of relocations reach %0.1. Thus the utility of statistical S-Hybrid algorithms do not change much. The comparison of algorithms with each other is similar as mentioned in the previous section.

We also made experiments regarding the multidimensional hybrid algorithm, M-Hybrid. M-Hybrid algorithms show a similar behavior. Due to space constraints, we refer the reader to [22] for experimental results.

## 6 Future Work

As a possible future work, new hybrid algorithms can be designed for other privacy metrics such as $\ell$-diversity, $(\alpha, k)$-anonymity or $\delta$-presence. This would be crucial in addressing different types of adversaries. There is also room for improvement for the hybrid algorithms proposed in this paper. For example, one can design hybrid algorithms that would theoretically bound the probability of identification against algorithm-aware adversaries. Hybrid techniques can also be evaluated with respect to different cost metrics and real applications so that utility gain can better be quantified under different scenarios.

## References

1. Agrawal, G., Feder, T., Kenthapadi, K., Khuller, S., Panigrahy, R., Thomas, D., Zhu, A.: Achieving anonymity via clustering. In: PODS 2006: Proceedings of the 25th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Chicago, IL, USA, June 26-28, pp. 153–162 (2006)
2. Bayardo, R.J., Agrawal, R.: Data privacy through optimal k-anonymization. In: ICDE 2005: Proceedings of the 21st International Conference on Data Engineering, pp. 217–228. IEEE Computer Society, Washington, DC (2005)
3. Brickell, J., Shmatikov, V.: The cost of privacy: destruction of data-mining utility in anonymized data publishing. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, pp. 70–78. ACM, New York (2008)
4. Byun, J.-W., Kamra, A., Bertino, E., Li, N.: Efficient k-anonymization using clustering techniques. In: Kotagiri, R., Radha Krishna, P., Mohania, M., Nantajeewarawat, E. (eds.) DASFAA 2007. LNCS, vol. 4443, pp. 188–200. Springer, Heidelberg (2007)
5. Ciriani, V., De Capitani di Vimercati, S., Foresti, S., Samarati, P.: k-anonymity. In: Secure Data Management in Decentralized Systems, pp. 323–353 (2007)
6. Domingo-Ferrer, J., Torra, V.: Ordinal, continuous and heterogeneous k-anonymity through microaggregation. Data Mining and Knowledge Discovery 11(2), 195–212 (2005)
7. Fung, B.C.M., Wang, K., Yu, P.S.: Top-down specialization for information and privacy preservation. In: ICDE 2005: Proceedings of the 21st International Conference on Data Engineering, pp. 205–216. IEEE Computer Society, Washington, DC (2005)
8. Ghinita, G., Karras, P., Kalnis, P., Mamoulis, N.: Fast data anonymization with low information loss. In: VLDB 2007: Proceedings of the 33rd International Conference on Very Large Data Bases, pp. 758–769. VLDB Endowment (2007)

9. Gionis, A., Mazza, A., Tassa, T.: k-anonymization revisited. In: IEEE 24th International Conference on Data Engineering, ICDE 2008, pp. 744–753 (April 2008)
10. Standard for privacy of individually identifiable health information. Federal Register, 66(40) (February 28, 2001)
11. Hore, B., Ch, R., Jammalamadaka, R., Mehrotra, S.: Flexible anonymization for privacy preserving data publishing: A systematic search based approach. In: Proceedings of the 2007 SIAM International Conference on Data Mining (2007)
12. Iyengar, V.S.: Transforming data to satisfy privacy constraints. In: KDD 2002: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 279–288. ACM, New York (2002)
13. Kifer, D., Gehrke, J.: Injecting utility into anonymized datasets. In: SIGMOD 2006: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, pp. 217–228. ACM, New York (2006)
14. LeFevre, K., DeWitt, D.J., Ramakrishnan, R.: Incognito: Efficient full-domain k-anonymity. In: SIGMOD 2005: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, pp. 49–60. ACM, New York (2005)
15. LeFevre, K., DeWitt, D.J., Ramakrishnan, R.: Mondrian multidimensional k-anonymity. In: ICDE 2006: Proceedings of the 22nd International Conference on Data Engineering, Atlanta, GA, April 3-7, pp. 25–35 (2006)
16. Li, N., Li, T.: t-closeness: Privacy beyond k-anonymity and l-diversity. In: ICDE 2007: Proceedings of the 23rd International Conference on Data Engineering, Istanbul, Turkey, April 16-20 (2007)
17. Lin, J.-L., Wei, M.-C., Li, C.-W., Hsieh, K.-C.: A hybrid method for k-anonymization. In: Asia-Pacific Services Computing Conference, APSCC 2008, pp. 385–390. IEEE (2008)
18. Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M.: $\ell$-diversity: Privacy beyond $k$-anonymity. In: ICDE 2006: Proceedings of the 22nd IEEE International Conference on Data Engineering, Atlanta Georgia (April 2006)
19. Nergiz, M.E., Atzori, M., Clifton, C.: Hiding the presence of individuals in shared databases. In: SIGMOD 2007: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, Beijing, China, June 11-14 (2007)
20. Nergiz, M.E., Clifton, C.: Thoughts on k-anonymization. Data and Knowledge Engineering 63(3), 622–645 (2007)
21. Nergiz, M.E., Clifton, C.: $\delta$-Presence without complete world knowledge. IEEE Transactions on Knowledge and Data Engineering, 868–883 (2009)
22. Nergiz, M.E., Gok, M.Z., Ozkanli, U.: Preservation of utility through hybrid k-anonymization. Technical Report TR 2013-001, Department of Computer Engineering, Zirve University (2013)
23. Samarati, P.: Protecting respondent's identities in microdata release. IEEE Transactions on Knowledge and Data Engineering 13(6), 1010–1027 (2001)
24. Samarati, P., Sweeney, L.: Generalizing data to provide anonymity when disclosing information (abstract). In: PODS 1998: Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, p. 188. ACM, New York (1998)
25. Tamersoy, A., Loukides, G., Nergiz, M.E., Saygin, Y., Malin, B.: Anonymization of longitudinal electronic medical records. IEEE Transactions on Information Technology in Biomedicine 16(3), 413–423 (2012)

26. Wong, R.C.-W., Fu, A.W.-C., Wang, K., Pei, J.: Minimality attack in privacy preserving data publishing. In: VLDB 2007: Proceedings of the 33rd International Conference on Very Large Data Bases, pp. 543–554. VLDB Endowment (2007)
27. Wong, R.C.-W., Li, J., Fu, A.W.-C., Wang, K. ($\alpha$, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing. In: KDD 2006: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 754–759. ACM, New York (2006)
28. Xiao, X., Tao, Y.: Anatomy: Simple and effective privacy preservation. In: VLDB 2006: Proceedings of 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, pp. 139–150 (2006)
29. Zhang, L., Jajodia, S., Brodsky, A.: Information disclosure under realistic assumptions: privacy versus optimality. In: CCS 2007: Proceedings of the 14th ACM Conference on Computer and Communications Security, pp. 573–583. ACM, New York (2007)

# The Security of Information Systems in Greek Hospitals

George Aggelinos and Sokratis K. Katsikas

Department of Digital Systems, University of Piraeus,
150 Androutsou St., Piraeus 185 32, Greece
gaggelinos@yahoo.gr, ska@unipi.gr

**Abstract.** Hospital information systems increasingly handle information related to the health status of citizens, due to the proliferation of the use of the electronic health record. Because of this, hospital information systems constitute a part of a country's critical information infrastructure; hence, the state of affairs regarding their security is of paramount interest. In this paper we report on the results of a survey performed among all public Greek hospitals regarding the security of their information systems. Comparisons to the situation in other countries as this is manifested in similar studies are made and conclusions are drawn; these indicate that there is much room for improvement.

**Keywords:** Hospital information systems, health information security, security policies.

## 1 Introduction

A fundamental change regarding ICT in health care is the transition from the traditional model of a stand-alone HIS, that is the HIS operating within the boundaries of a single Healthcare Organization (HO), to the networked HIS, that is a HO's HIS interconnected to HISs of other HOs or even of third parties, over national or international Wide Area Networks (WANs). Moreover, web-based e-health services are already been regularly provided and the healthcare sector has started exploiting the cloud computing paradigm [1]. Additionally, mobile devices like laptops, PDAs and even mobile phones are being increasingly used by the healthcare industry to access, store or transmit health information within the framework of providing health services [2].

In Greece, hospitals started using information systems considerably later than in other European countries. The process not only started late, but proved to be slow and difficult. This was mainly due to the frequent re-organization of the public health system of the country in the past fifteen years. Whereas initial designs called for isolated information systems per hospital, subsequent changes in the administrative hierarchy of HOs were only partially reflected to information systems. In such a landscape, the security of hospital information systems was (and still is) often not given the required attention. This is why the results of the survey may prove to be very useful to policy makers and administrators, as they allow the identification of areas where attention should be given and, accordingly, possible directions for

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 112–123, 2013.

management priorities. As comparisons with findings from other countries reveal, many findings are not dissimilar, thus highlighting the need for intensifying efforts in securing information systems operating in HOs on a global rather than local level.

This paper reports on the findings of a recent survey of hospital information systems in Greek public hospitals that pertain to the security of their information systems. The remaining of the paper is structured as follows: In section 2 we describe the survey process and the questionnaire used. In section 3 we discuss the survey findings with references made to respective findings in similar studies performed in other countries. Finally, section 4 summarizes our conclusions and recommendations.

## 2 The Survey Process and Questionnaire

The questionnaire we used consisted of 162 questions. Possible responses are structured (yes/no) or of a Likert type scale. At the end of each thematic area there is space available for free text to be used for further explanations or details on the answers given.

The questionnaire was structured in two sections; each section was further structured in three thematic areas. The first section is designed to explore the hospital information system. The three thematic areas herein are the organization of the information system, wherein the structure of the information system and some generic administrative operations are explored (30 questions); the hospital functions covered by the information system (33 questions); and the security of the information system (43 questions). The second section was designed to explore the business continuity and the disaster recovery capacity of the hospital. In the sequel, we will only discuss findings pertinent to the first section of the survey and particularly to the last thematic area, even though use of responses to questions posed in other sections/areas of the survey will be also utilized, so as to get as clear and full picture as possible on the security of the information system.

The survey questionnaire was aimed at all public hospitals in Greece; there are 139 of these, scattered all over the country. We decided to extend the survey only to public hospitals, as private HOs in Greece are mostly small clinics [3].

Even though the security of information systems should ideally be the responsibility of the top management in any organization, this is far from being the case in Greek hospitals, where top management is usually unaware of such issues. Consequently, the survey questionnaire was sent by e-mail to the IT Director of each hospital, following a telephone contact with them. Respondents were encouraged to consult with the competent personnel when formulating their responses rather than providing their personal (perhaps misinformed) opinion. Monitoring of the progress of responding to the questionnaire was also done over the phone. Responses were also received by e-mail. Statistical processing of the responses was made using SPSS v.19.

### 2.1 Response Rates and Profile of Respondents

Questionnaires were sent to all 139 public hospitals in the country. Completed questionnaires were received from 100 hospitals (71.94%), 14 hospitals (10.07%)

either declined taking part in the survey or stated that they do not have an IT Department and the remaining 25 hospitals (17.99%) did not respond. This response rate is similar to those reported in [4], [5] and far larger than that reported in [6]; however, the absolute number of respondents of the present survey is closer to that of [6] and far larger than those of [4] and [5]. The returned questionnaires contained answers to more than 99% of the questions.

From a health administration point of view, the country is divided into seven administrative health regions. The response rate among hospitals in different regions ranged from 61.29% to 85%.

The questionnaire included six identifying questions, as well as the identity and contact information of the IT Director were included. These questions were answered by 99.66% of the respondents, thus allowing full identification of all hospitals and, consequently, association of responses to hospitals. 63% of the responses came from general hospitals, 13% from general University hospitals, 7% from regional general hospitals and 14% from other types of HOs, such as health centers.

The distribution of the hospital size, in terms of number of beds, is given in Fig. 1.

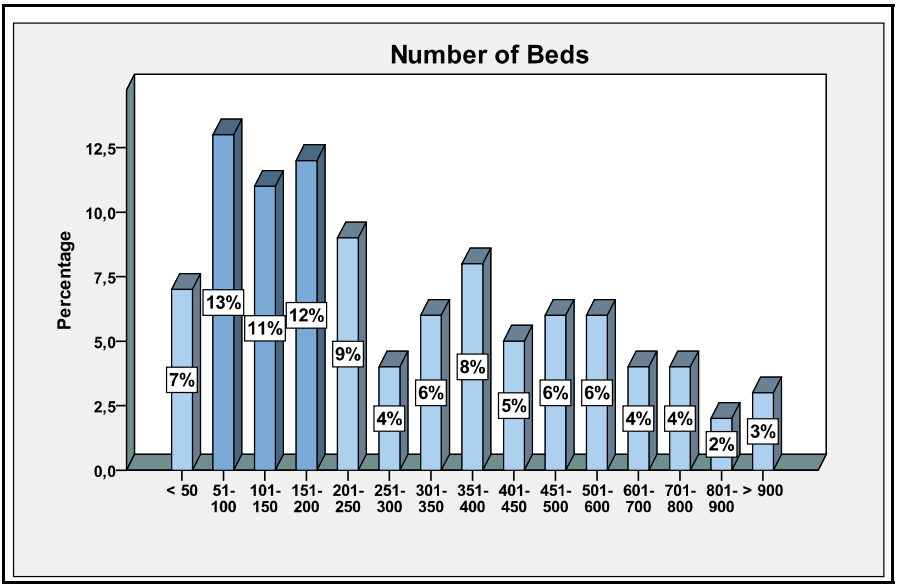



**Fig. 1.** Hospital size by number of beds

## 2.2 Information Systems' Configuration

Initially, the strategy followed to develop hospital information systems was to develop individual information systems (e.g. Patient administration information system, ERP, Laboratory Information System, radiology information system etc.) which would be federated (sometimes loosely) into a whole. This strategy proved to

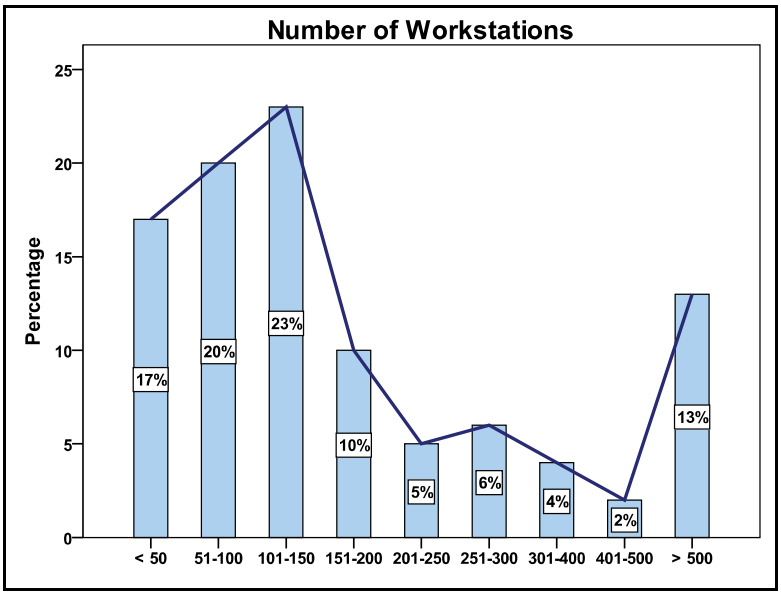



**Fig. 2.** IS workstations

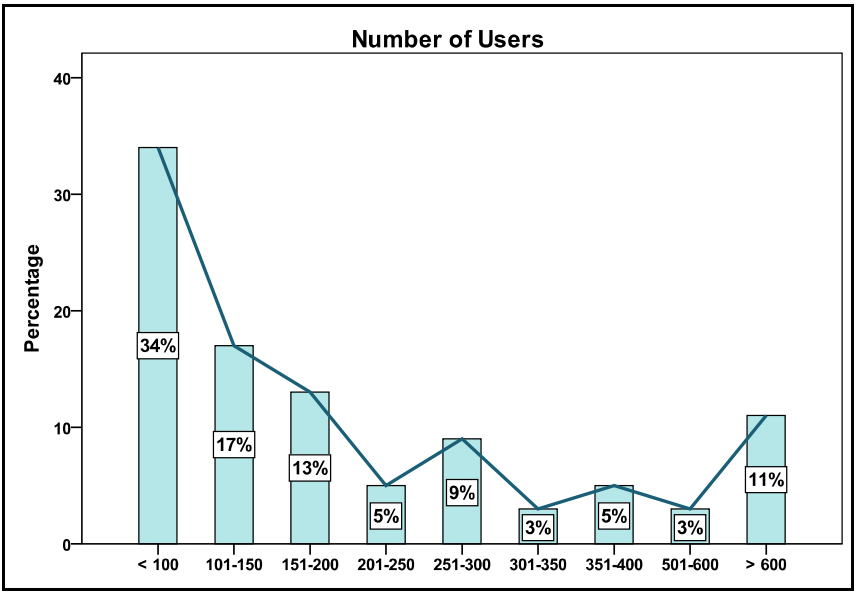



**Fig. 3.** IS users

be inefficient, mainly due to the lack of interoperability standards and of wide consensus on issues related to the medical practice itself (e.g. standardization of the contents of the medical record). The survey revealed that this strategy has already started to change, towards the direction of developing, right from the start, one and

only integrated hospital information system; this is found to be the case in 51% of the hospitals. Answers to related questions on the age of the installed information systems indicate that 12% are less than two years old and 39% are little more than three years old. This compares to the report [4] that in 2005, in the U.S., 22% of the installed systems were less than ten years old whereas 56% were older than ten years.

The client/server architectural model is followed in 91.9% of the cases; servers appear to be physically positioned in one space only (server room) in 85% of the cases. The remaining 15% is clearly moving towards the same direction of positioning all servers in the same location. The number of workstations connected to the information system is shown in Fig. 2 and the number of users is shown in Fig. 3. It is interesting to note that in 9% of the hospitals the number of workstations is between 500 and 600, while the number of users is larger than 600.

## 3 Security of Information Systems

The findings in this section are grouped into categories matching the security control clauses of ISO 27002:2005 [7] and ISO 27799:2008 [8].

### 3.1 Risk Assessment and Treatment

ISO 27002 stipulates that management actions and priorities for managing information security risks and for implementing controls selected to protect against these risks should be guided and determined by risk assessments. Risk assessment should be performed according to ISO 27005:2011 [9]. Interestingly, only 37% of the hospitals had performed a risk assessment and treatment exercise for their information systems, whereas only 7.5% reported having used a specific risk analysis methodology; this was CRAMM [10] in all cases. This means that the vast majority of Greek hospitals face information security risks which have not been measured and which are probably completely unknown.

### 3.2 Security Policy

Any serious effort towards securing an organization's information systems starts with the formulation, establishment and enforcement of a security policy that provides clear guidance on all matters related to the security of systems and information [7]. According to ISO 27002, "management should set a clear policy direction in line with business objectives and demonstrate support for, and commitment to, information security through the issue and maintenance of an information security policy across the organization". When asked whether a security policy exists, only 57% of the respondents responded affirmatively. Of these, a very small (5) number of hospitals do not enforce their security policy, even though they do have one in place. Therefore, only 54.7% of Greek hospitals have a security policy in place, which is also enforced.

The above notwithstanding, it is possible that some security controls are enforced without the formal establishment of a security policy. This was also investigated by means of a question "whether there exists a way for making the personnel aware of security controls and procedures and whether the personnel are being trained on these"; only a small percentage of 20% responded affirmatively. 15% of those hospitals that do have a means of making personnel aware of security controls and procedures and 17% of those that train personnel on these have a security policy in place. This finding is in line with the conclusion arrived at in [11], that one of the three major issues relevant to security among clinical staff in hospitals is the change in the approach taken towards training such personnel on data security in hospital environments.

### 3.3 Organization of Information Security

In order to manage information security within the organization, a management framework should be established to initiate and control the implementation of information security. Within this framework, security roles should be defined and assigned [7].

Greek hospitals have defined a role of IT Security Officer (ITSO) only in 10% of the cases. It is, however, possible that other hospitals have delegated these responsibilities and duties to the IT Directors. As it will be clearly seen in the sequel, the definition and assignment of this role is paramount to adhering with other crucial areas of security controls.

### 3.4 Asset Management

In order to achieve and maintain appropriate protection of organizational assets, all assets should be accounted for and have a nominated owner [7].

In our case, 32% of the hospitals report that responsibility for IT assets has been assigned to specific staff members.

### 3.5 Human Resources Security

In order to ensure that employees, contractors and third party users understand their responsibilities and are suitable for the roles they are considered for, and to reduce the risk of theft, fraud or misuse of facilities, security responsibilities should be addressed prior to employment in adequate job descriptions and in terms and conditions of employment. All candidates for employment, contractors and third party users should be adequately screened, especially for sensitive jobs and employees, contractors and third party users of information processing facilities should sign an agreement on their security roles and responsibilities [7].

Only 7.1% of the hospitals require their staff to sign a confidentiality agreement when employment starts. Prior-to-employment personnel screening for criminal acts is applied in 12.6% of the cases.

In Greece, there is no legislation specifically mandating the protection of health data, as is the case, for example of the Health Information Privacy and Accountability Act (HIPAA) in the U.S. [12]. Instead, as in other EU-member countries, such protection is mandated by a generic law on the protection of personal data. In the case of Greece this is law 2472 of 1997 that transposed the European directive on the protection of personal data [13] into the Greek legal order. According to this law, the person responsible for processing health data may be sanctioned if such data are inadvertently or purposefully disclosed to unauthorized recipients. Sanctions are also foreseen for the organization hosting the data (in our case the hospital).

Organizational sanctions in the form of disciplinary action for those employees violating the security policy exist in only 6.1% of the cases; therefore, in practice, abiding by the security policy is mostly on a voluntary basis.

Only 16% of the respondents report that a procedure ensuring non repudiation of staff actions exists.

Security standards also require assigning security roles and responsibilities, regardless of the establishment of an ITSO role [7-9]; 21% of the hospitals have done so.

24.4% of the hospitals intend to require external providers with whom exchange of information takes place to be security certified.

### 3.6 Physical and Environmental Security

Physical access to information systems facilities is one of the main areas of security concern. In order to prevent unauthorized physical access, damage and interference to the organization's premises and information, critical or sensitive information processing facilities should be housed in secure areas, protected by defined security perimeters, with appropriate security barriers and entry controls [7].

When asked whether information systems are located in controlled –from a physical access viewpoint- spaces, 79% of the respondents responded affirmatively. Still, a substantial percentage of hospitals keep their systems in mostly uncontrolled spaces. Clearly, allowing unauthorized internal or even external to the organization persons to physically access information systems is a risky practice that can lead to several security breaches, including disclosure of sensitive personal data or damage to equipment.

In 41% of the hospitals, spaces with special security specifications (e.g. server rooms) exist. However, procedures for controlling access to these spaces exist in only 19% of the cases, even though more than half of the hospitals acknowledge the need for such procedures.

12% of the respondents report that information processing equipment is being used outside the premises; this is usually mobile equipment (e.g. notebooks). A formal policy for such use is in place in only 17.2% of the cases. A cross examination of the answers to these responses reveals that most (75.8%) of the hospitals where processing outside premises is taking place do not have an appropriate policy in place.

32% of the hospitals have controls in place for ensuring the secure disposal and/or replacement of storage media.

### 3.7 Communications and Operations Management

To ensure the correct and secure operation of information processing facilities, responsibilities and procedures for the management and operation of all information processing facilities should be established [7].

A percentage of 56% is recorded for those hospitals that have a policy for using the network resources of the hospital that prohibits the personal use of such resources for purposes other than those within the employees' duties. Even worse is the situation with policies on the use of the internet and of the e-mail service: only 39% of the respondents have such a policy in place. It is interesting to notice that all hospitals that have an ITSO also have a security policy and a policy for the use of network resources in place. Additionally, seven out of ten of these hospitals have a policy for the use of the internet and of the e-mail service.

On the other hand, 55.6% of the respondents have set procedures for using the system control tools; 27.3% of the remaining ones do not have a security policy in place. However, 17.2% have not set such procedures, even though a security policy exists. 36.4% use authorization procedures for accessing system documentation; this reveals an often neglected security issue. Most of these hospitals have included pertinent controls in their security policy. Similar findings exist for hospitals having established controls for the correct use of information systems and installations, e.g. on conditions for allowing (or completely disallowing) the local installation of small applications by the users themselves on selected workstations.

Daily activity logs of IT staff are maintained in 15% of the cases, whereas the percentage in case of a security event is 17%.

87% of the hospitals report the use of antiviral software; this compares favorably with the reported percentage of 67% over all enterprises [14]. 55% of these hospitals also check externally developed or procured software for side channels and Trojan horses.

37.4% of the respondents report the existence of specialized software for checking system files for corruptions. On the other hand, 59 hospitals do not keep regularly backup copies of their system software. Overall then, the majority of hospitals that do keep backups, do not check these for corruptions. This finding is particularly useful when assessing the capacity of the hospitals to recover following a disaster.

43.4% of the hospitals have established pre-specified procedures for checking the security of their networks; the respective percentage among those hospitals that have an ITSO is 90%.

56% of the hospitals monitor the use of their information systems to avoid unused open connections, whereas 70% monitor their systems to identify future bandwidth needs.

Clock synchronization for auditing is used in 54.5% of the cases. Clear desk and screen policies exist in only 20% of the cases, even though this has been identified as one of the three major security issues among clinical personnel [11].

### 3.8 Access Control

Access to information, information processing facilities, and business processes should be controlled on the basis of business and security requirements [7].

Set procedures for assigning user rights exist in 76% of the cases; 16% of the remaining cases are small hospitals with less than 150 users. Unfortunately, only 24.5% of the respondents report the existence of such procedures to remote users of organizations affiliated to but nevertheless discrete from the main hospital (e.g. health centers). 39% of those hospitals having a security policy in place have not set such procedures.

71.4% of the respondents report that the hospital requires of its suppliers certified conformance to quality standards. However, only 14.1% have an established policy requiring the inclusion of access controls to its systems in outsourced service contracts.

Only 49% of the hospitals have an established password policy; this percentage rises to 80% among hospitals with an ITSO.

Only 48% of the hospitals have a user access control policy in place, either standalone or as part of the organizational security policy.

### 3.9 Information Systems Acquisition, Development and Maintenance

The design and implementation of the information system supporting the business process can be crucial for security; hence, security requirements should be identified and agreed prior to the development and/or implementation of information systems [7].

In the case of Greece, public hospitals' information systems have been designed, developed and implemented centrally, with little interaction between the developers and the hospital itself. As such, security requirements have not been set by the hospitals themselves.

Only 19% of the hospitals use cryptography, whereas no hospital has established key management procedures; this is equally true even among those with an existing security policy and those with an ITSO. Digital signatures are used only in 4% of the cases.

### 3.10 Information Security Incident Management

To ensure information events and weaknesses associated with information systems are communicated in a manner allowing timely corrective action to be taken, formal event reporting and escalation procedures should be in place [7].

A percentage of 15% of affirmative responses is recorded to the question whether procedures have been established and roles and responsibilities have been assigned for handling security incidents. A little better is the situation with assessing security reports received by the users, where 25% of the respondents report the existence of pertinent procedures. However, procedures for reporting security events or vulnerabilities exist in only 21% of the cases; this percentage rises to 90% among hospitals having an ITSO. Similar findings have been reported in [5] in Swiss hospitals.

### 3.11 Business Continuity Management

A business continuity management process should be implemented to minimize the impact on the organization and recover from loss of information assets to an acceptable level [7].

In our case, only 33% of the responses indicate that the term "disaster" has been defined by the hospital. The same number of hospitals report that they have formulated a disaster recovery plan. However, both (i.e. definition of the term and existence of plan) appear in only 24% of the cases. This is considerably less than the percentage (79%) reported in [15], [16] for 2007 and 2010 respectively over enterprises in all sectors of the economy.

### 3.12 Compliance

Breaches of any law, statutory, regulatory or contractual obligations and of any security requirements should be avoided [7].

We find that 76.8% of all respondents have not established secure information storage and processing procedures. Of the remaining respondents, 18.2% have a security policy in place.

Controls for protecting personal data exist in 46.5% of the cases; of these, 32.3% do have a security policy in place.

19% of the hospitals reports full knowledge and recording of laws and regulations pertinent to the operation of their information system. Less than half (46%) of the hospitals maintain contact with the Authority for the Protection of Personal Data, an independent agency competent by law to oversee all issues relevant to data protection, including those being processed by health care providers. Therefore, awareness on developments regarding the pertinent legislation is limited.

A commonly used approach to achieve an acceptable level of security in an organization is to certify its security-related policies and practices against pertinent international standards. Such certification also enhances the reputation of the organization and consequently positions it better with respect to its competitors. Additionally, security certification is known to assist in developing and maintaining a security culture within the organization. Only 14.4% of Greek hospitals report their intention to certify the security of their information systems against international standards, ISO 27799:2008 being the most preferable one. However, even though determining the aim of the certification is known to be one of the crucial issues to resolve before engaging in a certification exercise, only 12.6% of the hospitals have done so.

## 4 Conclusions

Hospital Information Systems largely process sensitive personal information. Protecting the security of this information needs to be based on a security plan that has been formulated following a risk analysis and management exercise in the organization. This is not the case in the vast majority of Greek hospitals, where

security policies have been formulated largely ad hoc and organizational security policies exist in slightly more than half of the hospitals. This might have been very different if more hospitals had established the ITSO role; this is the case only in very few hospitals.

The overall picture emerging from surveying the security management practices of information systems operating in Greek hospitals is one with serious security gaps, which are accentuated even more by the fact that existing policies are sometimes not enforced and/or enforcement is left at the discretion of the individual members of staff. Very few hospitals make their employees aware of security issues and train their staff accordingly, even though they may have a security policy in place. However, it is fair to say that the situation as compared to that of hospitals in other countries is not significantly different, with the exception of particular areas; it seems that information security still remains an issue among healthcare organizations everywhere.

It is clear that there is much room for improving the situation. A good starting point that could immediately make significant difference in the measured levels of security would be the establishment of a strict legal or regulatory requirement to address security issues in healthcare in particular, rather than including this into the generic legal requirements for the protection of personal data.

When modern technologies like mobile devices, body sensors connected through WSNs to the hospital network, cloud computing processing and visualization will be incorporated into the everyday life of a hospital, issues related to the security and privacy of health data will become even more pronounced and important, particularly because hospitals are part of a country's critical infrastructure. Hence, the need to keep track of the status of hospital information systems with regards to security and privacy issues becomes apparent. This can be, at least partly, achieved by surveys similar to the one herein.

## References

1. Transparency market research, http://www.transparencymarketresearch.com/healthcare-cloud-computing.html
2. Frost & Sullivan, http://www.frost.com/prod/servlet/press-release.pag?docid=267265445
3. Tountas, Y.: Economics of Health: Comparative Analysis of the Health Systems of Ten Developed Countries. Archives of Hellenic Medicine 20(1), 76–87 (2003)
4. Smith, E., et al.: Managing Health Information During Disasters: A Survey of Current Specialized Health Information Systems in Victorian Hospitals. Health Information Management Journal 36(1), 23–29 (2007)
5. Landolt, S., et al.: Assessing and Comparing Information Security in Swiss Hospitals. Interactive J. of Medical Research (i-JMR) 1(2), 11 (2012)
6. Bandyopadhyay, K.: Disaster –preparedness of Health Maintenance Organizations. Disaster Prevention and Management 11(4), 289–298 (2002)
7. ISO/IEC 27002 Information Technology – Security techniques – Code of practice for information security management (2005)
8. ISO/IEC 27799 Health Informatics – Information security management in health using ISO/IEC 27002 (2008)

9. ISO/IEC 27005 – Information technology – Security techniques – Information security risk management (2011)
10. CRAMM, Siemens Enterprise, http://www.cramm.com
11. Fernando, J., Dawson, L.L.: The Health Information System Security Threat Lifecycle: An Informatics Theory. Int. J. of Med. Inf. 78(12), 815–826 (2009)
12. US Department of Health and Human Services, http://www.hhs.gov/ocr/hipaa
13. European Commission, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:NOT
14. Geeknet. Enterprise Antivirus Security Survey. Bitpipe.com p. 6 (2012)
15. Balaouras, S.: The State of BC/DR Preparedness. Disaster Recovery J. 21(1), 14–22 (2008)
16. Dines, R.: The State of Disaster Recovery Preparedness. Disaster Recovery J. 22(1), 16–26 (2009)

# Risk Acceptance and Rejection for Threat and Opportunity Risks in Conflicting Incentives Risk Analysis

Lisa Rajbhandari and Einar Snekkenes

Norwegian Information Security Laboratory, Gjøvik University College, Norway
{firstname.lastname}@hig.no

**Abstract.** Classical methods for risk analysis usually rely on probability estimates that are sometimes difficult to verify. In particular, this is the case when the system in question is non-stationary or does not have a history for which reliable statistics is available. These methods focus on risks in relation to threats failing to consider risks in relation to opportunity. The Conflicting Incentives Risk Analysis (CIRA) addresses both these issues. Previously, CIRA has been investigated in analyzing threat risks. The paper contributes by illustrating the concept of opportunity risk in the context of CIRA. We give some theoretical underpinnings of risk acceptance and rejection of CIRA, addressing both risks. Furthermore, the paper explains the extension of CIRA to risk management by outlining the risk treatment (response) measures for threat (opportunity) risks.

**Keywords:** threat risk, opportunity risk, risk acceptance, risk rejection, risk analysis.

## 1 Introduction

The Conflicting Incentives Risk Analysis (CIRA) method provides an alternative notion of risk. That is, risk is specified in terms of conflicting incentives between the stakeholders (the risk owner and the strategy owner(s)) in regards to the execution of actions. The risk owner is the stakeholder whose perspective we consider when performing the risk analysis, i.e., he is the stakeholder at risk. The strategy owner is the stakeholder who is capable of triggering an action to increase his perceived benefit. For e.g., when analyzing risks to an end-user of a social networking service, the risk owner is the end-user while the strategy owners can be system administrator, hacker, etc.

Risk is the subjective concern that an individual can feel towards the outcome of events. Taking this perspective, we have two kinds of risks: concern that something undesirable might happen and concern that something desirable might not happen. In the following, we use the term threat (opportunity) to refer to the former (latter). To date, CIRA has been used in analyzing threat risks [9], [10]. In [10], CIRA is used for analyzing privacy risks faced by an end-user

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 124–136, 2013.

in a fictitious case study of an identity management system. Threat risks are caused by intentional execution of strategies by the strategy owner which results in gain for himself and loss for the risk owner. However, the use of CIRA for opportunity risk has not been investigated yet. Opportunity risks are caused by the strategy owners' potential failure to trigger a strategy that the risk owner could reasonably expect that he should trigger.

Most of the works on information security risks analysis and management such as NIST 800-30 [11], ISO/IEC 27005:2008 [7] and CORAS [3] consider only threat risks. This view is backed by researchers in [4], [12], [8]. Hillson [4] agrees that most of the classical risk management methods consider threats while the opportunities are ignored or addressed only reactively. In [8], Olsson also provides the evidence that the current risk management approaches focus on risk rather than opportunity. In economics, one tends to include opportunity risk. We agree with the economic perspective. For instance, we think that the risk that members of staff may fail to take advantage of new security technologies is a risk that must be included in the Chief Information Security Officer's bag of concerns.

To our knowledge, no works have been published addressing how risk acceptance and rejection criteria can be captured and analyzed in the context of CIRA. This paper gives some theoretical underpinnings of risk acceptance and rejection of CIRA method, addressing both threat and opportunity risks. In particular, it highlights and goes some way towards resolving a serious limitation present in other works on risk management- identification and management of opportunity risk in the context of information security management. Furthermore, the paper explains the extension of CIRA to risk management by outlining the risk treatment (response) measures for threat (opportunity) risks.

The remainder of the paper is organized as follows. In Sect. 2, we discuss the related work followed by the overview of CIRA. We explain the threat and opportunity risks in the context of CIRA in Sect. 4 and the details on computing the risk acceptance and rejection bounds are provided in Sect. 5. Sect. 6 outlines the risk treatment (response) measures for threat (opportunity) risks. In Sect. 7, we discuss some issues for further research. Finally, we conclude the paper in Sect. 8.

## 2 Related Work

In classical risk management, risk is often calculated as a combination of the likelihood of an incident and its consequence. The events are usually associated with having adverse/ unfavorable effect. This is further endorsed by the definition: *"risk is a function of the likelihood of a given threat-source's exercising a particular potential vulnerability, and the resulting impact of that adverse event on the organization"* [11]. Thus, typically most of the risk analysis and risk management approaches such as NIST 800-30 [11], ISO/IEC 27005:2008 [7], CORAS [3], OCTAVE [1] and RAMCAP [2] focus on threat risks. At the top level, the risks sources are categorized into human and non-human. Human threat sources

further include both the intended and unintended actions of human beings. The non-human threat sources consist of natural threats (e.g. flood, earthquake) and environmental threats (e.g. power failure, system failure). In NIST 800-30, risk from the given threats sources are considered: human (unintentional or deliberate actions), natural and environmental [11]. The ISO/IEC 27005:2008 standard also categorizes the origin of threat into accidental, deliberate and environmental (natural) [7].

The usage of the terms risk and opportunity varies among the risk analyst/ researchers. Some view risk as a term capturing both opportunity and threat [4]. On the other hand, some view opportunity as the opposite of risk [13], [8]. The latter view is usually captured by the term uncertainty; risk is defined as the uncertainty with negative consequences while opportunity is defined as the uncertainty with positive consequences. Even though the view on the definition of opportunity differs, most of the researchers agree that opportunity should be considered, whether being integrated into risk management [4], transforming risk management to uncertainty management [12] or as a separate field which is referred to as opportunity management [13], [8]. This issue has been emphasized mainly in the field of project management.

Hillson [4] explains how an existing risk management method can be extended to incorporate both threats and opportunities by: (1) adding new identification ways to effectively identify opportunities, (2) using double probability-impact matrix for representing both risks and (3) incorporating new strategies to respond to opportunities which are exploit, share, enhance and ignore. Ward et al. [12] argue that both threats and opportunities should be managed and proposes to transform the current project risk management processes into project uncertainty management. Further, White [13] suggests that at the enterprise level, more attention should be given to opportunity management than risk management. The Risk IT [5] framework looks at both IT risk and opportunity in an enterprise. The opportunity is concerned with the benefits that can be achieved (for e.g. identifying new business opportunities from using IT). In ISO 31000 [6], risk is defined as the "effect of uncertainty on objectives" whether positive or negative. Thus, the guideline can be used to determine risks having both positive or negative consequences.

## 3 Overview of CIRA

In this section, we provide an overview of CIRA explaining the terms, concepts and procedure. CIRA identifies stakeholders, their actions and perceived expected consequences that characterize the risk situation. As mentioned before, there are two classes of stakeholders: the strategy owner and the risk owner. Typically, each stakeholder has associated a collection of actions that he owns. Risk is modeled in terms of conflicting incentives between the risk owner and the strategy owners in regards to the execution of actions. The actions of the strategy owners may cause threat/ opportunity risks to the risk owner.

Human related risks is the focus of CIRA. This corresponds to understanding the human behavior and incentives that influence their actions. An incentive is something that motivates a stakeholder to take an action to increase his expected/ predicted utility. Utility is the benefit as perceived by the corresponding stakeholder and it comprises of utility factors. Each utility factor captures a specific aspect of utility e.g. prospect of wealth, reputation, legal compliance, ego. Thus, utility can be approximated as the sum of weighted values for utility factors using Multi Criteria Decision Analysis.

When we (the risk analysts) interact with the stakeholders, we assume that their preferences have been determined according to their thinking/ perception about the available options and what is good for them with respect to how others choose their actions. Thus, game theory comes into play in our model as it helps us understand these strategic settings that influence the stakeholders' behavior. However, we do not employ game theoretic modeling (i.e. constructing models of strategic settings) and computations.

People consider their self-interest when they interact in a strategic setting and this often gives rise to conflicts. However, we cannot ignore the fact that individuals acting in their self interest sometimes do cooperate to maximize the benefit of all involved. In our setting, it is assumed the stakeholders choose their utility factors according to their self interest and in turn their strategies/ actions to maximize their utility. In other words, the stakeholder will make the move that he thinks will give him the highest gratification.

When identifying and modeling strategy owner actions, the actions modeled correspond to the first move of a game and the effect of the move is modeled as the value of the game as perceived by the strategy owner and risk owner. Note that the strategy owner may be playing a game with multiple stakeholders. For e.g., an attacker will play a game with law enforcement and the legal system. However, we will model this as a strategy that modifies the utility factors of the strategy owner and the risk owner taking into account how the attacker perceives the outcome of the attack including the uncertainty relating to his capture and prospect of penalty. Thus, we do not engage in game theoretic computations of stakeholder behavior but rely on data collection to capture expectations relating to strategy outcomes. The procedure in CIRA is divided into: structural data collection phase (1-6), numerical data collection phase (7-9) and analysis phase (10-13) as depicted in Fig. 1.

## 4 Explaining Risk in the Context of CIRA

In this section, we explain the risks caused by intentional execution of strategies by the strategy owner: threat risk and opportunity risk in the context of CIRA. In classical risk analysis, risk is usually determined as the combination of likelihood and consequence resulting in the unit of $Ut^{-1}$, where U represents utility and t represents time. On the other hand, in CIRA, risk is the result of conflicting incentives and its unit is $U^2$.

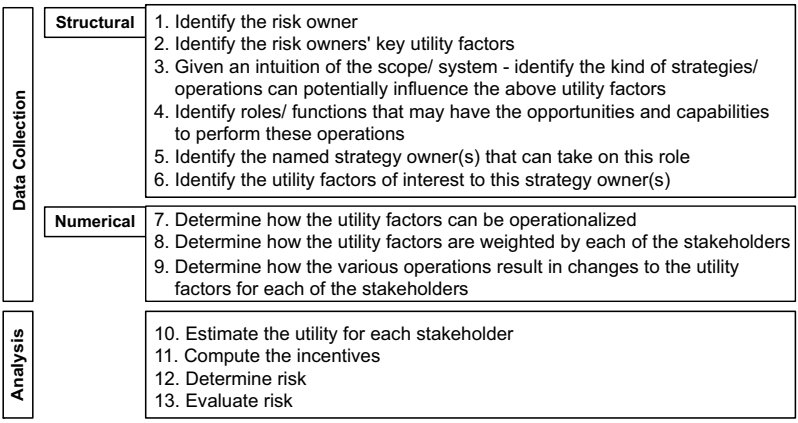



**Fig. 1.** Procedure in CIRA

### 4.1 Risk Visualization

In CIRA, we visualize risk using an incentive graph which is a simple 2 axis coordinate system corresponding to the set of incentives $(s, r)$, where $s(r)$ corresponds to the incentive of the strategy owner (risk taker) as shown in Fig. 2. Note that all events above (below) the $X$-axis belong to the collection of opportunity (threat) events. These concepts are described in Table 1. The graphs defined by $R_O, A_O, A_T, R_T, X$-axis in Fig. 2 partitions the risk plane into 6 non-overlapping areas as described in Table 2.

### 4.2 The Threat Risk

Previously, CIRA [9], [10] has been restricted to analyzing threat risks i.e. risk facing the risk owner caused by the intentional execution of strategies by the strategy owner which results in gain for himself and loss for the risk owner. The idea being that, risk is the combination of the strength of the force that motivates the strategy owner to send the risk owner to an undesirable state and the magnitude of this undesirability. These risks are usually the consequence of some personal motivations of the strategy owner such as gaining wealth, status, free time, etc. It is reasonable to make the assumption that the strategy owner will be rational in a behavioral economic sense. For e.g., for an end-user of a social networking service, there is uncertainty to the protection of his privacy as his information could be exploited for secondary purposes. In this case, he is facing a threat risk.

### 4.3 The Opportunity Risk

Opportunity risk is the concern that something desirable might not happen. The risk owner is facing opportunity risk when he is concerned that the strategy

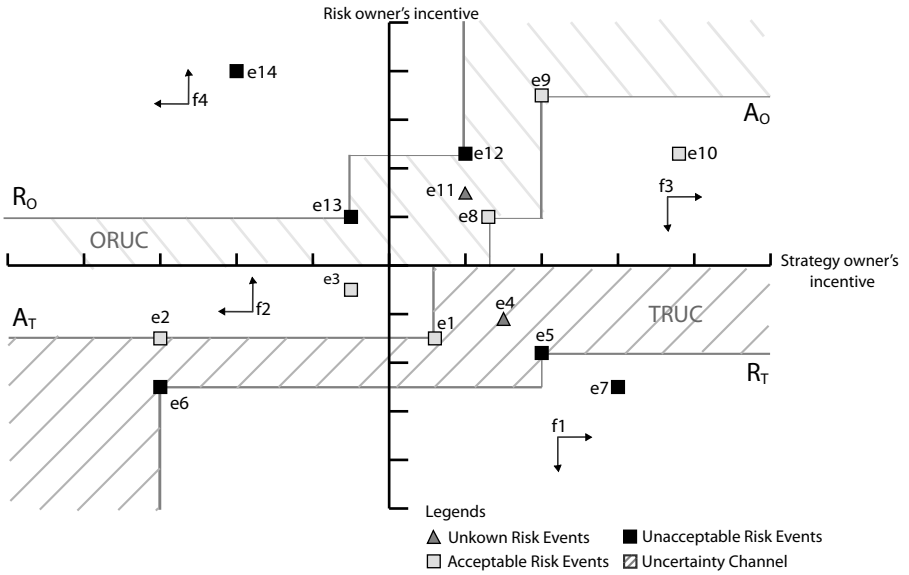



**Fig. 2.** CIRA risk visualization using the incentive graph

**Table 1.** Legend for CIRA Risk Visualization

| | |
|---|---|
| **For Threat Risk** | |
| $R_T$ | Rejection boundary of threat risks |
| $A_T$ | Acceptance boundary of threat risks |
| $e_1$ - $e_3$ | Outcome of a typical threat event having an acceptable risk |
| $e_4$ | Outcome of a typical threat event for which it is not known if the risk is acceptable or not |
| $e_5$ - $e_7$ | Outcome of a typical threat event having an unacceptable risk |
| $f_1$ | Rejection rationality for threat risks |
| $f_2$ | Acceptance rationality for threat risks |
| **For Opportunity Risk** | |
| $R_O$ | Rejection boundary of opportunity risks |
| $A_O$ | Acceptance boundary of opportunity risks |
| $e_8$ - $e_{10}$ | Outcome of a typical opportunity event having an acceptable risk |
| $e_{11}$ | Outcome of a typical opportunity event for which it is not known if the risk is acceptable or not |
| $e_{12}$ - $e_{14}$ | Outcome of a typical opportunity event having an unacceptable risk |
| $f_3$ | Acceptance rationality for opportunity risks |
| $f_4$ | Rejection rationality for opportunity risks |

**Table 2.** CIRA Plane Partition Legend

| **Area bounding** | **Explanation** |
|---|---|
| $A_O$ and $X$-axis | Acceptable risk from opportunity events ($e_{10}$) |
| Left and above $R_O$ | Unacceptable risks from opportunity events ($e_{14}$) |
| $A_O$ and $R_O$ | Information is lacking with respect to the acceptability of the risk of these opportunity events ($e_{11}$). It is called as the Opportunity Risk Uncertainty Channel (ORUC). |
| $A_T$ and $X$-axis | Acceptable risk from threat events ($e_3$) |
| Below and to the right of $R_T$ | Unacceptable risk from threat events ($e_7$) |
| $A_T$ and $R_T$ | Information is lacking with respect to the acceptability of the risk of these threat events ($e_4$). It is called as the Threat Risk Uncertainty Channel (TRUC). |

owner may fail to trigger a strategy that he could reasonably expect that the strategy owner should trigger. The reason being that the strategy owner would have to take a loss in utility and the risk owner would have the prospect of a gain. Likewise, in the threat risk case, it is reasonable to make the assumption that the strategy owner will be rational in a behavioral economic sense. For e.g., if there is uncertainty as to the willingness of a member of staff to spend some effort to identify and deploy more cost effective security products (while maintaining its security posture), the organization is facing opportunity risk.

## 5 Computing Risk Acceptance and Rejection Bounds

Assume we have a collection point sets $P = 2^{U \times U}$, where $U$ denotes the set of utilities. We select a set $D \in P$ of (Incentive, Consequence) pairs and for each of these pairs, we ask the risk owner if he finds the risk associated with this pair acceptable or unacceptable (reject). A risk pair with a negative (positive) consequence is referred to as a threat (opportunity) risk. Then we define $D_{AT}, D_{RT}, D_{AO}, D_{RO} \subseteq D$ such that $D_{AT}$ denotes the set of threat risks to be accepted, $D_{RT}$ the set of threat risks to be rejected, $D_{AO}$ the set of opportunity risks to be accepted and $D_{RO}$ the set of opportunity risks to be rejected. Thus, the risk owner partitions $D$ into the disjoint subsets $D_{AT}, D_{RT}, D_{AO}, D_{RO}$. We require that $D_{AT}$, $D_{RT}$, $D_{AO}$, $D_{RO}$ are non-empty.

We stipulate that the risk owner is rational in the following sense: Let $i, j, d$ be any positive utilities and $c$ be negative for threat consequences and positive for opportunity consequences, then

R1 If a threat consequence $c$ is accepted by the risk owner for a given strategy owner incentive $i$, then the consequence $c + d$ is acceptable for the incentive $i - j$.

R2 If a threat consequence $c$ is unacceptable (i.e. rejected) by the risk owner for a given strategy owner incentive $i$, then the consequence $c - d$ is unacceptable for the incentive $i + j$.

R3 If an opportunity consequence $c$ is accepted by the risk owner for a given strategy owner incentive $i$, then the consequence $c - d$ is acceptable for the incentive $i + j$.

R4 If an opportunity consequence $c$ is unacceptable (i.e. rejected) by the risk owner for a given strategy owner incentive $i$, then the consequence $c + d$ is unacceptable for the incentive $i - j$.

For example, in the case of opportunity risks, if the gain is relatively modest, you may be prepared to forfeit the gain generated by the strategy. However, if you stand to gain a lot, you will be less inclined to accept the possibility that you may not receive the benefit. Thus, you will require that the strategy owner has strong incentives to implement strategies that would give you large benefits.

The functions defined below compute bounds for acceptance and rejection for both opportunity and threat risks under the assumption that the risk owner is rational in the above sense.

**Definition 1.** *Risk acceptance and rejection bounds*

*The lower bound of consequences for threat risks to be accepted, specified as a function of strategy owner incentive is:*

$$A_T(x) = Min(\{y' | \exists x' \cdot (x', y') \in D_{AT} \wedge x \leq x'\})$$

*The upper bound of consequences for threat risks to be rejected, specified as a function of strategy owner incentive is:*

$$R_T(x) = Max(\{y' | \exists x' \cdot (x', y') \in D_{RT} \wedge x' \leq x\})$$

*The upper bound of consequences for opportunity risks to be accepted, specified as a function of strategy owner incentive is:*

$$A_O(x) = Max(\{y' | \exists x' \cdot (x', y') \in D_{AO} \wedge x' \leq x\})$$

*The lower bound of consequences for opportunity risks to be rejected, specified as a function of strategy owner incentive is:*

$$R_O(x) = Min(\{y' | \exists x' \cdot (x', y') \in D_{RO} \wedge x \leq x'\})$$

We can then define the rationality closures for the risk acceptance and rejection sets as follows:

**Definition 2.** *Risk acceptance and rejection rationality closures*

$$A_T^c = \{(i, c) | c \geq A_T(i)\} \quad \textit{All acceptable threat risks (R1)}$$
$$R_T^c = \{(i, c) | c \leq R_T(i)\} \quad \textit{All unacceptable threat risks (R2)}$$
$$A_O^c = \{(i, c) | c \leq A_O(i)\} \quad \textit{All acceptable opportunity risks (R3)}$$
$$R_O^c = \{(i, c) | c \geq R_O(i)\} \quad \textit{All unacceptable opportunity risks (R4)}$$

**Theorem 1.** *All elements in $D_{AT}, D_{RT}, D_{AO}, D_{RO}$ belong to the corresponding rationality closures, i.e.*

$$D_{AT} \subseteq A_T^c$$
$$D_{RT} \subseteq R_T^c$$
$$D_{AO} \subseteq A_O^c$$
$$D_{RO} \subseteq R_O^c$$

*Proof.* By expansion and noting that for any closed boolean expressions $P(.)$, $a$ and $b$:

$$P(a) \Rightarrow a \leq \text{Max}(\{y|P(y)\})$$
$$P(b) \Rightarrow b \geq \text{Min}(\{y|P(y)\})$$

**Theorem 2.** *The rationality closures extends the acceptance and rejection bounds. I.e. for all $i, j, d \geq 0$ then*

$$c \leq 0 \wedge (i, c) \in A_T^c \Rightarrow (i - j, c + d) \in A_T^c \text{ (R1)}$$
$$c \leq 0 \wedge (i, c) \in R_T^c \Rightarrow (i + j, c - d) \in R_T^c \text{ (R2)}$$
$$c \geq 0 \wedge (i, c) \in A_O^c \Rightarrow (i + j, c - d) \in A_O^c \text{ (R3)}$$
$$c \geq 0 \wedge (i, c) \in R_O^c \Rightarrow (i - j, c + d) \in R_O^c \text{ (R4)}$$

*Proof.* By expansion and noting that $a + b \geq a$ for $b \geq 0$, $e - f \leq e$ for $f \geq 0$ and for any closed boolean expression $P(.)$:

$$\text{Min}(\{y|\exists x \cdot P(x, y) \wedge (i - j) \leq x\}) \leq \text{Min}(\{y|\exists x \cdot P(x, y) \wedge i \leq x\})$$
$$\text{Max}(\{y|\exists x \cdot P(x, y) \wedge x \leq i\}) \leq \text{Max}(\{y|\exists x \cdot P(x, y) \wedge x \leq i + j\})$$

when $j \geq 0$.

The acceptance and rejection sets are mutually consistent for threats (opportunities) iff their rational extensions are non-overlapping. Given a risk acceptance (rejection) set $A$ ($R$), we define opportunity (OC(.)) and threat (TC(.)) risk acceptance/rejection consistency as

**Definition 3.** *Risk acceptance and rejection consistency*

$$OC(R, A) = TC(A, R) = \forall i, j, c, d \cdot (i, c) \in A \wedge (j, d) \in R \Rightarrow i < j \vee c > d$$

**Theorem 3.** *The rationality closure is consistency preserving. I.e.*

$$TC(A_T, R_T) \Rightarrow TC(A_T^c, R_T^c)$$
$$OC(A_O, R_O) \Rightarrow OC(A_O^c, R_O^c)$$

*Proof.* Since the acceptance and rejection sets are mutually consistent, noting that there is a transitivity property between position of the point in $A_T$, $R_T$ and the point in the closure, it suffices to show that each element in the accept (reject) closure is 'on the correct side' of some point in the corresponding partition of $D$. But this holds by Lemma 1.

**Lemma 1.** *All elements in a closure are bounded by some element in the corresponding partition of D:*

$$\forall i, c \cdot (i, c) \in A_T^C \Rightarrow \exists j, d \cdot (j, d) \in D_{AT} \wedge j \geq i \wedge d \leq c$$
$$\forall i, c \cdot (i, c) \in R_T^C \Rightarrow \exists j, d \cdot (j, d) \in D_{RT} \wedge j \leq i \wedge d \geq c$$
$$\forall i, c \cdot (i, c) \in A_O^C \Rightarrow \exists j, d \cdot (j, d) \in D_{AO} \wedge j \leq i \wedge d \geq c$$
$$\forall i, c \cdot (i, c) \in R_O^C \Rightarrow \exists j, d \cdot (j, d) \in D_{RO} \wedge j \geq i \wedge d \leq c$$

*Proof.* By expansion, using the existential witness obtained from the antecedent, we can easily construct the existential witness required in the consequent. Noting that $\forall x \cdot x \leq max(y|P(y)) \Rightarrow \exists z \cdot x \leq z \wedge P(z)$ and $\forall x \cdot x \geq min(y|P(y)) \Rightarrow \exists z \cdot x \geq z \wedge P(z)$.

We can easily restrict acceptance and rejection closure sets to the corresponding consistent subset as follows:

$$A_{TC}^c = A_T^c \setminus R_T^c$$
$$R_{TC}^c = R_T^c \setminus A_T^c$$
$$A_{OC}^c = A_O^c \setminus R_O^c$$
$$R_{OC}^c = R_O^c \setminus A_O^c$$

However, in practice, rather than restricting the acceptance and rejection sets, one would engage in a dialogue with the risk owner such as to ensure that the partitions of $D$ (i.e. $D_{AT}, D_{RT}, D_{AO}, D_{RO}$) are mutually consistent.

## 6 Risk Treatment (Response) Measures for Threat (Opportunity) Risks

In this section, we explain the risk treatment (response) measures for threat (opportunity) risks. The overall process for risk management in CIRA is depicted in Fig. 3. Risk analysis helps to identify and estimate risks, and provide insight suitable for deciding if risk exposure needs to be changed. That is, if a treatment/ response action is needed, or risk exposure may be increased. Further, risk management is taking actions to treat/ respond to those risks that are not within the risk acceptance criteria. The treatment measures for the threat risks include: mitigate, avoid, transfer and accept (i.e. accept the risk without taking any action). On the other hand, the response measures for the opportunity risks include enhance, exploit, share and ignore [4].

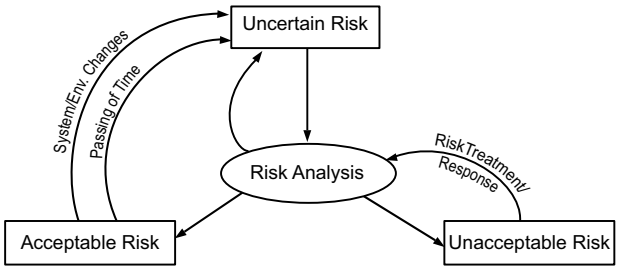



**Fig. 3.** CIRA Risk Exposure States and Transitions

The risk exposure is either acceptable, unacceptable or uncertain/ unknown. By doing risk analysis, we can determine if the exposure is acceptable or not. If the system or the environment changes, we may no longer have sufficient evidence to conclude that the exposure is acceptable. Similarly, over time, the system and its environment may be exposed to changes that are not easily discoverable. Thus, we are drifting towards a state of unknown exposure. When implementing a risk treatment/ response measure, this measure may only have a partial effect and it may also have side effects giving rise to new vulnerabilities. The risk analysis process further helps to decide the risk exposure.

In classical risk management, risk is managed through reduction of incident likelihood and consequence. Since CIRA does not adopt the likelihood and consequence paradigm, our risk management goals will be somewhat different. In CIRA, risk treatment/ response amounts to the modification of perceived utility caused by the strategies in questions. That is, a risk treatment/ response measure aims to modify the weights that the stakeholders assign to the relevant utility factors or modify the incentives of the stakeholders. This is illustrated in Fig. 4 and described in more detail below. Threat risks can be mitigated through a combination of strategy owner incentive reduction and risk owner incentive increase. In Fig. 4, the arrows in $m_1$ represent the desired direction that we want to move the outcome through mitigation. In the above threat risk example, risk faced by the end-user can be reduced if privacy rules and regulations (that govern when and how the services use/ collect personal information of customers) are established and enforced. In the case of opportunity risks, the primary risk response strategy is to increase the strategy owner incentives to implement the strategy (represented by the rightward arrow in $m_2$). We may also respond to the risk by reducing the utility of the risk owner (represented by the downward arrow in $m_2$). However, sometimes we may also want to increase the utility of the risk owner. In the above opportunity risk example, the staff can be incentivized by enforcing and communicating rules within the organization or providing free training so that they are willing to deploy more cost effective security products which in turn benefits the organization.

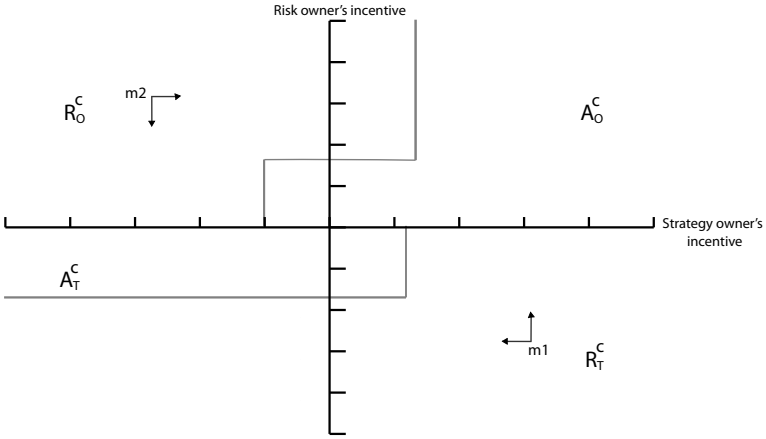



**Fig. 4.** CIRA Risk Management Strategies

## 7 Future Work

Some of the potential issues for further work include the exploration of the newly introduced theoretical concepts on opportunity risk, risk treatment and response measures. More case studies needs to be conducted to explore the different risks, and to validate and improve the method.

In CIRA, risk acceptance criteria is determined after the decision input parameters are determined unlike in most of the classical methods. For instance, in the ISO/IEC 27005:2008 standard, it is required that the risk acceptance criteria should be determined before the threat and vulnerability discovery is carried out. However, this may not be economically rational. For e.g., we may have a situation where a risk that is low is unacceptable because the mitigation effort required to control the risk is very low. Similarly, we may accept a very high risk if all response options have little or no effect. Thus, an economically rational actor will determine his risk acceptance threshold on at least the following information: incident risk, risk mitigation cost and the effectiveness of mitigation measures. Thus, it is questionable if it is a good strategy to fix the risk acceptance criteria as an expected value before the relevant decision input parameters have been determined. This issue can be further investigated.

## 8 Conclusion

This paper has explained the key concepts of risk acceptance and rejection in the CIRA method. Definitions have been formalized and we have included some theorems establishing some consequences of our definitions. Our definitions and model introduce the concept of opportunity risk in the context of information

security management. The opportunity risk remains overlooked and needs more emphasis by current research on risk management. A comprehensive Conflicting Incentives Risk Analysis and Management (CIRAM) method which considers and addresses both threat and opportunity risks has the potential to enhance the overall risk management process.

**Acknowledgement.** The work reported in this paper is part of the PETweb II project sponsored by The Research Council of Norway under grant 193030/S10.

## References

[1] Alberts, C., Dorofee, A.: Managing information security risks, The OCTAVE approach. Addison Wesley (2002) ISBN 0-321-11886-3
[2] ASME Innovative Technologies Institute, LLC. Risk Analysis and Management for Critical Asset Protection (RAMCAP): The Framework, Version 2.0 (May 2006)
[3] Braber, F., Hogganvik, I., Lund, M.S., Stølen, K., Vraalsen, F.: Model-based security analysis in seven steps — a guided tour to the CORAS method. BT Technology Journal 25(1), 101–117 (2007)
[4] Hillson, D.: Extending the risk process to manage opportunities. International Journal of Project Management 20(3), 235–240 (2002)
[5] ISACA. The Risk IT Framework (2009)
[6] ISO 31000. Risk Management – Principles and Guidelines. ISO (2009)
[7] ISO/IEC 27005. Information technology -Security techniques -Information security risk management. ISO/IEC, 1st edn. (2008)
[8] Olsson, R.: In search of opportunity management: Is the risk management process enough? International Journal of Project Management 25(8), 745–752 (2007)
[9] Rajbhandari, L., Snekkenes, E.: Intended Actions: Risk Is Conflicting Incentives. In: Gollmann, D., Freiling, F.C. (eds.) ISC 2012. LNCS, vol. 7483, pp. 370–386. Springer, Heidelberg (2012)
[10] Rajbhandari, L., Snekkenes, E.: Using the Conflicting Incentives Risk Analysis method. In: Janczewski, L.J., Wolf, H., Shenoi, S. (eds.) SEC 2013. IFIP AICT, vol. 405, pp. 315–329. Springer, Heidelberg (2013)
[11] Stoneburner, G., Goguen, A., Feringa, A.: NIST SP 800-30, Risk Management Guide for Information Technology. NIST (July 2002)
[12] Ward, S., Chapman, C.: Transforming project risk management into project uncertainty management. International Journal of Project Management 21(2), 97–105 (2003)
[13] White, B.E.: Enterprise Opportunity and Risk. In: INCOSE Symposium, Orlando, FL (July 2006)

# ProCAVE: Privacy-Preserving Collection and Authenticity Validation of Online Evidence

Efthymios Lalas[1], Lilian Mitrou[2,3], and Costas Lambrinoudakis[1]

[1] Department of Digital Systems,
University of Piraeus,
GR-18534, Greece
lalas@webmail.unipi.gr, clam@unipi.gr
[2] Department of Information & Communication Systems Engineering,
University of the Aegean,
GR-83200, Greece
l.mitrou@aegean.gr
[3] Department of Informatics,
Athens University of Economics and Business,
GR-10434, Greece

**Abstract.** It is an undisputable fact that nowadays many different types of crime are conducted by utilizing some type of electronic device - communication. To address this new situation, modern forensics tools evolved, becoming sophisticated enough to handle almost all kinds of digital content. However, surprisingly enough, collecting and validating the authenticity of online content remains, until now, a problem to resolve. The common practice is to capture (screen-shot) or save a web page, the authenticity of which is usually validated in a judicial process by an expert's testimony. In this paper, we introduce *ProCAVE* , a simple software architecture with a set of accompanying procedures, and we argue that their combined use can deliver evidence from online sources in the court, in a sound and privacy-preserving manner.

## 1 Introduction

The web today is used by billions of people, facilitating business, communication and exchange - dissemination of information. However, this new reality also has a "dark side" [5], since a series of crimes are committed through and with the use of it. Such crimes are known as: (a) *"Computer Crimes"* i.e. those posing a direct threat to data, information systems and networks, (b) *"Computer Related Crimes"*; i.e. crime perpetrated by means of a computer such as computer fraud, property rights infringements, and (c) *"Content-related Crime"*; such as child pornography, defamation via internet, etc.

In the legal system of most countries, guilt in a criminal proceeding is the summation of *means*, *motive* and *opportunity* [4]. However, the presence of those three elements is often not sufficient to convict someone; appropriate evidence must be presented that will prove that the opportunity was indeed taken by the accused for the crimes that he/she is charged (a U.S. example is stated in [4]).

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 137–148, 2013.

Nevertheless, legal systems are mostly tailored to traditional types of crime. Any information obtained from the Internet has to convince the truth of a deed, which means that it must have all the attributes of conventional evidence. It has to be "irrefutable authentic", i.e. it must be possible to positively tie evidentiary material to the incident [5] and collected in accordance with formal requirements to establish its reliability [13] and admissibility. As for conventional crime, in order to prove the guilt and convict a person for having committed an offense, it is required to recognize, recover, reconstruct and present the digital evidence in a way that it renders it admissible in legal proceedings [11].

World Wide Web has become not only a crime scene, but also a breeding ground for primary and secondary sources of evidence. However, internet evidence, as internet content, is ex natura ephemeral and volatile, since it can be easily altered or deleted. Even though it should be stressed that, in general, with the advent of European digital signature legislation (Directive 2000/31/EC on electronic commerce), electronic material has gained a comparable legal status as paper material.

Considering crimes in which online content plays a crucial role, the main challenge is its collection, preservation and admissible presentation. In order to achieve its goal, i.e. allow and assist the court to form and pass a judgment, this online content has to be properly collected and its authenticity validated.

To this end, a number of different procedures have been employed around the world. Perhaps the most common one is taking a screenshot of a web site, printing it and validating its authenticity in the court by a witness' testimony [2,6]. The same procedure is followed with web pages that have been saved using the "Save As" function of the browser, or with other means of web page downloading. Recently, there are some tools that have also been proposed, mainly for forensics purposes, that save online content locally and perform some hashing and/or apply the current timestamp to the downloaded content [26].

There are cases where the aforementioned procedure, based on a witness' testimony, is not enough [3]. Furthermore, it is possible the content in question to be removed, so an expert from a certain company (usually WebArchive) is being called to provide evidence of what a web page looked like at a certain point of time. When this is not feasible, then the owner of the web site in question is called to testify, in order to provide evidence of the online content [3].

It is well known among the people involved in judicial processes, that the approaches described above cause too many controversies and concerns which often result in non-admissibility of the digital evidence. Some examples are: "The witness altered/photoshoped the screenshot", "The witness changed the html code of the saved web page before signing it", "The web page was not publicly available at that time", "The site's administrator/owner is in another country and cannot testify". Such statements demonstrate that the key element for online content admissibility is the validation of its authenticity; as stated in [7], once the authenticity of the electronic evidence has been validated, all other evidentiary problems are the common problems lawyers face all the time.

This paper proposes a novel way for collecting online evidence and validating its authenticity so as to be acceptable for evidentiary purposes. Our approach is based on the notion of web proxy, which has never been used before in collecting and validating the authenticity of web site content. It is argued that, as soon as someone identifies online content related to a crime, he/she can pass the request through a web proxy (belonging to a trusted authority) that will "freeze" the content, apply current timestamp and signatures and deliver the evidence to the user (a natural person, an organization or a public authority). Moreover, the same request can be simultaneously performed by other web proxies in order for the evidence to be securely stored in more than one servers. To demonstrate the applicability of the proposed solution, *ProCAVE* has been implemented and tested extensively with various web sites.

The rest of the paper is organized as follows: Section 2 reviews the current legal and technological status regarding the collection and validation of authenticity of online content. An overview of the proposed software solution is presented in Section 3. The experimental evaluation and results are discussed in Section 4 whereas Section 5 refers to some technical and other considerations.

## 2 Current Status and Motivation

Evidence has been present in legal systems since the first trial in the human history. All countries have incorporated rules and procedures that are deemed to be appropriate and legally robust for validating the authenticity of evidence in the court. Judge Grimm [14] has codified the rules governing the validation of the authenticity of web page evidence under the US Law. According to this categorization the most common rules for web pages are the following:

- 901(b)(1): witness with personal knowledge,
- 901(b)(3): expert testimony,
- 901(b)(4): distinctive characteristics of a web site,
- 901(b)(7): public records - usually from government web sites,
- 901(b)(9): system or process capable of producing a reliable result, and
- 902(5): official publications.

The important part of validating the authenticity (and thus admissibility) of the evidence is that one cannot rely on a simple method, since the degree of foundation which is appropriate in any given case is in the judgment of the court [14]. Thus, if multiple methods are used it is more likely that a court will deem the online content as authentic [9]. This is the main reason that all of the solutions that are currently used fail to produce undisputable results.

First of all, the authenticity of printed copies or captured images of a web site has to be validated by the witness in order to be admitted [20], a procedure that many times is still questionable especially when the witness is not independent (e.g. is not a police officer). Moreover, the exact time that the specific content was accessible online cannot be proven, since standard time-signing techniques cannot be applied [10].

Certain commercial tools [6,26] claim that they support a sound mechanism for collecting online evidence; the time issue may be resolved, however once again the user needs to validate the authenticity of the evidence in the court. The cost is another drawback, and of course when someone comes across online content that needs to be collected for legal purposes, no one expects him to buy a software only for saving an instance of a page!

Expert testimony (as is usually the case with Web Archive) is always an option. However companies that archive web pages do not store dynamic or personalized content, and there are also cases that some content has been removed before the page was archived. Moreover, it is surely very difficult (nearly unfeasible) for experts from Web Archive to testify in another country.

The last option is to have the web site's owner testify about the web site's content at a certain date/time. But this is very rarely done and of course even if such a testimony exists, it is still necessary to employ third party tools to prove the existence of the content.

Another essential feature missing from all above online content preservation techniques, and which has partially motivated our work, is the notion of privacy. More specifically, as described in [11], the use of forensics methods may itself constitute a violation of citizens' fundamental right to privacy; that's why digital evidence must - among others - comply with the respective provisions guaranteeing data privacy.

Finally, as extensively documented over the past few years [8,11,16], forensics, in general, lack standardization of methods and formats, a fact that causes many procedural problems. In terms of collection of online content, this is impossible with existing techniques, since they are proprietary, diverse and depend on each user's technical knowledge regarding evidence acquisition.

The above-described needs for a privacy-preserving, standardized method of independent collection of online content nurtured the seeds of our current work. To this end, our paper makes the following contributions:

- The first privacy-preserving web-based tool for collecting evidence from web pages, namely *ProCAVE* , is presented. It is demonstrated that it fulfills all the aforementioned requirements.
- A prototype of *ProCAVE* has been implemented and together with a dataset of known web sites has been used for the evaluation of its effectiveness and accuracy.

## 3 Solution Overview

This Section unravels the logic behind the proposed *ProCAVE* software solution. To accomplish its goals, *ProCAVE* is practically comprised of two elements: the *Web Proxy* and the *Collection and Validation of Authenticity (CVA) Engine*.

### 3.1 Web Proxy

As soon as a user discovers a web site, e.g. *www.abc.com*, whose content can be used as evidence, he/she visits the web site *www.xyz.com*, which serves as the

*Web Proxy* of *ProCAVE*. There he/she is presented - among other options - with an address field, where he/she can enter the url that he/she wants to navigate to. He/she enters *www.abc.com* and the *Web Proxy* receives the request, forwards it to *www.abc.com* and receives and forwards back the result. This procedure is depicted in Figure 1.

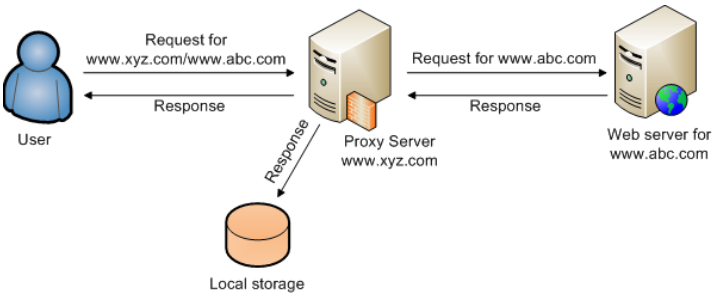



**Fig. 1.** Proxy HTTP request

Therefore, the user requests url *www.xyz.com\www.abc.com*, and he/she is presented with the contents of url *www.abc.com*. However, contrary to a conventional HTTP request, this specific request passes through the *ProCAVE Web Proxy* that keeps a copy of the response (i.e. HTML code, CSS, images, scripts etc.) locally. This means that whenever the user wishes, and provided that the content is the one that he/she wants to collect, he/she can proceed to the next step which is the *Collection and Validation of Authenticity* described next.

### 3.2 Collection and Validation of Authenticity (CVA)

Currently the user possesses content (a response), which is stored locally in the *Web Proxy*. He/she can now proceed with the validation of the authenticity of the stored content. This will be done, as shown in Figure 2, with the help of privacy, hashing and digital signing modules.

As soon as the user selects the CVA option, the id of the response is sent to the *Web Proxy*, along with some other parameters. These parameters reflect the privacy level that the user is requesting and correspondingly the confidentiality level for the collected content. To this end, if the user decides to use the privacy option, a blacklist/whitelist option is adopted; i.e. the user is allowed to choose some content and scramble all the other, or scramble some content and leave all the other intact. This scrambling is accomplished by sending the chosen HTML element ids, together with the selected option, to the *Web Proxy*, that modifies the content accordingly. The scrambling process engages the public key of the user, and is represented in Figure 2 with a dashed rectangular due to its optional use.

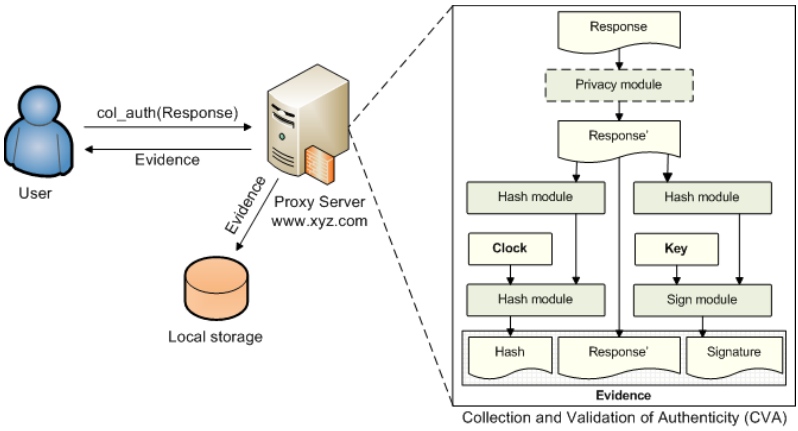



**Fig. 2.** Collection and Validation of Authenticity (CVA) Workflow

Next, a hashing of the (scrambled or unscrambled) content is performed, producing the "digest", which will be later utilized for the verification of the content. A timestamp (which has been produced at the time of the request) is concatenated to the digest, in order for the date and time to be bind to it, according to [10]. The concatenated "digest + timestamp" is fed once more to the hash module, producing the final hash value.

Up to now the proposed mechanisms ha maintained the original message while a message digest has been produced. In the final step of the CVA procedure, the original message is signed with the engine's private key, thus validating the authenticity of the evidence's creator, which is the engine itself. The final (signed) elements will be grouped together, forming the evidence of the online content. This evidence will be returned to the user, and it will be also stored locally for future reference.

### 3.3 Multiple Requests

As already discussed in previous sections, a single copy of digital evidence is not necessarily sufficient to prove a crime. It may be necessary to prove how the specific web site was visible to various locations around the world. Moreover, it may be necessary to store the evidence in various, geographically spread, locations.

The design of *ProCAVE* satisfies the above requirement. When a certain *Web Proxy* receives a request, it creates a copy of this request and forwards it to other Web Proxies. As a result, each one of these Web Proxies will collect and validate the authenticity of the *www.abc.com* web site's content, returning the resulting evidence. The returned evidence may be different from one *Web Proxy* to another, since the content of the same web site may differ from country to country. Nevertheless, the collection of online content from different locations

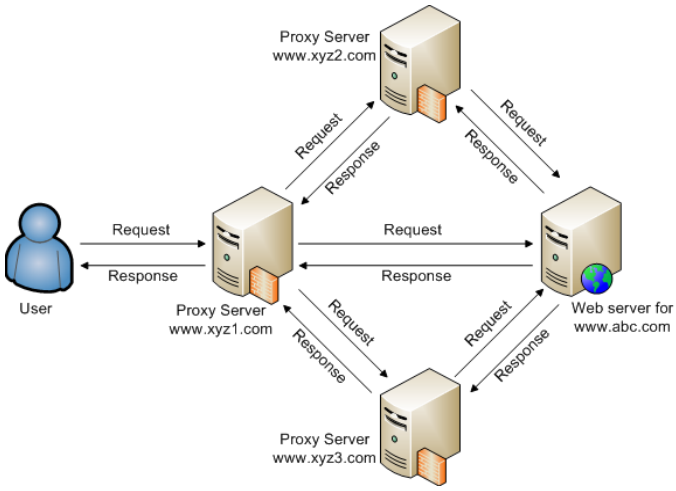



**Fig. 3.** Performing multiple requests

serves the purpose of: (a) storing evidence multiple times for security reasons and (b) depicting how a web site was appearing to the users around the world. Figure 3 illustrates this procedure.

### 3.4 Putting It All Together

Let's consider a simple use case scenario for *ProCAVE*.

**Scenario.** *The web site www.abc.com contains content that can be presented as evidence in a judicial process. It must be properly collected and its authenticity must be validated.*

**Procedure.** The user visits web site *www.xyz.com* (the *Web Proxy*) and through that proxy performs a request to the web site *www.abc.com*. The result of the latter request is shown on his/her browser. If this is the content that, according to the user, is related to a crime, then the user selects the content that he/she wishes (or does not wish) to be anonymized. Consequently, he/she chooses to store the content, and the website returns the evidence from multiple locations, along with validation of its authenticity.

The above procedure addresses all the problems related to evidence from online content. Specifically, the user has a fully-functional html file that can be presented to the court as its authenticity is proved. Moreover, the user can prove the exact date and time that the evidence was created and by which engine. Therefore if someone questions the initial evidence, he/she can take a copy from the corresponding engine to see if something has changed. In addition, there is evidence from multiple locations, something that provides the user with the opportunity to prove that this content was visible from around the world.

The way *ProCAVE* handles digital evidence covers more than one of the requirements stated by [14,24]; they can be admitted to court by a witness, an expert (administering the local copy of *ProCAVE*) can testify in the court regarding the locally stored content and this content can also fall under rule 901(b)(7) as public record, if *ProCAVE* is run by a governmental organization [15].

Last but not least, the entire procedure respects the privacy of the user, since he/she is given the option to scramble all the content that he/she wants to be invisible or not accessible to third/not authorized persons.

## 4 Implementation and Experimental Evaluation

### 4.1 Implementation

To evaluate the effectiveness and applicability of the proposed solution, *ProCAVE* was implemented as a PHP/MySQL software tool [22]. The implementation was based on a simple web proxy, described in [17], extended to include the collection and authenticity validation functions described in the previous sections.

More specifically, each time a request for a web site is performed through the *Web proxy*, a random 6-digit hex number is assigned to it and to the subsequent requests made for downloading the other components of the page (stylesheets, images, scripts, etc.). Thus, when the viewed page needs to be stored, the CVA engine uses this number to group all these files together (response). In terms of hashing, the standard `hash_file`() php function, with the 320-bit version of the RIPEMD algorithm, is utilised. On the other hand, digital signing is performed with the SHA-1 algorithm for hashing, followed by encryption with a private-key generated with the help of OpenSSL [23].

### 4.2 Results

*ProCAVE* was tested with various web sites of diverse content (news, academics, entertainment). For each of those web sites, evidence was collected through the proposed tool and was then compared to the content resulted from the usual "Save As" procedure supported by the browsers. The comparison was performed with the Similarity Analyzer [19], which examines two web pages and computes the percentage of HTML similarity and Text Similarity. The results for 10 web pages are depicted in Table 1.

It can be noticed that in most cases, HTML Similarity is close to 90%. This is an expected result, since both *ProCAVE* and the browser's "Save As" function change the links in the downloaded HTML code to point to local locations, which is different for those two methods. On the other hand, Text Similarity is close to 98%, which means that what someone can see through the browser is very close to what is saved through *ProCAVE*. The 2% difference on average is mainly due to text parts that are downloaded/altered in the browser, with standard or asynchronous scripting and not in the content saved through *ProCAVE*.

A graphical representation of the results, together with their mean values, is depicted in Figure 4.

**Table 1.** Results of the comparison between *ProCAVE* and the "Save As" function

| Hex | Web page | HTML Similarity | Text Similarity |
|---|---|---|---|
| 298168 | inria.fr | 99,86% | 99,98% |
| 558404 | ssl.ds.unipi.gr | 100% | 100% |
| 20d62f | news.yahoo.com | 93,23% | 96,72% |
| 337ec6 | spiegel.de | 94,68% | 99,38% |
| 408faa | behind-the-enemy-lines.com | 48,56% | 90,59% |
| 418f7e | maawg.org | 94,57% | 98,75% |
| 75a048 | ansa.it | 97,31% | 98,05% |
| 79031e | bbc.co.uk | 87,58% | 95,79% |
| 7a777b | slashdot.org | 96,14% | 99,13% |
| 82400b | xkcd.xom | 84,03% | 100% |

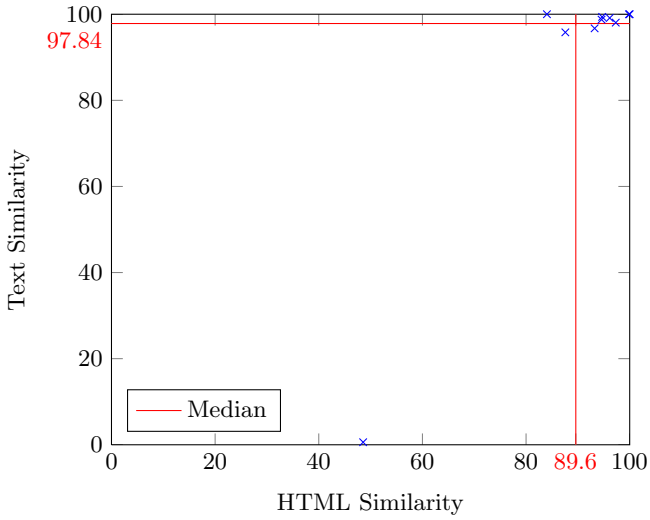



**Fig. 4.** Graphical representation of the results and their mean values

## 5 Anti-forensics and other Considerations

As stated in [1], from the very first day that digital computers and networks appeared, data hiding was, and continues to be, an issue. Since it is not expected that *ProCAVE* will be an exception to that, this section refers to technical and other issues that have mainly to do with anti-forensics.

Considering that *ProCAVE* will run in centralized locations (servers), a perpetrator could find all the domain names and ip addresses used by the tool and deny all requests originating from them - or, even worse, present legitimate content to them. A simple solution to that would be the use of a dynamic ip

address pool for the server's outgoing traffic, or the use of ip proxy servers that periodically change ip address.

Another anti-forensic technique would be the use of programming tools to present content to the user in a way not supported by *ProCAVE*, e.g. synchronous/asynchronous scripts. However, this is something that can be easily addressed through minor modifications of the way *ProCAVE* works. For instance, in the current implementation a javascript/ajax request would produce a false result, since it would modify content locally without making the correct changes in the saved copy. A solution would be to support listening of local events and simulating them to the remote end (with the use of a scriptable Web browser, like in [18]). Especially for the AJAX case, it is not a major issue since according to [25] only 3.2% of all web sites use this technology.

There are certain applications that do not fall under the above mentioned case; those would be flash-based web sites, specific chatting technologies etc. We believe that those services are out of this work's scope, since they represent instantly available content which does not resemble the traditional online content that *ProCAVE* deals with.

Since the functionality of *ProCAVE* will be publically available, it will be also vulnerable to attacks like denial-of-service (DOS), abusing etc. To that end, standard techniques for protecting a web site must be adopted, like firewalls, intrusion prevention systems, etc. Moreover, access to the website can be limited to registered users (perhaps owning a digital certificate), who will be able to perform a certain number of requests per minute.

Furthermore, regarding the use of digital certificates in *ProCAVE*'s privacy option, a Public-Key Infrastructure (PKI) must be used for their creation, management and revocation. However, the level of trust that is achieved depends heavily on the chosen Certificate Authority (CA). Simple implementations can make use of open source tools, like OpenSSL [23], but for large-scale use a commonly trusted entity must be employed.

Last but not least, an important issue refers to the authority which could be considered as being trustworthy enough to be held responsible for running this tool and keeping local copies of evidence. We argue that this issue has to be handled in accordance to the legal framework and the jurisdiction of the Forensics Department of each Country or Region; however, the distributed design of our tool and the fact that multiple requests can be performed (and multiple copies of the evidence can be saved) by remote servers, makes it easy for every individual or organization to run an instance of *ProCAVE*. In any case, we may assume that if more *ProCAVE* instances are involved, the integrity and acceptability of evidence and the procedure is better served and preserved.

## 6 Conclusions and Future Work

In this paper a simple software solution, namely *ProCAVE*, that can collect and validate the authenticity of content from online sources has been presented. To the best of our knowledge, *ProCAVE* is the first system that avoids the usual

local copies or screenshots of web sites (and the resulting dispute). Instead, it is based on an online architecture that collects evidence from multiple locations at the same time and, most importantly, in a privacy-preserving manner.

To verify it's correctness, a simple implementation of *ProCAVE* was employed for conducting a series of representative tests. The results of the tests have proved that the resulting evidence was of great resemblance to the content that the user was presented through his/her browser and thus that *ProCAVE* can produce acceptable digital evidence in real-time; i.e. during the time that the user sees the content on his/her screen, without involving him in any complicated procedures.

Future work will include modifications of the software so as to implement currently unsupported features, like listening to local events and modifying remote content accordingly, grabbing videos etc. The employment of *ProCAVE* by Forensics Departments around the world would also be of great importance, since it would provide valuable feedback from real-life scenarios.

**Acknowledgements.** The first author would like to thank Stavros Niarchos Foundation (*www.snf.org*) for supporting this work.

## References

1. Berghel, H.: Hiding data, forensics, and anti-forensics. Communications of the ACM 50(4) (April 2007)
2. CanProve - Capture Online Evidence, `http://canprove.com` (last visited: March 2013)
3. Careless, J.: Collecting and authenticating online evidence, CBA Practicelink, (last visited: March 2013)
4. Commonwealth vs. Michael M. OLaughlin: Burglary, armed assault in a dwelling, assault and battery by means of a dangerous weapon, practice, criminal, required finding, Appellate Court Decision, No. 04-P-48 (2005)
5. Council of Europe (CoE), Explanatory Report to the Convention on Cybercrime, ETS 185 (2001), `http://conventions.coe.int/Treaty/en/Reports/Html/185.htm`
6. DVers - Digital Verification Services, `http://www.dvers.gr` (last visited: March 2013)
7. Fenner, M.G.: Evidentiary Problems Associated with the Introduction of Web-Based Evidence, LSN: Evidence (Public Law) (Sub-Topic) (December 2010), Available at SSRN: `http://ssrn.com/abstract=1722714`
8. Garfinkel, S.L.: Digital forensics research: The next 10 years. Digital Investigation: The International Journal of Digital Forensics & Incident Response 7, S64-S73 (2010)
9. Gibson, J.: A Primer on Admitting Web Page into Evidence, Nevada Lawyer Magazine, `http://nvbar.org/articles/content/primer-admitting-web-pages-evidence` (last visited: March 2013)
10. Hosmer, C.: Proving the Integrity of Digital Evidence with Time. International Journal of Digital Evidence 1(1) (2002)
11. Karyda, M., Mitrou, L.: Internet Forensics: Legal and Technical issues. In: 2nd Annual Workshop on Digital Forensics and Incident Analysis (WDFIA 2007), Samos, Greece (August 2007)

12. Kerzner, M.: Evidence Authentication: Web Site Content, Atkison Baker, http://www.depo.com/E-letters/TheDiscoveryUpdate/2008/October/Articles/website_authentication.html (last visited: March 2013)
13. Leroux, O.: Legal Admissibility of Electronic Evidence. International Review of Law Computers and Technology 18(2), 193–220 (2004)
14. Lorraine, J.R., Mack, B.: Plaintiffs v. Markel American Insurance Company, Defendants. Civil Action No. PWG-06-1893, United States District Court for the District of Maryland (2007)
15. Mylonas, A., Meletiadis, V., Mitrou, L., Gritzalis, D.: Dynamic Evidence Acquisition for Smartphone Forensics. In: Proceedings of the 27th IFIP International Information Security and Privacy Conference. AICT, vol. 267, pp. 245–256. Springer (2012)
16. Mylonas, A., Meletiadis, V., Mitrou, L., Gritzalis, D.: Smartphone Sensor Data as Digital Evidence. Computers & Security (Special Issue: Cybercrime in the Digital Economy) (to appear 2013)
17. Nixon, R.: Plug-In PHP: 100 Power Solutions: Simple Solutions to Practical PHP Problems. McGraw-Hill Education (2010)
18. PHP Scriptable Web Browser, http://www.simpletest.org/en/browser_documentation.html (last visited: March 2013)
19. Similarity Analyzer, http://tool.motoricerca.info/similarity-analyzer.phtml (last visited: March 2013)
20. Sommer, P.: Digital Evidence, Digital Investigations and E-Disclosure: A Guide to Forensic Readiness for Organisations, Security Adviders and Lawyers, 3rd edn. Version 3.0, Information Assurance Advisory Council (March 2012)
21. Tanenbaum, A.S.: Computer Networks, 4th edn. Prentice Hall Professional Technical Reference (2002)
22. ProCAVE Tool. Access available upon request
23. The OpenSSL Toolkit, http://www.openssl.org (last visited: March 2013)
24. U.S.Courts, Federal Rules of Evidence (December 1, 2010), http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/2010%20Rules/Evidence.pdf
25. Web Statistics - Key data of the Web, http://www.scriptol.com/web/statistics.php (last visited: March 2013)
26. X1 Social Discovery, http://www.x1discovery.com/social_discovery.html (last visited: March 2013)

# Assessing the Feasibility of Security Metrics

Bernhard Heinzle and Steven Furnell

Centre for Security, Communications and Network Research
Plymouth University, United Kingdom
bernhard.heinzle@gmail.com, s.furnell@plymouth.ac.uk

**Abstract.** This paper proposes a self-assessment framework that allows a user to determine security metrics that are feasible specifically for the user's ISMS. To achieve this, a metric catalogue containing 95 metrics from different sources was created. The catalogue was enhanced by ascertaining requirements that need to be fulfilled in order to be able to use the metric as well as ISO 27001 clauses and controls whose effectiveness is being measured by each metric. During an assessment, the user indicates which requirements are fulfilled. After conducting an assessment, a list of feasible metrics, the number of metrics per ISO 27001 clause and control, and other information are generated as assessment results. A software prototype was created and shows a proof of concept. The results of the study were evaluated by external experts, which has validated the composition of the metrics catalogue, the design of the self-assessment framework, the value of the prototype and helped to identify areas of improvement and future work.

**Keywords:** Security Metrics, Security Measurement, Feasibility, Effectiveness, ISMS.

## 1 Introduction

Being able to measure the level of security within an organisation is important to, and desired by, many organisations. In order to achieve this, the challenge of selecting the right and most feasible metrics has to be tackled first.

This study aims to develop an approach which allows an organisation to select feasible security metrics from a given metrics catalogue. Security Metrics are used to measure the effectiveness of an organisation's Information Security Management System (ISMS) as well as the sub-processes, activities and controls of the ISMS. Specifications and guidelines for establishing a so-called Information Security Measurement Programme within an organization are given by a variety of different guidelines, standard or other publications. Some of these publications also establish a list of example metrics.

With the self-assessment framework, an organisation can describe their ISMS or organisational structure and thereby determine a set of metrics, based on a metrics catalogue. This set of metrics is then supposed to be suitable for their organisational structure and given IT infrastructure.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 149–160, 2013.

The term metric generally refers to the process and methods of quantification of a given attribute, aspect or characteristic [1,2]. The overall aim is to "[...] simplify a complex socio-technical system into models and further to numbers, percentages or partial orders" [3]. According to this definition, information security metrics measure aspects of information security.

While security metrics are defined differently and can be categorised differently [1,2,4,5,6], this study focuses on security metrics according to ISO 27004 [7]: Metrics that measure the effectiveness of an ISMS and its sub-processes and controls. A variety of frameworks and guidelines on how to set up a so called information security measurement programme exist [4,5,7,8], although these publication only give little or no guidance on how to select the most feasible or adequate security metrics.

Further publications in the area of security metrics mostly agree that this is a difficult area and further research is strongly needed [2,6,9,10,11,12]. Two approaches for determining feasible security metrics were reviewed more into detail. Savola's approach [3] is a set of evaluation criteria with a scheme how to evaluate candidate metrics. Fruehwirth et al. [13] published an approach that tries to determine feasible metrics by considering the organisation's capabilities according to the Systems Security Engineering Capability Maturity Model (SSE-CMM). The two-way relationship between metrics and capabilities is weighted; on one side the extent to which the measurement process of a metric requires a capability to be in place, and on the other side to which extent a metric can help to assess or improve a capability. It was found that both reviewed approaches rely on subjective perceptions of individuals, which can be considered as a weakness. Furthermore the both approaches do not define which metrics shall be evaluated, while integrating widely used metrics could speed up the process.

## 2 Possibilities to Describe an ISMS

For the self-assessment framework that was developed during this study a formal method to describe an organisation's ISMS is vital in order to enable the determination of feasible security metrics. Rather than evaluating a list of security metrics from a catalogue and determine the best suitable or most feasible metrics according to an evaluation scheme such as Savola's [3] approach, a method to describe an organisation's ISMS and determine feasible metrics with the information about the ISMS was researched.

Similar to the approach published by Fruehwirth et al. [13], the maturity model used in CobiT 4.1 [14] and the ISO/IEC 15504 based process capability model used in COBIT5 [5] were reviewed. When it comes to the idea of using these models to determine feasible metrics, the difficulty seems to be the establishment of a relation between single processes at specific maturity levels and single metrics. In other words, it is difficult to say that metric A can be calculated easily if process B is at maturity level C. Theoretically, metrics can be linked to processes and as generally known, processes at specific maturity levels are considered measurable. The model of assessing processes with a capability

or maturity level seems to lack of the needed granularity to tell which metrics are feasible with this information. Process capability and maturity models can give a general indication, whether processes are well managed and measurable or not. Establishing a direct link to specific metrics and producing a list of feasible metrics seems to be very difficult.

An additional possibility to describe an organisation's ISMS is offered by catalogues of possible elements of an organisation's ISMS, such as the "IT-Grundschutz Catalogues" [15]. It was evaluated how the existence of specific components could indicate that specific metrics are feasible.

Linking specific metrics to specific modules of the catalogue was not found useful for the self-assessment framework for several reasons. On one side, a module might allow a metric to be calculated so that it can be considered feasible. On the other side, the existence of a module does not implicate that a metric is feasible. To tell which metrics are feasible from the mere existence of one or more components is difficult, owed to the type of components used in BSI catalogues. The existence of a module does not say a lot how measurable the module is. Additional to the mere existence it is very often a certain condition that needs to be fulfilled. A further issue that makes the use of BSI component catalogue unsuitable for this project is the lack of up-to-dateness of the english version of the catalogue. The most current english catalogue was published in 2005. Furthermore, using a set of components like the BSI catalogues is very specific to the used framework. This weakness was also discovered for process maturity and capability levels. An organisation which has used ISO 27001/2 to organize its ISMS and control will not likely use BSI Catalogues to determine feasible metrics.

It was decided to use a method that is closer oriented to metrics and less bound to ISMS frameworks like ISO 27001 [16] or the COBIT process capability and maturity model [5,14]. To describe an organisation within the self-assessment framework, requirements of each metric were worded without using a predefined model or formal language. Requirements are described as a condition that needs to be fulfilled by components of the ISMS or information that needs to be reported by components of the ISMS, e.g. "Inventory of assets indicates number of applications that are classified as critical to the organisation". The list of recorded requirements can then be used to build an organisational model. An organisation shall be described by the list of fulfilled requirements, which will be a subset of the overall list of requirements.

## 3 Metrics Catalogue

A metric catalogue was created containing the following information:

- Source of the metric
- Title of the metric
- An identifier which is unique within the source
- Brief description, e.g. "Percentage (%) of information systems that have conducted annual contingency plan testing"

- ISO 27001 processes and controls that are measured by the metric. At the end of an assessment, this allows to determine which controls are measured.
- Requirements of the metric, i.e. a condition that needs to be fulfilled for the metric to be feasible.

An overview of sources and the number of metrics used from each source is shown in Table 1.

**Table 1.** Number of metrics per source

| | |
|---|---|
| ISO 27004 [7] | 13 |
| NIST SP 800-55 [4] | 16 |
| Steve Wright [17] | 7 |
| The CIS Security Metrics [18] | 28 |
| Scott Berinato [19] | 5 |
| Robert Lemos [20] | 4 |
| COBIT5 [5] | 13 |
| security metametrics blog [21] | 9 |
| total number of metrics | 95 |

As mentioned earlier, requirements of each metric were worded without using a predefined model or formal language. The ascertained requirements differ in terms of the type of condition that needs to be fulfilled. The majority of requirements refer to the ability to report certain data. A different type of requirement refers to a condition of the data source, e.g. "Incident management differs between occurrence and detection of incident". A further type refers to an organisational condition, e.g. "Authorization from management to perform attack on password hashes" or "Availability of and expertise with password cracking software". Whenever a requirement is related to the possibility to collect data, this does not only refer to the availability of the data but also to its up-to-dateness. Although feasibility of a metric only requires data to be available and obtainable, a lack of up-to-dateness decreases meaningfulness and the metric's quality heavily.

While ascertaining requirements it was found that the feasibility of metrics depend heavily on the support given by software used for patch, asset, incident, identity, etc. management. This finding also reflects that some of these software solutions fully integrate the calculation of security metrics.

The list of ascertained requirements was then grouped into categories of requirements. Categories were made based on the area of the ISMS or the IT activities that are addressed. Categories are similar to ISO 27001 control sections or control objectives.

For each metric, the selection of ISO 27001 clauses and controls that are measured by the metric were assigned. Thus, for each clause and control the number of metrics that measure it can be determined. This also allows determining the most frequently measured clauses and controls. It was found that the controls

that are measured the most often are incident and vulnerability management related clauses and controls (4.2.2 h, 4.2.3 a, A.12.6.1, A.13.1.1, A.13.2.1, A.13.2.1), management commitment and security co-ordination (A.6.1.1 and A.6.1.2) and awareness and training (A.8.2.2).

Figure 1 shows, by the way of an example, how the metrics catalogue looks like. Metric shown is "Nr 2 - Vulnerability Measure 1" from NIST SP 800-55 [4]. The catalogue as it is displayed in the figure is created and its data is being managed by the software prototype which was used to implement the self-assessment framework (see section 4).

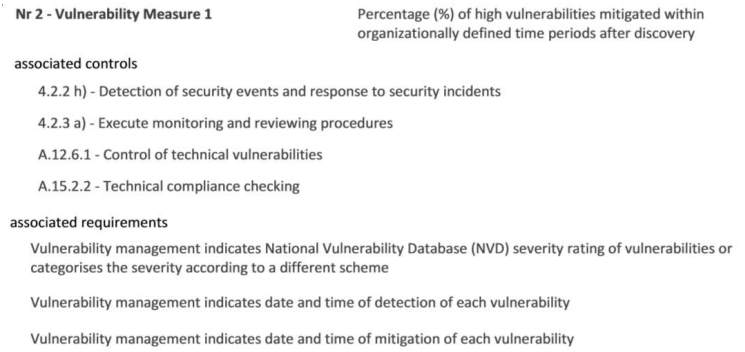



**Fig. 1.** Example metric from the metrics catalogue

## 4 Self-assessment Framework

Figure 2 shows an overview of the developed self-assessment framework, the data that is being used and how this data correlates.

As initial data the framework uses the metrics catalogue, the list of requirements, a list of ISO 27001 clauses and controls and the relationships between these three items, which are again stored in the metrics catalogue.

During an assessment, the user is asked to indicate which of the requirements are fulfilled within the ISMS. It might occur that a requirement is currently not fulfilled properly, but fulfilment can be achieved in near future. If this is possible with a reasonable effort and the user is willing to do so, the user can indicate this for each fulfilment of a requirement. In this case, the requirement is considered as fulfilled to the effect that the metrics that rely on this requirement are considered feasible as long as the metric's other requirements are fulfilled. Comments on how the requirement will be fulfilled in future should be added for documentation purposes.

The opportunity to add comments on the possibilities and modalities of data collection related to each requirement is given to the user. While indicating that a certain requirement is fulfilled, details on how often respectively easily data can be collected or other comments can arise, which is why this information can

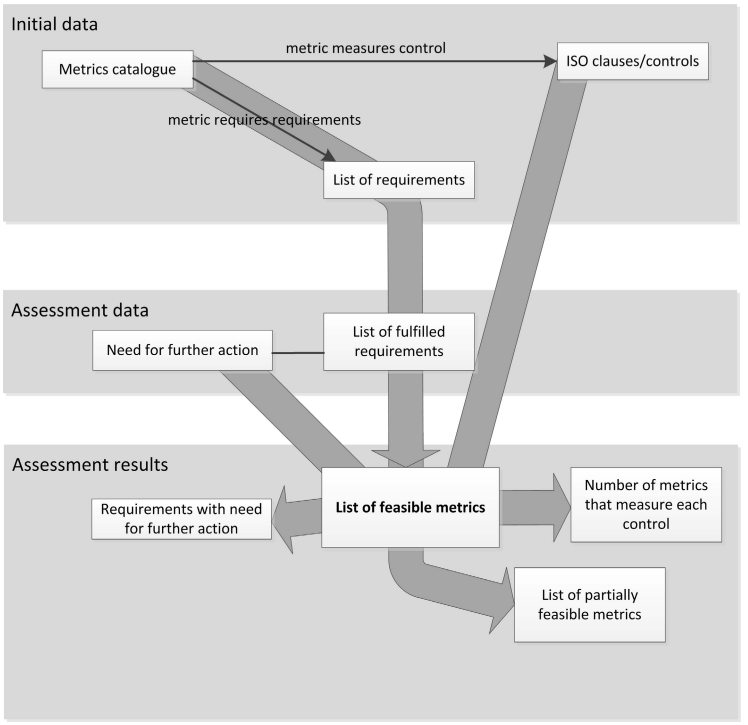



**Fig. 2.** Self-assessment framework

be added by the user. This option is given to user as the possible interval of data collection has impact on the possible interval of metric results. It would be possible to ask the user to enter the interval as a numerical value of days or select from weekly, monthly, quarterly, etc. for each requirement and then determine the measurement interval of the metrics automatically. However, in the current version this data is collected as comment only and can be used for documentation purposes.

For an assessment the scope is very important. Components with different scopes concur at an assessment: the ISMS, the measurement programme according to ISO 27004, a requirement and in particular a data source (e.g. patch management software) as well as a control. In the best case, all listed components have the same source, but this might not always be the case in practice. According to ISO 27001 [16], the scope of an ISMS and its controls should be same and be defined clearly. In practice, it might occur that a requirement's data source has a wider scope than an ISMS control, although these two components are related to each other or even the same (e.g. an IDS as data source for a metric and the same IDS as security event detection and monitoring control). This shows that a data source of a metric might have a different scope than the control it is measuring. In addition, the measurement programme has a scope that might

differ. According to ISO 27004 [7], the scope of the measurement programme is even recommended to be smaller at the beginning than the scope of the ISMS. This issue was found very difficult to be addressed in the framework. However, the user should keep this in mind and consider scope discrepancies while indicating which requirements are fulfilled and make a note of these discrepancies while conducting an assessment.

## 5 Assessment Results

Once the user has scrolled through the list of requirements and has indicated which ones are met, the following can be determined:

**Feasible metrics:** A list of feasible metrics is generated, i.e. metrics that have all their requirements that are not optional fulfilled. If one or more requirement of a metric needs some further action in order to be fulfilled properly, the metric is still considered feasible.

A feasibility score is calculated for each metric by dividing the number of fulfilled requirements by the number of need requirements. Optional requirements are not considered at this score. Using the feasibility score, a metric is feasible if the score is equal to 1.

All metrics on this list can be used with the ISMS that was assessed with the self-assessment framework, given that the needed action indicated in the next section is taken.

**Requirements that need further action:** A list of requirements that need further action to be fulfilled properly is generated, i.e. all the requirements for those it was indicated that the requirement is currently not fulfilled but fulfilment will be achieved with reasonable effort in near future. All issues on this list should be addressed by the user of the framework.

**Partially feasible metrics:** A list of partially feasible metrics is generated, i.e. some but not all the requirements were fulfilled. In terms of the previously mentioned feasibility score, this refers to metrics that have a score greater than 0 and less than 1. This list indicates which metrics could be used if more requirements were fulfilled.

**Measured ISO 27001 clauses and controls:** A list of ISO 27001 clauses and controls with the number of feasible metrics per clause and control is generated. For each clause and control, the number of feasible metrics that are assigned to the clause or control according to the metrics catalogue is counted and displayed. In this way, the user of the framework sees at a glance of which clauses and control he or she is capable of measuring the effectiveness.

## 6 Software Prototype

With the aim of providing an example how the self-assessment framework could be used in practice, a software prototype was implemented. This was achieved using Microsoft Access 2010. The prototype offers functionalities for editing the

metrics catalogue and the list of requirements as well as conducting assessments. Reports for metrics (i.e. the metrics catalogue), requirements and assessments can be generated as well. A screenshot of the assessment form that is used for conducting an assessment and collecting data about currently fulfilled requirements is shown in Figure 3.

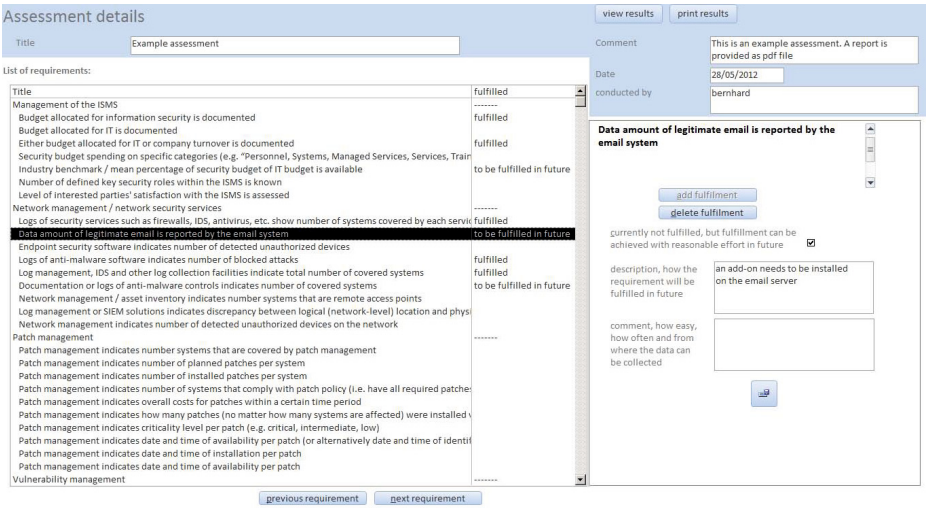



**Fig. 3.** Assessment form

In order to draw a line between the framework and the prototype, in can be said that the framework is the theoretic approach of using the metrics catalogue (including the requirements per metrics) and determining feasible metrics for a customer by using the entire list of requirements and indicating which metrics are fulfilled. The prototype is a software implementation of the framework, but the idea behind it resides with the framework. That means that it is not necessary to use the prototype in order to use the framework, one could make a different implementation or do it manually with paperwork. The self-assessment framework and the prototype are not the same thing but the prototype is strongly linked to the framework.

## 7 External Evaluation and Discussion

The results of the study were evaluated by 11 external experts. Evaluators are working in the following positions: Professor at University of Applied Sciences Upper Austria, CISO at Domestic & General, Manager at a leading Security Consulting Company in London, Sr. IT Auditor at General Motors UK, Security Manager at HCL Great Britain Ltd, GRC Consultant in InfoSec at RNG Conseil Limited, Digital Security Risk Consultant at BP, Information Security

Manager at Marie Curie Cancer Care, Audit Manager at Cofunds Limited, Security Manager at Hermes Fund Managers Limited as well as IT Risk and Compliance Manager at Sony Europe. The evaluation was done by asking for the evaluators' opinions about the metrics catalogue, the list of requirements, the self-assessment framework (as a theoretic description) and the software prototype via 13 open-ended questions. The questionnaires were sent to the experts, who then completed them and send back their responses.

In general, the evaluated components were found very useful. Some evaluators found that the catalogue contains too many metrics and is too lengthy. The number of metrics in the catalogue could be a problem if the catalogue was published as security metrics guideline or proposed set of security metrics. The difficulty to select the "right" metrics from this long list is addressed in the self-assessment framework by proposing feasible metrics based on the fulfilment of requirements. In other words, it is the idea of the project to take a list of metrics that would be to long if a user had to select metrics manually from the catalogue. The project foresees the self-assessment framework which uses the catalogue and in particular the list of requirements to determine feasible metrics.

No metrics were proposed for removal from the catalogue, although some metrics were critisized as not being meaningful or as bad quality metric in general. Some new metrics were recommended; ITIL metrics for service management related areas such as configuration management and some specific sources were proposed. The areas of SOX and other compliance-related aspects and business continuity management were identified as areas that need more metrics.

Some metrics were found as infeasible respectively the metric's requirements were found as very difficult or unlikely to be fulfilled. This is known and shall not be considered as a weakness. Metrics can be added to the catalogue, even though their requirements are very infeasible, as long as their requirements were ascertained and worded correctly. This solely results in the metrics being feasible during hardly any assessment.

It was also commented that metrics should be linked to business objectives and then be selected according to the metric's ability to fulfil relevant objectives, as it is done by other publications [4,7,13]. However, metrics are linked to control objectives, which are subordinate to business objectives. Although the framework in its current version does not select metrics according to a list of ISO 27001 control objectives that shall be achieved, it is possible to adapt both framework and prototype to allow this.

The self-assessment framework was found adequate and helpful by the majority of evaluators, although the description should be improved. The possibility to indicate that a requirement is currently not fulfilled but will be fulfilled in future was found very useful, as this – together with the ability to add comments how fulfilment will be achieved – allows a gap analysis where work needs to be done and allows the generation of an action plan, as it is done by the prototype.

The prototype was found useful by many evaluators, also the report and in particular the generated list of ISO 27001 clauses and controls with number of feasible metrics per control was found useful. However, the fact that it is rather

a prototype delivering a proof of concept of the framework than a software that is ready for release was addressed by many evaluators. Some evaluators detected a lack of usability, as it was commented that it takes some time to understand how the forms should be used and how the software works. It was commented that visualizing the flow of an assessment and indicating the current step in the assessment form would make it easier for the user to understand how the assessment works. As further proposals for future development, it was mentioned that more graphical output would be beneficial, in particular for showing the report to company executives. For example, a pie chart showing feasible metrics per source could be added to the catalogue. Some proposals were related to extending the software to support data collection and calculation of metric results as well. The results of the study leave room for further development, but these proposals address functionalities that were not part of the original aims of the study.

## 8 Conclusion and Future Work

The metrics catalogue delivers an extensive set of metrics for measuring the effectiveness of an ISMS or processes and controls of an ISMS. The use of the catalogue is not limited to the self-assessment framework; it can be used independently as a collection of security metrics. The catalogue is not only a collection of security metrics, also ISO 27001 clauses and control were assigned to each metric if their effectiveness is being measured. As essential information for the self-assessment framework requirements were ascertained for each metric. The metrics catalogue does not and could never claim completeness. As used sources can change or new sources can appear, constant monitoring of existing sources and updating of the catalogue is needed.

The self-assessment framework defines how feasible metrics can be determined. An assessment is conducted by presenting the list of requirements to the user, who indicates which requirements are fulfilled by the user's ISMS. Results of the assessment are not only feasible metrics: a list of partially feasible metrics can be created together with an action plan indicating which requirements need further action to be fulfilled properly and the number of feasible metrics per ISO 27001 clause and controls, which has the benefit that the user of the framework sees at a glance which parts of the ISMS have their effectiveness measured. With the self-assessment framework, anybody can determine feasible metrics; no special knowledge regarding security metrics is needed. The only prerequisite is being sufficiently informed about the ISMS or having enough information about the ISMS at disposal so that one can indicate which requirements are fulfilled.

The prototype provides a proof of concept of how an assessment according to the self-assessment could be conducted with tool support. Additionally, the prototype allows management of all data needed by the self-assessment framework and generates documents such as the metrics catalogue. The prototype is rather a proof of concept than a piece of software that is ready for release. Further work is needed before releasing it to the market. The framework and the prototype

could be developed further by adding graphical charts to the reports, allowing users to adapt metrics, include processes from data collection to presentation of metric results and offering more interactive methods than PDF files to explore data like the metrics catalogue or the assessment results.

The external evaluation showed the usefulness of the results of the study and additionally helped to identify limitations and future work. However, the number of 11 evaluators is too low to enable generalisation, a more extensive review should be conducted in future.

In common with other approaches for selecting metrics, this approach still relies on subjective perceptions of individuals. The ascertainment of requirements for each metric involves subjectivity. In addition, the assignment of controls to security metrics was done mainly based on the perception of the researcher. The metrics catalogue and in particular mappings to controls and requirements could be revised in a peer review process.

The reviewed frameworks and guidelines for establishing an information security measurement programme offer little guidance on how to select metrics. Therefore, the self-assessment framework could be integrated into those frameworks. Any source of metrics could list the requirements per metric and enable users to determine feasible metrics via the self-assessment framework and the prototype.

## References

1. Savola, R.: Towards a taxonomy for information security metrics. In: Proceedings of the 2007 ACM Workshop on Quality of Protection, QoP 2007, pp. 28–30. ACM, New York (2007)
2. Jansen, W.A.: NIST IR 7564: Directions in security metrics research, National Institute of Standards and Technology, U.S. Dept. of Commerce, Gaithersburg (2009), http://csrc.nist.gov/publications/nistir/ir7564/nistir-7564_metrics-research.pdf (accessed December 25, 2011)
3. Savola, R.: On the feasibility of utilizing security metrics in software-intensive systems. IJCSNS International Journal of Computer Science and Network Security 10(1), 230–239 (2010)
4. Chew, E., Swanson, M., Stine, K., Bartol, N., Brown, A., Robinson, W.: NIST Special Publication 800-55: Performance Measurement Guide for Information Security, National Institute of Standards and Technology, U.S. Dept. of Commerce, Gaithersburg (2008), http://csrc.nist.gov/publications/nistpubs/800-55-Rev1/SP800-55-rev1.pdf (accessed December 15, 2011)
5. COBIT5, Illinois, A Business Framework for the Governance and Management of Enterprise IT. ISACA (2012), http://www.isaca.org/COBIT/Pages/Product-Family.aspx (accessed May 16, 2012)
6. Saydjari, O.S.: Is risk a good security metric? In: Proceedings of the 2nd ACM Workshop on Quality of Protection, QoP 2006, pp. 59–60. ACM, New York (2006)
7. ISO 27004, Genf, ISO/IEC 27004:2009 – Information technology – Security techniques – Information security management – Measurement. International Organization for Standardization, ISO (2009)

8. Payne, S.C.: A Guide to Security Metrics, SANS Institute (2006), http://www.sans.org/reading_room/whitepapers/auditing/guide-security-metrics_55 (accessed December 17, 2011)
9. Bellovin, S.: On the Brittleness of Software and the Infeasibility of Security Metrics. Security & Privacy 4(4), 96 (2006)
10. Bayuk, J.: Alternative Security Metrics. In: Information Technology: New Generations, ITNG 2011, pp. 943–946 (2011)
11. Hinson, G.: Seven myths about information security metrics. The Information Systems Security Association ISSA Journal, 1–6 (July 2006)
12. Rosenquist, M.: Measuring the Return on IT Security Investments, Intel Corporation, Whitepaper (2007), http://communities.intel.com/docs/DOC-1279 (accessed December 02, 2011)
13. Fruehwirth, C., Biffl, S., Tabatabai, M., Weippl, E.: Addressing misalignment between information security metrics and business-driven security objectives. In: Proceedings of the 6th International Workshop on Security Measurements and Metrics, MetriSec 2010, pp. 6:1–6:7. ACM, New York (2010)
14. CobiT 4.1, Illinois, Control Objectives for Information and related Technology. IT Governance Institute (2007), http://www.isaca.org/Knowledge-Center/cobit/Documents/CobiT_4.1.pdf (accessed December 12, 2011).
15. BSI IT-Grundschutz Catalogues, Bonn, Federal Office for Information Security (BSI) (2005), https://www.bsi.bund.de/EN/Topics/ITGrundschutz/ITGrundschutzCatalogues/itgrundschutzcatalogues_node.html (accessed December 12, 2011)
16. ISO 27001, Genf, ISO/IEC 27001:2005 – Information technology – Security techniques – Information security management systems – Requirements. International Organization for Standardization, ISO (2005)
17. Wright, S.: Measuring the Effectiveness of Security using ISO 27001 (2006), http://wwww.iwar.org.uk/comsec/resources/iso-27001/measuring-effectiveness.pdf (accessed January 07, 2012)
18. The Center for Internet Security, The CIS Security Metrics (2010), https://benchmarks.cisecurity.org/tools2/metrics/CIS_Security_Metrics_v1.1.0.pdf (accessed November 29, 2011)
19. Berinato, S.: A Few Good Information Security Metrics (2005), http://www.csoonline.com/article/220462/a-few-good-information-security-metrics (accessed May 05, 2012)
20. Lemos, R.: Five Strategic Security Metrics To Watch (2012), http://www.darkreading.com/security-monitoring/167901086/security/perimeter-security/232601457/five-strategic-security-metrics-to-watch.html (accessed May 20, 2012)
21. Brotby, C., Hinson, G.: Security Metametrics: SMotW: Security Metrics of the Week (2012), http://securitymetametrics.blogspot.co.nz/search/label/SMotW (accessed June 23, 2012)

# The Influence of Social Media Use on Willingness to Share Location Information

Bjørnar Tessem and Lars Nyre

Department of Information Science and Media Studies,
University of Bergen, Postbox 7802, 5020 Bergen, Norway
{bjornar.tessem,lars.nyre}@uib.no

**Abstract.** Willingness to share personal information is a strong indication of trust in persons and confidence in institutions. Mobile phones, high connectivity, and social media have opened up new ways of sharing personal information, and many are using these possibilities eagerly. Here we present results from a study on how social media use influences people's willingness to share location and other personal information. We conducted a survey about willingness to share and in addition ran an experiment where the treatment was use of Foursquare, a location-oriented social medium. The analysis shows that frequent social media users are more inclined to share location and other personal information than others. The difference varies, as they are not much more willing to share location information with persons, but more willing to share with social media and other institutions. Frequent users seem to have more confidence in institutions, both public and commercial, and are more willing to share location in exchange for location oriented services. A main finding is that the experience with social media itself is an important cause for the increased confidence.

**Keywords:** privacy, confidence, information sharing, social media, location.

## 1 Introduction

With the World Wide Web privacy has become a prominent issue in the daily lives of people. People have access to technologies that enable them to share personal information with persons, social media, and institutions at any time and in any place. The need to protect personal information or control its propagation and use is what information privacy is about, and it has become something we relate to daily. We have to assess the risk of personal information being used in a way we do not appreciate, balancing this with the value gained from sharing with recipients ranging from family to distant organisations, all who may want our personal data to support their activities.

A particular form of personal information is the location a person is at. Knowledge about location has a potential for persons and institutions to provide valuable services like information about public transportation or nearby restaurants.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 161–172, 2013.

But location recipients may also have intrusive or threatening uses like targeted advertisement and private intelligence activity on their agenda. Hence our trust in persons or confidence in institutions are essential factors when it comes to sharing location information and other information types. In this article we analyse personal location information and how willingness to share it varies.

Central channels for sharing location information and other personal information are social media. We see that people to a large extent use social media to tell where they are, what they do at the place, and what their plans for moving around are. So, the main research focus in this article is to assess the effect of the use of social media on the willingness to share location information. We also take a look at how demographic variables like age and gender influence the willingness to share location information, and also contrast location information with people's willingness to share unspecified information types.

The article continues with a background section where we place our study in the context of trust, confidence, and privacy research. In section 3 we motivate our research and present research questions and methods, before we analyse the survey data as well as an experiment that we conducted (sections 4 and 5). The results presented in the discussion and conclusion sections are mainly showing that social media experience has a significant positive and causal effect on the willingness to share locations with institutions in general, and in particular with social media themselves, indicating that people's confidence in institutions increases with increased social media use.

## 2 Background

The concept of trust has been studied by many sociologists, and it is often contrasted with the term confidence, which refers to our relation to institutions [1–3]. Trust is for these researchers a concept used for the belief that a *person* will behave in a way we approve of. Trust thus implies a continuous risk assessment. Confidence on the other hand is the belief that an *institution* works according to society's expectations, playing a meaningful role. Predictibility is thus a main factor of people's sense of confidence. This distinction between trust as a personal relation and confidence as a societal relation to institutions is of importance when we investigate how willing people are to share their personal information, and is central in the analysis and discussion presented here.

When people have trust or confidence in the information recipient they are normally more willing to share their personal information. Beldad et al. [4] mention five factors that influence our willingness to share with institutions; those are perceptions about data relevance, the sensitivity of the data, perceptions of risk, trust (confidence), and benefits from sharing. Risk and confidence are for some based on privacy statements or security mechanisms like passwords [5], but most commonly the organisation's need for a good reputation[6]. Willingness to share is also influenced by the perceived value of services. So, services are traded for risk. When it comes to trust (in persons) and willingness to share, this is to a large extent connected to who the person is. Wiese et al.[7] have

shown that people's perception of (emotional) closeness to a particular person is almost the only factor deciding whether they are willing to share. In addition, the willingness to share is strengthened if a person is a family member.

In the context of privacy, social media play a particular role, as the social medium on the one hand is a communication channel to your friends, and to some extent a proxy for physical nearness to your friends. On the other hand, it may be a commercial service provided by an organization that will use the information gained from the social medium for their own purposes. This may be about using people's personal data for statistical purposes, but it is also about targeting people as individuals for commercial purposes.

In a review study, Nadkarni & Hoffmann [8] show that the two main reasons for people using Facebook is the need to belong, and the need for self-presentation. Belonging in a society means to give something of yourself, and self-presentation is about informing others, directly or indirectly, about personal qualities. But as people share personal information in the social medium, they also automatically provide information to the owner of the social medium about their interests, opinions, and needs. So, people somehow balance the trust they have in their friends to whom they promote themselves with the confidence that the medium owner will use the information in a way that is not harmful. It would be worthwhile trying to understand how people cope with this contrast, and in particular how the experienced social medium user relates to sharing information with personal and institutional recipients, as well as with social media.

In this article we focus on people's willingness to share location information. It has a temporal dimension and may be quickly outdated; it can be used to deduce other types of information about a person; and it may have a variety of uses for different recipients depending on their relation to you. Commercial companies may want to use it for pushing ads, whereas your friends need it to meet you quickly. Location is also sensitive; as soon as someone knows your location, this information can be used to deduce more information about you that may be used for harmful or annoying purposes.

The computer-human interaction field also relates to location privacy. Consolvo et al. [9] show that willingness to share location is dependent on who asks for the location, when the sharer is asked, what he or she is doing, and with whom. Their investigation is supported by several studies, but willingness to share is also influenced by the receiver's motive for wanting this information[10–12]. Olsen et al. [13] did a clustering analysis of willingness to share with different persons and established five distinct groups for our willingness to share: the public, coworkers in general, trusted coworkers and manager, family, and spouse. Willingness to share was clustered into 6 types of information (from least to most sensitive): work contact information, home contact info and formal social status, work documents and availability (including location), health and personal preferences, personal data and opinions, personal secrets. Khalil and Connelly [14] support these findings, indicating that location and activity information is something we are less willing to share than company (people we are with) and conversation (what we are talking about). Location is more

sensitive information type others because of the social dynamics in groups, and the conflict of disclosing versus not disclosing your location. Being secretive about location may be just as problematic as always being open about your location to a group that you have a close social relationship to[15].

## 3 Research Problem and Method

Factors like who, where, when, and what are essential when we assess whether we want to share data, but research on sharing has so far focussed on willingness to share with persons, and has often put institutions into a large collective category. On the institution side we also find the social media, where many share personal information for the purpose of belonging and self-promotion. With the role these media now have in our daily lives, it is relevant to see how the sharing behaviour of the experienced social medium user compares to others, and how it relates to social media experience. Merging the personal trust we have in social media friends with the confidence in the social medium company has the potential to create a particular sharing behaviour, and influence our opinion about the value of location information. We wanted to observe the variation in sharing behaviour, with social media use as a central explanation variable, also controlling for gender and sex.

More precisely formulated the main research questions of the study are:

**Q0.** How is willingness to share personal (location) information with persons, social media, and institutions influenced by social media use?
**Q1.** How does social media use relate to acceptance of recipients' using our location information?
**Q2.** How does social media use relate to willingness to share location information in exchange for different services?

The main dependent variables (willingness to share with varying categories of information recipients, accepting use of location information, willingness to share in exchange services) in our study are measured by answers to a questionnaire. The survey was conducted with 82 respondents from the Norwegian municipality of Sogndal. The respondents were mainly people we met in the streets during the day, recruited for instance at outdoor cafes. We also recruited people who filled in the survey form at office locations and at home after we had made appointments with them.

There are now thousands of location-based applications and services to available for smart phones. A well-known example is Foursquare, which is a location-oriented social medium that allows you to tell about a place you are at and what your are doing at the place when you are there. As we focused on location in this study, we also conducted an experiment to check if recent experience with a location-based social medium service (represented by Foursquare) on top of frequent use of social media in general would influence responses to questions about location information. We recruited 17 young persons in Sogndal in aged from 18-24 years, who were using Facebook frequently on their mobile, and asked

them to use Foursquare for 24 hours. These 17 persons answered the same survey form as the other 82, and the answers were compared with a control group consisting of 35 respondents in the main survey data material that shared age group and social media use pattern. The goal was to see if exposure to Foursquare led to immediate changes in the respondents' sharing behaviour.

## 4 Survey Analysis

The variables analysed are mainly index variables computed as sums of Likert scale values. Use of social media is for example operationalised as the sum of Likert scores to the respondent's degree of use of 8 types of social media (Facebook, Twitter, LinkedIn, Foursquare, ...). Willingness to share with institutions is given by one index variable adding up willingness to share with each of 6 institution categories (police, religious organisations, political organisations, newspapers, public organisations, commercial organisations). Willingness to share with persons is summarized by using answers to questions about willingness to share with family, friends, or colleagues. As mentioned above we consider social media to have a special function in this context, and have chosen to analyse responses about willingness to share with social media separately.

### 4.1 Willingness to Share Information

As a foundation for comparison we first analysed how willing our respondents were to share personal information of an unspecified category. We thus got a first understanding of the level of willingness of our respondents to share with various recipient categories by ranking the answers by recipient. The ranking found is family (median = very willing), friends, police, colleagues, government institutions, news media, social media, political organisation, commercial organisation, religious organisation (median = unwilling). With the exception of police and government, respondents are mainly unwilling to share with institutions, but are quite willing to share with persons they know well. This is a confirmation of results from earlier studies [7, 13, 16].

To get a first impression on how social media influence willingness to share we split our respondents into 2 groups, the first consisting of respondents who answered very often or often to a question about use of Facebook, and the remaining respondents were placed in the second group. We found that people who do not use Facebook or use it seldom are highly unwilling to share with social media, placing social media lowest in a ranking of recipients. Social media got a slightly higher ranking in the frequent users group than in the combined groups. For other recipients the differences in rankings are small and not significant, both in the frequent and non-frequent user groups.

With respect to our three independent variables, willingness to share personal information with persons, willingness to share personal information with social media, and willingness to share personal information with institutions, we have some correlations between age and social media use and willingness to share,

whereas gender does not seem to have any significant influence on people's responses. We also found no relationship between gender and social media use or other variables in our analysis, so for the remaining part of the article, we will not discuss gender as an explanatory factor.

There is an expected strong negative sample correlation [17, Ch. 15-1] between age and social media use, as younger people use social media much more than older people (corr = -0.572, N = 82, p = 0.000). Age correlates with negatively with willingness to share with persons (corr = -0.305, N = 74, p = 0.008), willingness to share with social media (corr = -0.411, N = 79, p = 0.000), and willingness to share with institutions (corr = -0.230, N = 75, p = 0.047). At the same time social media use correlates positively with these three willingness to share variables (persons (corr = 0.290, N = 74, p = 0.012), social media (corr = 0.424, N = 79, p = 0.000), institutions (corr = 0.306, N = 75, p = 0.008)).

It is not obvious how age and social media use interact to create the effect we see on willingness to share. However, from the literature we know that age is a demographic variable that does not influence trust and confidence [18, 19]. One would expect that this also goes for their willingness to share information, making older people as willing to share personal information as younger people. When doing a partial correlation analysis [17, Ch. 15-4] between social media use and the three willingness to share variables controlling for age, we still see a positive correlation between social media use and willingness to share with social media (corr = 0.28, N = 68, p = 0.015) and with institutions (corr = 0.261, N = 68, p = 0.029), but not with persons. When repeating the partial correlation analysis, now controlling for social media use, we find no significant correlations between age and the three willingness to share variables. Our results thus confirm the findings in the literature, and allow us to make the conclusion that social media use is the main factor influencing willingness to share. The correlation with age is a direct effect from the fact that older people do not use social media as much as the young ones.

### 4.2 Willingness to Share Location Information

When looking into the data on willingness to share location information, we first did a ranking of recipient categories. We found few differences in the list compared to the one for unspecified information, except that colleagues and social media each moved one place up on the list. A Wilcoxon test [17, Ch. 16-3] comparing willingness to share unspecified information versus location information showed that there is overall less willingness to share location information. This was significant for persons (p = 0.023), institutions (p = 0.001), and for social media (p = 0.002).

A sample correlation analysis looking for relations between age, social media use, and willingness to share location showed correlations between both age (corr = -0.442, N = 80, p = 0.000) and willingness to share with social media as well as social media use (corr = 0.532, N = 80, p = 0.000) and willingness to share with social media. We found no significant differences in willingness to share locations with persons and institutions for the eager social media users (p-values

around 0.10 for both persons and institutions). The higher willingness to share location with social media remains for eager social media users when controlling for age (corr = 0.344, N = 66, p = 0.004), but the relation to age disappears when controlling for social media use.

### 4.3 Sharing Location for Services

To get more insights into which factors that make people willing to share location information, we asked the respondents to assess to what extent they would share location data in exchange for different services. We found no correlation between willingness to share location for service and willingness to share location with persons. But there is a correlation with willingness to share location with institutions (corr = 0.379, N = 71, p = 0.001). This means that those who find location services useful, are also those who are most willing to share location with institutions in general. This also applies to specified services, as correlation is also found for 10 out of the 11 specific types of services we queried about.

When ranking the 11 services according to willingness to share location in exchange for the service we got the following list:

1. Get updated information about public transportation schedules
2. Get updated information about traffic conditions
3. Find a nearby restaurant
4. Make an appointment with a precise meeting points
5. Find facts about the place
6. Get information about public services at the place
7. Get news about the place
8. Learn about local history
9. Get tourist information
10. Buy or sell things at the place
11. Get advertisements at the place

Notice that the top four concern information that is relevant both at the location and in real time. The next ones are less temporally oriented, and are related to information you might just as well search for at home. At the bottom of the list we have two services related directly to commercial activities, which probably is due the fact that spending money is something we want to control ourselves, and such services may be felt as more intrusive than the others.

When we see these findings in relation to the findings that age and social media use explain willingness to share location with institutions, it is tempting to explain young, social media users' willingness to share location as an implication of higher appreciation of location-oriented services.

### 4.4 Opinions about Recipients' Use of Location Information

To get an understanding of how willingly the respondents accepted that recipients used their location information for varying purposes, we asked them

about how they would feel about a particular recipient (the same 10 recipients as before) using their location for particular purposes. We listed five different purposes: selling the information, giving it away or using it for directed advertising, political activities, intelligence, and statistical purposes. The respondends are most accepting to family members regarding use of location information, whereas they are least accepting of religious organisations. It must be said that the respondents were rather sceptical to any use of their personal location information, but still there were differences regarding purpose. It was considered most acceptable to use the location information for collecting statistics, then political purposes, commercial purposes, intelligence, and the least acceptable purpose was to sell the location information. An explanation is that use for statistics is least intrusive and least personal and also considered to be of benefit to society; political and commercial purposes are intrusive, but not threatening; whereas intelligence or selling your personal information to unknown parties is felt threatening.

There is a strong correlation between age and accepting that others use your location information in the sense that young people are more accepting than older people (corr = -0.417, N = 71, p = 0.000). We do not in general find the same correlation between social media use and acceptance of use. However, it is interesting that there is a strongly significant relation between social media use and acceptance of social media using your location information, and at the same time there is a strong correlation with accepting that your location is used for directed advertising. This suggests that eager social media users are more accepting of their personal information being used for commercial purposes, i.e., they are more inclined to accept the business model with targeted advertising as a necessity for social media. They have experienced the predictability of the social media, and as a result confidence grows.

Furthermore, people who more willingly share their location in social media are more accepting of others' use of their location information, and for those

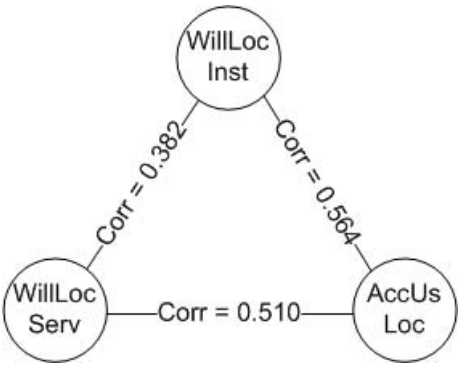



**Fig. 1.** Correlation between three variables, willingness to share location information, willingness to share location for services, and acceptance of others' use of location information

who more willingly share with institutions we find an even stronger correlation. Figure 1 shows how the three variables willingness to share location information with institutions (WillLocInst), willingness to share location for services (WillLocServ), and acceptance of others' using location information (AccUseLoc) relate to each other.

It is not obvious that this has a causal explanation, but it seems to strengthen the argument that people who like location-based services are more willing to give away their location to institutions. Probably the respondents thought about 'institutions' in the sense 'institutions that provide services' when responding to the questionnaire. These have behaved according to the respondents' expectations, and they therefore become more accepting of institutions' use of their personal data.

## 5 The Foursquare Experiment

In addition to the survey, we ran an experiment where we tested the immediate effect of use of Foursquare on people's willingness to share. The treatment was to make the 17 respondents, selected due to their age and their experience with Facebook on the mobile, use Foursquare for 24 hours. The respondents installed Foursquare on their mobiles, and checked in on Foursquare five times at different locations during a period of at least 24 hours. Foursquare was set up so that it also posted all their Foursquare check-ins on Facebook. After the 24 hours these respondents filled in the same questionnaire as the 82 other respondents. The Foursquare users were paid 200 kroner to participate in the experiment.

As a control group for the experimental group we used respondents from the survey who have the same profile as the Foursquare users, i.e., is a frequent user of Facebook, and use mobile Facebook. The control group consisted of 35 respondents. We tested the effect of the experimental treatment using Mann-Whitney tests [17, Ch. 16-3]. These showed that the Foursquare users were even more willing to share location with institutions than the control group ($p = 0.024$). They were similar to the control group when it came to sharing unspecified information, and also sharing location information with social media and persons.

Considering opinions about willingness to share in exchange for services, the Foursquare users are even more positive than the control group ($p = 0.022$), indirectly confirming that experienced social media users share more willingly with institutions. We also compared how the two groups responded to the questions about acceptance of use. We found a clear correlation in the sense that our Foursquare users are more accepting of the use of their personal location information for all recipient categories.

We have observed the effect that respondents are more willing to share locations with institutions when having used a new social medium. The value of the Foursquare service seems to have become evident for the users after one day's use, and willlingness to share location increased with their experience of using the medium. When our experiment group used Foursquare, they got a feeling of

what it in fact means to share the location with institutions, an experience that made it feel less threatening.

## 6 Discussion

The main finding of this study is that people's willingness to share personal information and location information is positively influenced by the use of social media. We found no relations between willingness to share and gender, but there is a relation between willingness to share and age. However, this was explained away as we controlled for age in the analysis.

If we accept that willingness to share personal information represents personal trust and institutional confidence, our results show that people's trust in other persons does not change significantly with use of social media when it comes to location, and neither when using a location-oriented social medium. Regarding confidence in social media, people get higher confidence when using social media, but confidence does not increase more when using Foursquare in addition. As last, institutional confidence increases with social media use and even more through the use of Foursquare. This finding was strengthened with the observation that the Foursquare users were more willing to share location in exchange for services, and had less problems with information recipients making use of the location.

It seems that the potential for personal trust is reached before the introduction of social media, whereas the institutional confidence has a potential to increase with use of social media. The experience with using social media in fact builds up people's confidence towards service providers, including providers of location services. The effect may come about because people who use social media are also more used to relating to institutions on the computer, as they for instance use the web for services, meet these institutions in that form, and have few bad experiences with them.

However, the Foursquare experiment shows that the effect is seen for two identical respondent groups, with experience in disclosing the location to others as the experimental treatment distinguishing the groups. The actual act of sharing locations on the mobile phone, the experience that it was not felt threatening, and that it was considered a valuable service, contributed to increased confidence. That is, the social media experience caused the increased confidence.

According to our observations, the increasing use of social media in the population will most likely contribute positively to trust and institutional confidence in the (Norwegian) population. The confidence in institutions may in the future be just as dependent on high quality services and presentations on the web and in social media, as it previously depended on well kept buildings, and service-minded personnel at a reception desk.

Three other observations: First, location information in itself seemed to be more sensitive for people than unspecified information types. This means that location is kept more secret than personal information in general, and should have a special role in how people's privacy issues are handled. Second, people who do not use

social media a lot are strongly against sharing information on social media. Their lack of knowledge and experience with the technology seems to create a lack of confidence in social media, and it makes these people critical of sharing personal information in social media. Third, people value most the location services that are important both at the place and in real time. Developing other types of location-based information services may not have the same potential.

A couple of validity issues must be mentioned. First, going from answers to queries about willingness to share personal location information to conclude something about trust and confidence, is perhaps a long step. Trust and confidence may not be very precisely defined as theoretical concepts in the literature, but trust and confidence in communication activities is certainly a significant contributor to willingness to share personal information. Willingness to share information is thus a clear indication of trust and confidence. Second, the respondents we have all live in one single, mainly monocultural, town in Norway, in a rural district. This is a fact which to some extent weakens the study's generalizability. It is a goal for us to redo the study in a highly urban area like London, to establish better generality of results.

## 7 Conclusion

The survey and the Foursquare experiment conducted in Sogndal, Norway show that willingness to share personal information and in particular location information is influenced by people's experience with social media. Those of our respondents that were active on social media were more willing to share with institutions, thus showing an increased confidence in institutions. The benefits seen in using location-oriented services further increased the acceptance of institutions' use of their personal location information. So, the most eager social media users are more inclined to accept the business model of location service providers companies, and will more often disclose their location in exchange for services.

It remains to be seen if social media use will increase the overall confidence in institutions on a long term basis, but as social media and other digital services on the mobile phone are embraced by all generations, we will probably see this effect. For business and governmental organisations this implies that predictable behaviour on the web reinforces confidence and opens up for collecting personal information with the aim to provide even more advanced user-adapted services. Still, this will also depend on whether institutions are careful not to abuse the confidence obtained. The scepticism to sharing personal information is still high, and infringements may turn the tide in these matter.

## References

1. Luhmann, N.: Trust and Power. Wiley (1979)
2. Seligman, A.: The Problem of Trust. University Press (1997)
3. Giddens, A.: The Consequences of Modernity. In: Sociology. Stanford University Press (1990)

4. Beldad, A., de Jong, M., Steehouder, M.: A comprehensive theoretical framework for personal information-related behaviors on the internet. The Information Society 27(4), 220–232 (2011)
5. Koufaris, M., Hampton-Sosa, W.: The development of initial trust in an online company by new customers. Inf. Manage. 41(3), 377–397 (2004)
6. Chen, C.: Identifying significant factors influencing consumer trust in an online travel site. J. of IT & Tourism 8(3-4), 197–214 (2006)
7. Wiese, J., Kelley, P.G., Cranor, L.F., Dabbish, L., Hong, J.I., Zimmerman, J.: Are you close with me? Are you nearby?: Investigating social groups, closeness, and willingness to share. In: Landay, J.A., Shi, Y., Patterson, D.J., Rogers, Y., Xie, X. (eds.) Ubicomp, pp. 197–206. ACM (2011)
8. Nadkarni, A., Hofmann, S.G.: Why do people use Facebook? Personality and Individual Differences 52, 243–249 (2012)
9. Consolvo, S., Smith, I.E., Matthews, T., LaMarca, A., Tabert, J., Powledge, P.: Location disclosure to social relations: why, when, & what people want to share. In: van der Veer, G.C., Gale, C. (eds.) CHI, pp. 81–90. ACM (2005)
10. Adams, A., Sasse, M.A.: Taming the wolf in sheep's clothing: privacy in multimedia communications. In: Proceedings of the Seventh ACM International Conference on Multimedia (Part 1), MULTIMEDIA 1999, pp. 101–107. ACM, New York (1999)
11. Lederer, S., Hong, I., Dey, K., Landay, A.: Personal privacy through understanding and action: five pitfalls for designers. Personal Ubiquitous Comput. 8(6), 440–454 (2004)
12. Anthony, D., Henderson, T., Kotz, D.: Privacy in location-aware computing environments. IEEE Pervasive Computing 6(4), 64–72 (2007)
13. Olson, J.S., Grudin, J., Horvitz, E.: A study of preferences for sharing and privacy. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2005, pp. 1985–1988. ACM, New York (2005)
14. Khalil, A., Connelly, K.: Context-aware telephony: privacy preferences and sharing patterns. In: Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work, CSCW 2006, pp. 469–478. ACM, New York (2006)
15. Mancini, C., Rogers, Y., Thomas, K., Joinson, A.N., Price, B.A., Bandara, A.K., Jedrzejczyk, L., Nuseibeh, B.: In the best families: tracking and relationships. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, pp. 2419–2428. ACM, New York (2011)
16. Beldad, A., van der Geest, T., de Jong, M., Steehouder, M.F.: Shall I tell you where I live and who I am? Factors influencing the behavioral intention to disclose personal data for online government transactions. Int. J. Hum. Comput. Interaction 28(3), 163–177 (2012)
17. Wonnacott, R.J., Wonnacott, T.H.: Introductory Statistics, 4th edn. Wiley (1985)
18. Sutter, M., Kocher, M.G.: Trust and trustworthiness across different age groups. Games and Economic Behavior 59(2), 364–382 (2007)
19. Listhaug, O.: Confidence in institutions: findings from the Norwegian Values Study. Acta Sociologica 27(2), 111–122 (1984)

# A Qualitative Metrics Vector for the Awareness of Smartphone Security Users

Alexios Mylonas, Dimitris Gritzalis, Bill Tsoumas, and Theodore Apostolopoulos

Information Security and Critical Infrastructure Protection Research Laboratory,
Dept. of Informatics, Athens University of Economics & Business (AUEB),
76 Patission Ave., Athens, GR-10434 Greece
{amylonas,dgrit,bts,tca}@aueb.gr

**Abstract.** The proliferation of smartphones introduced new challenges in the users' security and privacy. Currently, the literature concentrates mainly on the 'nuts and bolts' of their security models. Not extensive work is available on the security awareness of smartphone users, even though their role in the ecosystem is important. This is so, as users' actions directly affect their security and privacy. This paper extends a previous work on the awareness of smartphone users who install apps from official repositories. More specifically, we explore if a security background affects the smartphone security awareness of the survey participants by splitting them in two groups, comprising of security savvy and non-security savvy users. The results of the statistical analysis indicate, inter alia, that the participants' security background has slight impact on their security awareness in the smartphone ecosystem.

**Keywords:** Smartphone, Security, Awareness, User Survey.

## 1 Introduction and problem definition

Smartphones are mobile devices that combine the functionalities of cell phones and portable computers [22]. The popularity of smartphones is constantly increasing together with the number of smartphone third-party applications (apps). Apps are distributed in a centralized fashion from *app repositories* or *app markets*, such as Google Play, OVI Store, etc. The app repositories may be either official, i.e. maintained by the smartphone platform (e.g. Google Play), or unofficial (e.g. Amazon Appstore for Android).

Each platform's security model enforces different rules concerning whether apps from unofficial repositories are allowed to be installed in a device [2,16]. For instance, Android allows app installation from unofficial repositories, whereas unmodified iOS (i.e. not 'jailbroken') does not. In addition, new apps undergo different submission procedures in each smartphone platform [2,19].

Finally, the platforms have different expectations from users during app selection. They range from delegating users to infer whether an app will impair their security and privacy - which involves them to scrutinize any available security messages,

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 173–184, 2013.

the app's reputation score and review, etc. - to make simple authorization decisions, e.g. enable push notifications.

Smartphone security literature currently focuses primarily on the platforms' security details (e.g. [6,8,9,11,20]) and on malware (e.g. [5,12,16-17,23]). In this context, less effort focuses on the security awareness of users, even though their role in the smartphone ecosystem is vital. This holds true, as their actions directly affect their security and privacy (e.g. installing malware), while it may also indirectly affect other users. For instance, thorough and technically sound reviews of a suspicious app may aid a non-security savvy user to avoid its installation.

Herein, we extend the analysis of the user survey in [18], by splitting the sample population in two groups, comprising from security savvy and non-security savvy smartphone users. Our goal is to examine if a security background [21] - at minimum comprehension of the notion of threat, risk, safeguards - deriving from sources either academic (e.g., university information security courses), or industrial (e.g., information security certifications) - affects the participants' security awareness. Our results suggest that the participants' security background has a slight impact on their awareness about security and privacy in the smartphone ecosystem.

The rest of the paper is organized as follows. Section 2 presents related work. Section 3 presents the survey's methodology. Section 4 presents our results. Section 5 presents our limitations. The paper concludes in Section 6.

## 2 Related Work

The literature that focuses on the security awareness of smartphone users is rather limited and focuses on Android [3,4,7,14]. Authors in [3] argue that current risk signals employed by an app repository become ineffective over time, as users tend to click through them.

In [4], users were found not to consider security and privacy issues during app selection, as they tend to ignore privacy policies and EULAs. Users also tend to ignore security messages during app installation from app repositories [7,14]. Moreover, they are unable to comprehend selected permissions and the risks that are related to them. Thus, security messages cannot assist smartphone users to make appropriate security decisions.

Our previous analysis [18], which was concurrent with the above works, was not limited to Android users. Nonetheless, we found similar results regarding the lack of security awareness by all participants, irrespective of the device platform. In specific, users ignored the presented security and agreement messages and did not consider security issues during app selection. They poorly adopted common physical security controls (e.g. device password), as well as other third-party security software. Also, they consider apps that are available in the official app repository as risk-free.

In this paper, we show that this misconception about apps from the official app repository, the poor adoption of security controls, as well as ignorance of the prompted messages, do exist in both groups of participants.

# 3 Methodology

## 3.1 Data Collection and Demographics

The survey took place from September-December 2011 in Athens, Greece. We conducted structured interviews [10], which lasted on average 5-8 min (due to space limitations the questionnaire is archived in [1]). Discussion with the user during the interview aimed to ensure the validity of responses, as well as the comprehension of questions and technical terms.

The responses were collected from random people (on the street, or from public transportation means, i.e., train stations, underground, etc.). Participants were asked if they owned a smartphone and if they wished to participate in a survey about smartphone app usage. Then, they were asked if they knew what an app repository (or 'app market') is, and if they had installed any third-party app from it. A questionnaire was given to a user only if he/she had installed at least one smartphone app.

We analyzed data from 458 smartphone users ($min_{age}$=15, $max_{age}$=52). Among them, ~70% were male and 81% were aged [15-30]. The 56.3% were non-security savvy (Q4), i.e. not having a generic security background [21] from a source either academic (e.g. university security courses) or industrial (e.g. IT security certifications). Users classified themselves as: (a) 10.3% non-technically savvy ('moderate' IT), (b) 41.3% with good IT skills, and (c) 48.5% technically savvy ('excellent' IT). The 81.4% were aware of smartphone malware and 95.2% were privacy concerned. The 75.8% stored personal data and 35.8% stored business data in their devices. Finally, an in-depth discussion of the sample's descriptive statistics is available in [18].

## 3.2 Data Analysis

We splitted the sample into two groups (*SECSAV*, *NSECSAV*), according to the participants' responses in Q4, in order to examine any differences in their responses. *SECSAV* included users with a generic security background [21] and *NSECSAV* included non-security savvy users, respectively. Initially, we examined if the groups responded similarly in the same questions. For doing so, we computed the appropriate $x^2$ distribution test of independence (significance level a=0.05), between the responses regarding the security background question (*sec_background_question*) and the responses of the rest instrument's questions. Thus, we tested the hypotheses that:

$H_0$: *sec_background_question and question$_i$ are independent*
$H_1$: *sec_background_question and question$_i$ are not independent*

for every pair {*sec_background_question*, *question$_i$*} in the instrument.

Then, we further analyzed the groups by identifying and comparing correlations between their responses. We computed for each group of smartphone users, the $x^2$ distribution test of independence (significance level a=0.05) for the responses of every pair of questions {a,b} in the instrument. We tested the null hypothesis $H_0$: *question$_a$ and question$_b$ are independent* and $H_1$: *question$_a$ and question$_b$ are not independent*. Then, we examined the direction (i.e. positive, negative) of all correlations that were statistically significant, by computing the φ coefficient. We filtered out findings that did not offer anything to the security discussion (e.g. correlation between

device misplace and efficiency as criterion for app selection). Finally, we compared the different associations of the two groups. The next section includes the results of our analysis.

# 4 Findings

## 4.1 Response Diversity

This section focuses on the two groups' diversity of responses. Our analysis revealed significant differences in the two groups' responses in the questions that are presented in Table 1. Descriptive statistics for the rest responses of the survey are available in [18] and will not be repeated here.

As one would initially assume, significantly less *SECSAV* users were unaware of smartphone malware ($x^2=8.623$, $p=0.003$). Also, considerably more *SECSAV* users regard smartphone security software (hereto 'secsoft') as essential ($x^2=10.803$, $p=0.001$) and password protect their devices ($x^2=5.743$, $p=0.017$). Regarding the security messages, significantly more *SECSAV* users scrutinized them ($x^2=8.097$, $p=0.004$), while significantly more *NSECSAV* users occasionally (i.e. sometimes) inspected them ($x^2=6.485$, $p=0.011$). However, this attitude is orthogonal with the expectations of smartphone platforms, in which users make informed decisions by scrutinizing any displayed security messages [16].

The results revealed that the majority of *NSECSAV* users consider apps that are indexed in the app repository as secure for installation in their devices (hereinafter this is referred to as 'trust in the app repository'). This trust in the repository is a serious vulnerability according to the analysis in [18]. Significantly less ($x^2=10.893$, $p=0.001$) - but still a majority of - *SECSAV* users trust the app repository. Moreover, even though significantly more *SECSAV* users considered the reputation ($x^2=6.319$, $p=0.012$) and security and privacy issues ($x^2=12.949$, $p<0.001$) during app selection, they were the minority of users in both groups. This clearly suggests that current risk signals in app repositories are not effective [3,18].

Finally, the results indicate that significantly more *SECSAV* users are technically savvy ($x^2=20.566$, $p<0.001$), while significantly more *NSECSAV* users have good ($x^2=11.257$, $p=0.001$), or moderate ($x^2=4.102$, $p=0.043$) IT skills.

**Table 1.** Differences in responses between user groups

| Question | SECSAV | NSECSAV |
|---|---|---|
| IT skills (*excellent*), (Q3c) | 60.5% | 39.1% |
| IT skills (*good*) (Q3b) | 32.5% | 48.1% |
| IT skills (*moderate*) (Q3a) | 7.0% | 12.8% |
| Privacy or security app criterion (Q8e) | 7.0% | 0.8% |
| Reputation app criterion (Q8d) | 12.5% | 5.8% |
| Security messages (*always*) (Q11c) | 46.0% | 32.9% |
| Security messages (*sometimes* ) (Q11b) | 41.5% | 53.5% |
| Smartphone malware existence (Q16) | 87.5% | 76.7% |
| Smartphone security software essential (Q19) | 74.0% | 59.3% |
| Trust app repository (Q9) | 68.5% | 81.8% |
| Use device lock (Q20b) | 70.5% | 59.7% |

### 4.2 Correlation Diversity

This section focuses on the two groups' correlation diversity in the user responses. Table 2 includes the details of the tests of independence in each group. It also summarizes the findings of this section.

**Poor Adoption of Smartphone Security Software.** The analysis revealed that users in *both groups* ignore endpoint security (*Finding 1, Fd1*), as: (a) they regard smartphone secsoft essential, but tend to use them only in their PC and not in their smartphone ($Fd1_a$). This finding shows a clear asymmetry of security awareness in the two platforms (i.e. PC, smartphone). Our results revealed that (b) even though users in *both groups* ignore endpoint security they tend to be unsure if submitted apps in the repository undergo security analysis ($Fd1_b$). Also, (c) iOS *SECSAV* users tend to regard that smartphone secsoft is not essential ($Fd1_c$). This erroneous security posture, which may stem from Apple's *walled garden* approach (and marketing strategy), has been proven invalid since malware has been found in iOS [5].

On the other hand, we found that some participants (the minority of the sample, according to [18]) opt for endpoint security (Fd2): (a) users in both groups tend to use smartphone secsoft when they respond that app testing takes place in the repository (*Fd2a*), thus not relying solely on the centralized protection, (b) *SECSAV* users consider smartphone secsoft essential and search for free secsoft in the app repository (Fd2b), implying that these users try to protect their devices, and (c) *NSECSAV* Android users tend to use smartphone secsoft ($Fd2_C$), which may result from the fact that various smartphone malware families target Android [23].

*NSECSAV* users who inspect an app's reputation are less likely to regard smartphone secsoft as essential (*Fd3*). This is a rather interesting finding, since smartphone secsoft can be used as an additional line of defense against malware, especially the ones from mediocre attackers [16-17]. Reputation score alone cannot guarantee that an app is benign, as popular apps do not necessarily respect a user's privacy [3].

**Unauthorized Access.** Our results revealed that users in *both groups* are exposed to unauthorized *remote access* (*Fd4*). This is because they were found not to: encrypt their data, use third-party secsoft, and scrutinize security messages. Thus, if they grant malware or greyware access to sensitive resources or an attacker gains unauthorized remote access via vulnerability exploitation, then their resources will be unprotected. Specifically: (a) users who do not use smartphone secsoft are less likely to encrypt their data ($Fd4_a$) and (b) users who occasionally read security messages (i.e. responded sometimes) are less likely to encrypt their data ($Fd4_b$). Also, we found *SECSAV* users who (c) are unaware of smartphone malware and are less likely to encrypt their data ($Fd4_c$) and (d) tend to have encryption and remote wipe disabled ($Fd4_d$). Finally, *NSECSAV* users who ignore security messages are less likely to encrypt their data ($Fd4_e$).

Furthermore, our results revealed multiple cases where users in *both groups* are exposed to unauthorized physical access (*Fd5*). This happens since users do not adopt physical security controls (i.e. device password, encryption, remote device locator, remote wipe) and/or third-party security software, which can proactively (e.g. encryption), or reactively (e.g. remote wipe) protect against this threat. Specifically: (a) users

are less likely to use device password lock when they are not using encryption ($Fd5_a$), (b) users who do not password protect their devices are less likely to enable remote wipe ($Fd5_b$), (c) users who do not password protect their devices tend not to use the remote locator ($Fd5_c$), (d) users tend to have the security mechanisms remote wipe and remote locator both disabled ($Fd5_d$), and (e) users who do not use smartphone secsoft are less likely to encrypt the data ($Fd5_e$). Moreover, our results revealed additional occasions where *NSECSAV* users were exposed to physical access, namely: (f) users tend to disable both device encryption and remote device locator ($Fd5_f$), (g) users who do not password protect their device are less likely to use smartphone secsoft ($Fd5_g$), and users who were ignorant about smartphone secsoft were found not to h) password protect their devices ($Fd5_h$), and i) use encryption ($Fd5_i$).

The analysis revealed cases where the impact of unauthorized access attacks in the *NSECSAV* groups, increases (*Fd6*) since users who: (a) do not encrypt their data tend to store personal data in their devices ($Fd6_a$), (b) store personal data are less likely to scrutinize security messages ($Fd6_b$), (c) store personal data are less likely aware of smartphone malware ($Fd6_c$), and (d) do not enable remote device locator tend to have misplaced their device in the past ($Fd6_d$).

Our analysis revealed that users who scrutinize security messages opt for physical security controls (*Fd7a*), since: a) users in *both groups* who scrutinize security messages tend to encrypt their data (*Fd7a*) and (b) *SECSAV* users who scrutinize security messages tend to password protect their smartphone (*Fd7b*). Thus, these users have an encouraging security posture against unauthorized access.

**Trust in the App Repository.** The analysis in [18] indentified trust in the app repository as a severe vulnerability, as it lowered the sample's security posture. This analysis revealed that this trust may expose users to malware/greyware residing in the repository [6,23] (*Fd8*). This is so, since: (a) users in *both groups* who are unaware of smartphone malware tend to trust the app repository ($Fd8_a$), (b) *SECSAV* users who trust the repository tend to occasionally inspect the prompted security messages ($Fd8_b$), and (c) *NSECSAV* users who trust the repository are less likely to search the app repository for free secsoft ($Fd8_c$). The impact in case a *NSECSAV* user installs malware/greyware from the repository increases, because *NSECSAV* users who trust the repository tend to store personal data in their devices ($Fd6_e$).

Furthermore, the results suggest that *NSECSAV* users trust the app repository, but tend to be unaware whether apps are securely analyzed during their submission (*Fd9*). In this context, the results suggest that the trust in the repository does not stem from the perceived efficiency of the app vetting mechanism. On the other hand, our analysis revealed some *NSECSAV* users who are not complacent with security provided by the app repository and try to amend it. This is so, as users who do not trust the app repository tend to search in it for free secsoft (*Fd10*).

**Inspection of Security Messages.** The security models of smartphone platforms assume that users scrutinize any messages that they prompt, in order to make informed security decisions. Our results suggest that in *both groups*, users exist who scrutinize

security messages and tend to scrutinize agreement messages (*Fd11*). Although this suggests that these users are concerned about their security and privacy, they are the sample's minority according to the analysis in [18].

**Table 2.** Findings from correlations between responses

| # | **Finding** (*Fd*) | **SECSAV** | **NSECSAV** |
|---|---|---|---|
| | ***Adoption of smartphone security software*** | $(x^2,p,\varphi)_1,..,(x^2,p,\varphi)_n$† | $(x^2,p,\varphi)_1 \ldots (x^2,p,\varphi)_n$ |
| **1.** | Users ignore endpoint security | $(4.863,0.027,0.156)_a$,$(5.654,0.017,-0.168)_b$, $(4.056,0.044,-0.142)_c$ | $(4.885,0.027,0.138)_a$, $(6.735,0.009,-0.162)_b$ |
| **2.** | Minority of users opt for endpoint security | $(8.344,0.004,0.37)_a$, $(8.337,0.004,0.204)_b$ | $(7.975,0.005,0.199)_a$, $(7.594,0.006,0.172)_C$ |
| **3.** | Users who inspect app's reputation tend to not regard smartphone secsoft essential | - | (4.450,0.035,-0.131) |
| | ***Adoption of physical controls*** | $(x^2,p,\varphi)_1,..,(x^2,p,\varphi)_n$ | $(x^2,p,\varphi)_1 \ldots (x^2,p,\varphi)_n$ |
| **4.** | Users are exposed to unauthorized remote access | $(5.770;0.016;0.170)_a$,$(14.305,p<0.001,-0.267)_b$, $(6.491, 0.011, 0.180)_c$, $(5.931,0.015, 0.172)_d$ | $(15.978,p<0.001,0.249)_a$,$(5.112,0.024,-0.141)_b$, $(5.878,0.015, -0.151)_e$ |
| **5.** | Users are exposed to unauthorized physical access | $(5.4145,0.020,0.165)_a$,$(10.525, 0.001,0.229)_b$, $(8.776,0.003, 0.209)_c$, $(34.333, p<0.001, 0.414)_d$, $(5.770,0.016,0.170)_e$, | $(11.984,0.001,0.216)_a$,$(6.828, 0.009,0.163)_b$, $(9.959,0.002, 0.196)_c$, $(137.857, p<0.001, 0.731)_d$, $(15.978,p<0.001, 0.249)_e$, $(7.731,0.005,0.173)_f$, $(8.292,0.004, 0.179)_g$, $(13.001, p<0.001,0.224)_h$, $(4.472,0.034, 0.132)_i$ |
| **6.** | Increased impact of unauthorized access | - | $(4.593,0.032,-0.133)_a$, $(6.889, 0.009,-0.163)_b$, $(6.643,0.010,-0.160)_c$,$(7.755,0.005,-0.173)_d$, $(8.768,0.003,0.184)_e$ |
| **7.** | Users who scrutinize security messages opt for physical security controls | $(19.715, p<0.001,0.314)_a$, $(8.085,0.004,0.201)_b$ | $(17.352, p<0.001, 0.259)_a$, |
| | ***Trust in app repository*** | $(x^2,p,\varphi)_1,..,(x^2,p,\varphi)_n$ | $(x^2,p,\varphi)_1 \ldots (x^2,p,\varphi)_n$ |
| **8.** | Users are exposed to malware indexed in the app repository | $(5.035,0.025,-0.159)_a$, $(9.824, 0.002,0.222)_b$, | $(9.167,0.002,-0.188)_a$, $(5.337,0.021,-0.144)_c$ |
| **9.** | Users trust the app repository but are unaware of app testing | - | (5.057,0.025, 0.140) |
| **10.** | Uses who do not trust app repository tend to search in it for free secsoft | - | (5.337,0.021,-0.144) |
| | ***Inspection of security messages*** | $(x^2,p,\varphi)_1,..,(x^2,p,\varphi)_n$ | $(x^2,p,\varphi)_1 \ldots (x^2,p,\varphi)_n$ |
| **11.** | Minority of users scrutinize prompted messages | $(21.480, p<0.001,0.328)_a$ | $(40.338, p<0.001,0.395)_a$ |
| **12.** | Users ignore prompted messages | $(22.688,0.000,0.337)_a$,$(12.195, p<0.001,-0.247)_b$, | $(34.262,0.000,0.364)_a$,$(20.452,p<0.001,-0.282)_b$, |
| **13.** | Users who ignore security messages are uncertain if apps undergo security analysis | - | (7.297,0.007, 0.168) |
| | ***User's background influences their security*** | $(x^2,p,\varphi)_1,..,(x^2,p,\varphi)_n$ | $(x^2,p,\varphi)_1 \ldots (x^2,p,\varphi)_n$ |

**Table 2.** (*continued*)

| | | | |
|---|---|---|---|
| **14.** | User's security posture is influenced by IT skills | (6.029,0.014,-0.174)$_a$, (4.538, 0.033,0.151)$_b$, (12.122, 0.000,0.246)$_c$ | (12.272,p<0.001,-0.218)$_a$, (8.471,0.004, 0.181)$_b$, (5.853, 0.016,0.151)$_d$,(8.839,0.003,-0.185)$_e$, (8.628,0.003, 0.183)$_f$, (12.093,0.001,0.217)$_g$,(12.051,0.001,0.216),2)$_h$,(16.502,p< 0.001,0.253)$_i$ |
| **15.** | Users are ignorant of smartphone secsoft and smartphone malware | (41.976, p<0.001, 0.458)$_a$ | (31.629, p<0.001,0.350) $_a$ |
| **16.** | Users disregard security notions in the smartphone app ecosystem | - | (7.297,0.007,0.168)$_a$, (10.377,0.001,0.201)$_b$, (19.988, p<0.001,-0.278)$_c$ |
| | ***Other findings*** | $(x^2,p,\varphi)_1$,..,$(x^2,p,\varphi)_n$ | $(x^2,p,\varphi)_1 \ldots (x^2,p,\varphi)_n$ |
| **17.** | The device is used for personal and business purposes | (19.682, p<0.001,0.314) $_a$ | (27.087, p<0.001,0.324) $_a$ |
| **18.** | Users disregard security indicators during app selection | (5.283,0.022,-0.163) $_a$ | (4.584,0.032,-0.133)$_a$, (11.194,0.001,-0.208)$_b$ |

† $x^2$ test of independence value, significance level, association's direction (positive, negative).

On the other hand, there are users in *both groups* who ignore prompted messages (*Fd12*). Namely: (a) there are users who ignore both security and agreement messages ($Fd12_a$) and (b) users who occasionally inspect security messages are less likely to scrutinize agreement messages ($Fd12_b$). Also, *NSECSAV* users who ignore security messages tend to be uncertain if apps are securely tested during their submission (*Fd13*). This indicates that the ignorance of security messages is not a result of the trust in the efficiency of the app testing mechanism.

**User's Background Influences Their Security Posture.** The results suggest that user's security posture is influenced by their IT skills (*Fd14*). Firstly, in *both groups*: (a) technically savvy users are less likely to trust the app repository ($Fd14_a$), and (b) users who scrutinize security messages tend to be technically savvy ($Fd14_b$). In *SECSAV*, (c) non-technically savvy users tend to occasionally inspect security messages ($Fd14_c$). In *NSECSAV* the results suggest that: (d) non-technical savvy users tend to be unsure if the repository's apps are security analyzed ($Fd14_d$), whereas (e) technically savvy ones are less likely to respond that they do not know if apps undergo security analysis ($Fd14_e$). Also, (f) non-technical savvy users are less likely to search the repository for free secsoft ($Fd14_f$) and (g) users who inspect app reviews tend to be technically savvy ($Fd14_g$). Finally, participants who use: (h) remote wipe, and (i) remote locator, tend to be technically savvy, ($Fd14_h$) and ($Fd14_i$) respectively.

Furthermore, the results revealed users in *both groups* who are ignorant of smartphone secsoft and smartphone malware (*Fd15*). This is alarming, since in both groups there are users exposed to smartphone malware/greyware.

Finally, our results suggest that *NSECSAV* users disregard important aspects of security in the smartphone app ecosystem (*Fd16*), as users who are unaware if app testing happens in the app repository tend to: (a) ignore security messages

($Fd16_a$), (b) ignore agreement messages ($Fd16_b$), and (c) be unaware of smartphone malware ($Fd16_c$).

**Other Findings.** The results revealed users in *both groups* who store business data in their device and also tend to store personal data in it (*Fd17*). This suggests that the device is being used for business and personal purposes, thus the impact of unauthorized access increases.

In addition, the results suggest that users disregard security indicators during app selection (*Fd18*), as (a) users in *both groups* who consider an app's usefulness tend to disregard its reputation score during app selection ($Fd18_a$) and (b) *NSECSAV* users who consider an app's usefulness tend to disregard its reviews during app selection ($Fd18_b$). The analysis in [18] revealed usefulness as the most popular app selection criterion and the above negative correlations suggest that users tend to disregard the two risk signals.

## 5 Limitations

A limitation of our study is that it may be affected by the sample's demographics. As the majority of respondents are Android or iOS users, male, and aged [15-30], the results may be biased towards the population's device, gender, or age. During our analysis we found only a few statistically significant differences in the responses between users of different gender and platform. Also, users aged [15-30] tend to be the early adopters of technology and, hence, we consider that our findings provide considerable insight about smartphone security awareness. This is validated from the common findings regarding the smartphone security awareness reported in related surveys, which had different demographics.

Also, our data collection relies on self-reported statistics. For instance, we asked users to classify their IT skills and if they possessed a security background [21] (at minimum comprehension of the notion of threat, risk, safeguard) - deriving from a source either academic or industrial. Although we decided to avoid a direct validation of these responses, to keep the length of the instrument short and avoid the user's fatigue, our discussion with the user ensured the validity of responses, as well as the comprehension of the questions and technical terms.

Moreover, during the discussion the researchers were cautious not to reveal that the survey's purpose was to measure the user's security awareness, as this would inflate their responses. This was the reason why in the survey's beginning we claimed to conduct a survey about smartphone app usage. The findings showed that we successfully avoided such misleading behaviors, since the majority of the collected answers are alerting in terms of the user's security awareness. Furthermore, the instrument included control questions (i.e. Q5, Q7, Q17-Q18), to ensure that the researcher did not accidentally overlook any ambiguous responses from outliers during the interview. Each time a control question identified an outlier during the analysis, the relevant data were excluded, thus leaving us with 458 cases.

Another limitation is that the survey is cross-platform, thus, it suffers from the heterogeneity of each platform's security controls. In this context, a security control can be provided as a platform service (e.g. remote wipe in iOS), whereas it may require

the installation of a third-party app in others (e.g. in remote wipe Android). Software controls exists in some platforms that protects selected device assets (e.g. corporate documents). This software may be available either as a standalone app or as part of a Mobile Device Management solution. Moreover, the restrictions of the platform's security model may hinder a security control's availability (e.g. antivirus in iOS). In such cases, we excluded the relevant smartphone population from the analysis.

This survey does not focus on the above mentioned details regarding security controls. It examines whether users adopt security controls, without focusing on the implementation details or the software origin (i.e. third-party or offered by the platform), finding that they are not applied by the majority of participants.

## 6 Discussion and Conclusions

The popularity of smartphones and their respective apps, which are being downloaded from 'app repositories' or 'app markets', is increasing fast. This popularity has drawn the attention of both attackers, who attempt to impair the security and privacy of smartphone users via malicious apps, and, relevant research that focuses on the protection of smartphones. Currently, the security awareness of smartphone users is not adequately explored in literature. Nonetheless, users' role in the smartphone ecosystem is crucial, since they are often delegated by the platforms to make informed decisions that impact their security and privacy. This is, in particular, true in case someone uses a smartphones in order to access critical applications or infrastructures, etc. [13, 15].

In this paper, we extended our work on smartphone security awareness by examining whether a security background, deriving from sources either academic (e.g., university security courses) or industrial (e.g., information security certifications), affects the survey participants' security awareness. For doing so, we split the sample in two groups, comprising of security savvy users and non-security savvy users and then examine and compare their security posture.

Our results suggest that the participants' security background has slight impact on their awareness about the security and privacy in the smartphone ecosystem.

Firstly, users ignored endpoint security even in cases where they were uncertain about centralized security (i.e. app testing in the repository). We discovered occasions that even security savvy users were exposed in unauthorized physical access, which is a critical finding due to the device's small size and mobility. Furthermore, we found users in both groups who were exposed to smartphone malware/greyware, which is indexed in the app repository, due to their trust to the app repository, disregard of security messages and in some cases due to the unawareness of the threat itself.

Our results also suggest that current security indicators that are being used by app repositories are ineffective. Therefore they must be redesigned and the users must be trained to use them.

Overall, our results provide proof that smartphone users require awareness training specifically tailored for smartphone security. To this end, the current common body of knowledge for the security domain (e.g. [21]) must be extended to include the necessary background (e.g., impact of authorization decisions via security messages, unique attacks, significance in the adoption of security controls, etc.) to enable smartphone

users cope with the challenges in the smartphone ecosystem. We plan to extend the common body of knowledge in our future work.

**Acknowledgment.** This work has been co-funded by the European Union (European Social Fund, ESF) and Greek national funds, through the Operational Program *Education and Lifelong Learning* of the National Strategic Reference Framework (Program Heraclitus II: Investing in Knowledge Society through the ESF).

The authors would like to thank A. Kastania (AUEB) for her valuable suggestions.

## References

1. Questionnaire, `http://www.aueb.gr/users/amylonas/instrument.pdf`
2. Barrera, D., Van Oorschot, P.: Secure Software Installation on Smartphones. IEEE Security & Privacy 9(3), 42–48 (2011)
3. Chia, P., Yamamoto, Y., Asokan, N.: Is this App Safe? A Large Scale Study on Application Permissions and Risk Signals. In: Proceedings of the 21st International World Wide Web Conference, pp. 311–320 (2012)
4. Chin, E., Felt, A.P., Sekar, V., Wagner, D.: Measuring User Confidence in Smartphone Security and Privacy. In: Proceedigns of the 8th Symposium on Usable Privacy and Security (2012)
5. Egele, M., Kruegel, C., Kirda, E., Vigna, G.: PiOS: Detecting privacy leaks in iOS applications. In: Proceedigns of the 18th Annual Network and Distributed System Security Symposium (2011)
6. Enck, W., Octeau, D., McDaniel, P., Chaudhuri, S.: A study of Android application security. In: Proceedings of the 20th Usenix Security Symposium (2011)
7. Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E., Wagner, D.: Android Permissions: User Attention, Comprehension, and Behavior. In: Proceedigns of the 8th Symposium on Usable Privacy and Security (2012)
8. Felt, A.P., Wang, H.J., Moshchuk, A., Hanna, S., Chin, E.: Permission Re-Delegation: Attacks and Defenses. In: Proceedings of the 20th Usenix Security Symposium (2011)
9. Felt, A., Chin, E., Hanna, S., Song, D., Wagner, D.: Android permissions demystified. In: Proceedings of the 18th ACM Conference on Computer and Communications Security, pp. 627–638. ACM, USA (2011)
10. Flick, U.: An introduction to qualitative research. Sage Publications, London (1998)
11. Grace, M., Zhou, Y., Wang, Z., Jiang, X.: Systematic Detection of Capability Leaks in Stock Android Smartphones. In: Proceedings of the 19th Network and Distributed System Security Symposium (2012)
12. Hornyack, P., Han, S., Jung, J., Schechter, S., Wetherall, D.: These aren't the droids you're looking for: Retrofitting Android to protect data from imperious applications. In: Proceedings of the 18th ACM Conference on Computer and Communications Security, USA, pp. 639–652 (2011)
13. Iliadis, J., Gritzalis, D., Spinellis, D., Preneel, B., Katsikas, S.: Evaluating certificate status information mechanisms. In: Proceedings of the 7th ACM Computer and Communications Security Conference, pp. 1–8 (2000)
14. Kelley, P., Consolvo, S., Cranor, L., Jung, J., Sadeh, N., Wetherall, D.: A Conundrum of Permissions: Installing Applications on an Android Smartphone. In: Blyth, J., Dietrich, S., Camp, L.J. (eds.) FC 2012. LNCS, vol. 7398, pp. 68–79. Springer, Heidelberg (2012)

15. Lekkas, D., Gritzalis, D.: Long-term verifiability of healthcare records authenticity. International Journal of Medical Informatics 76(5-6), 442–448 (2006)
16. Mylonas, A., Dritsas, S., Tsoumas, B., Gritzalis, D.: Smartphone Security Evaluation: The Malware Attack Case. In: Proceedings of the 8th International Conference of Security and Cryptography, pp. 25–36 (2011)
17. Mylonas, A., Dritsas, S., Tsoumas, B., Gritzalis, D.: On the Feasibility of Malware Attacks in Smartphone Platforms. In: Obaidat, M.S., Sevillano, J.L., Filipe, J. (eds.) ICETE 2011. CCIS, vol. 314, pp. 217–232. Springer, Heidelberg (2012)
18. Mylonas, A., Kastania, A., Gritzalis, D.: Delegate the smartphone user? Security awareness in smartphone platforms. Computers & Security 34, 47–66 (2013)
19. Mylonas, A., Tsoumas, B., Dritsas, S., Gritzalis, D.: A Secure Smartphone Applications Roll-out Scheme. In: Furnell, S., Lambrinoudakis, C., Pernul, G. (eds.) TrustBus 2011. LNCS, vol. 6863, pp. 49–61. Springer, Heidelberg (2011)
20. Shabtai, A., Fledel, Y., Kanonov, U., Elovici, Y., Dolev, S., Glezer, C.: Google Android: A Comprehensive Security Assessment. IEEE Security Privacy 8(2), 35–44 (2010)
21. Lambrinoudakis, C., Gritzalis, D., Tsoumas, V., Karyda, M.: Secure electronic voting: The current landscape. In: Gritzalis, D. (ed.) Secure Electronic Voting, pp. 110–122. Kluwer (2003)
22. Theoharidou, M., Mylonas, A., Gritzalis, D.: A risk assessment method for smartphones. In: Gritzalis, D., Furnell, S., Theoharidou, M. (eds.) SEC 2012. IFIP AICT, vol. 376, pp. 443–456. Springer, Heidelberg (2012)
23. Zhou, Y., Jiang, X.: Dissecting Android malware: Characterization and evolution. In: Proceedings of 33rd IEEE Symposium on Security and Privacy, pp. 95–109 (2012)

# Trustworthy Selection of Cloud Providers Based on Security and Privacy Requirements: Justifying Trust Assumptions

Michalis Pavlidis[1], Haralambos Mouratidis[1], Christos Kalloniatis[2], Shareeful Islam[1], and Stefanos Gritzalis[3]

[1] School of Architecture, Computing and Engineering, University of East London, U.K.
m.pavlidis@ieee.org, {haris,shareeful}@uel.ac.uk
[2] Cultural Informatics Laboratory, Dept. Of Cultural Technology and Communication, University of the Aegean, Greece
chkallon@aegean.gr
[3] Laboratory of Information and Communication Systems Security, Dept. of Information and Communications Systems Engineering, University of the Aegean, Greece
sgritz@aegean.gr

**Abstract.** Cloud computing is a new paradigm with a promising potential. However, issues of security, privacy, and trust raise concerns and discourage its adoption. In previous work we presented a framework for the selection of appropriate cloud provider based on security and privacy requirements criteria. However, the adoption of cloud includes release of control over valuable assets, which constitutes trust in the cloud provider of paramount importance. In this paper we extend the framework by incorporating trust and control concepts in its language and adding a new activity to properly identify and reason about trust assumptions during the selection of appropriate cloud provider. Also, the CASE tool was extended to support the new activity. A case study is used to illustrate the usefulness of our approach.

**Keywords:** Cloud Computing, Security, Privacy, Requirements, Trust, Control.

## 1 Introduction

Cloud computing is an evolving paradigm that is radically changing the way humans store, share and access their digital files. Its promise is the introduction of a rapid elastic and unlimited computation, storage, and bandwidth with a significant lower cost. However, to fully realize the potential of the cloud, appropriate security and privacy solutions must be adopted. Many organisations and individuals are still avoiding cloud services mostly because they are not sure if the services provided, by various providers, are suitable for their security and privacy requirements [1]. This is especially true since organisations and individuals would have to hand in their personal and organizational data into service providers over which they have no control.

S. Furnell, C. Lambrinoudakis, and J. Lopez (Eds.): TrustBus 2013, LNCS 8058, pp. 185–198, 2013.

It is therefore important, that appropriate software engineering techniques must be developed to support the structured and systematic identification of security and privacy requirements that an organization might have for their systems and based on those requirements to support selection of appropriate cloud services. However, and despite the recent research interest in developing software engineering techniques to support systems based on the cloud, the literature fails to provide a systematic and structured approach that enables software engineers to identify security and privacy requirements and select a suitable service provider based on such requirements.

To this end, in previous work [2] we proposed a novel framework to support elicitation of security and privacy requirements and selection of a service provider based on those requirements. The framework consists of a modeling language, a process and a tool. The analysis performed by that framework, trusts that the cloud provider will deliver the required security and privacy mechanisms needed for the identified security and privacy requirements. However, blind trust is not ideal, but trust should be supported by appropriate justification. We want to be able to feel confident, in as higher degree as possible, that the cloud provider will deliver as promised and reasonably rely on them to care for our valuable assets. In order to be able to understand that, we need to clearly understand the relevant underlying trust assumptions, make them explicit and justify them.

The work presented in this paper, extends our previous work to address the above challenges and to support justified trust assumptions through a systematic trust based process. In other words, we want to support the decision making process by identifying underlying trust assumptions and justifying the trust that we place on cloud providers. The language is extended with trust and control concepts, new activities are added into the process to identify direct and indirect trust relationships, and also the tool is extended to support the activities. The rest of the paper is organized as follows. Section 2 presents an overview of our previous work. In section 3 we present the extended framework that incorporates the trust process. An illustration of the framework is presented in section 4 using a case study while section 5 presents the related work and section 6 concludes the paper.

## 2 Background Information on the Framework

The framework we already presented in [1] consists of a language and a process that is focused on the requirements engineering stage. The language employs concepts from the requirements, security and privacy engineering domains, and it is based on our previous work on security requirements engineering, and in particular Secure Tropos [3] and privacy requirements engineering, and in particular PRiS [4]. However, the language is enriched with new concepts, such as cloud actor, measure, and mechanisms, which are necessary to support the selection of cloud providers. The process supported by the framework is iterative and it is based on the development of a set of models that are incrementally refined to include further details. It provides a structured way of eliciting and analysing security and privacy requirements, identifying relevant security and privacy mechanisms and of selecting an appropriate cloud service provider based on these mechanisms. It comprises of three main activities: *the Security*

*and Privacy Cataloguing*, the *Security and Privacy Analysis,* and the *Selection of Cloud Service Provider*. Each one of these activities has specific inputs and it results in specific outputs. The first two activities enable developers to understand the security and privacy requirements of the system and identify relevant security and privacy mechanisms that the cloud providers should deploy to support the identified security and privacy requirements. Once all the security and privacy mechanisms have been identified, the third activity supports the selection of an appropriate service provider based on the degree of satisfaction of these mechanisms by potential cloud providers. Our framework makes use of an analysis technique based on an independent probabilistic model, which uses the measure of *satisfiability* [5]. In our work, satisfiability represents the probability that the security and/or privacy mechanism will be satisfied. Thus, the evaluation results in contribution relationships from the cloud provider to the probability of satisfying the security and/or privacy mechanisms of the system identified in the previous activity of our process.

To express the contribution of each provider to the satisfiability of each security/privacy requirement of the system, a weight is assigned. Weights take a value between 0 and 1. The allocation of such weights is performed by the security, privacy and cloud experts after studying the required security and privacy mechanisms and the various characteristics and provisions that a potential cloud provider has in place to support these mechanisms. The overall satisfiability level is calculated by summing up all the satisfiability values of an individual cloud provider and dividing that sum by the number of security and privacy mechanisms required by the system. The cloud provider with the highest satisfaction level is the preferred provider.

The framework is supported by a tool that has been developed based on the Open Models Initiative ADOxx Platform (www.openmodels.at). The tool provides an environment for developers to create a number of diagrams that support the described process. In particular, the process described in the previous section results in the development of four artefacts represented in terms of four diagrams. These are the *Security and Privacy Reference Catalogue Diagram*, the *Security and Privacy Organisational Diagram*, *the Security and Privacy System Requirements Diagram* and the *Cloud Provider Selection Diagram* respectively.

## 3 Framework Extension

The above described framework, helps to select among potential cloud providers based on the probabilities. However, trust is more than subjective probabilities [6], and the selection of a cloud provider should not only be based on calculation of probabilities. Even, if there is a probability that the cloud provider has the capability to support the required security and privacy mechanisms it does not mean that this will happen. What is required is a structured process that can reveal underlying trust relationships, reason about them and enable their justification.

In previous work [7] we have presented a process for trustworthy information systems development that uses a language [8] based on trust and control concepts. We incorporate this work into the framework described in the previous section, to enhance its cloud provider selection activity by considering trust relationships. In

particular, the extension of the framework is threefold. The language is extended with trust and control concepts, new activities are added into the process to identify direct and indirect trust relationships, and also the tool is extended to support the new activities.

### 3.1 Language Extension

The language has been enhanced with the following trust-related concepts that allow a better understanding of the factors that affect confidence: **Resolution.** Resolution of a dependency is the indication of how the uncertainty in the fulfillment of a dependency is removed in order to build confidence in the dependency. It is necessary to be identified as a dependency implies a vulnerability for the depender because the dependee might not fulfill the dependency. There are two types of resolution, i.e., trust and control, that can be identified to feel confident in the fulfilment of a dependency. Also, there can be more than one resolution. **Trust.** Trust is the positive expectation of one actor about the behaviour of another actor by whom she/he might be positively or negatively affected [9]. In the context of a dependency, the depender is the trustor and the dependee is the trustee. There are four types of trust resolution:

- **Experiential Trust.** Experiential trust is trust that originates from previous direct experience with the trustee. The depender then is actually depending on himself and there can be only one instance of experiential trust, as there is only one instance of someone's self.
- **Reported Trust.** Reported trust is trust that originates from a third party (the reporter) who reports that the trustee is trustworthy. Therefore the depender depends on the reporter to trust the dependee. There can be more than one of reporters who are reporting whether the dependee is trusted. Apart from human the third party can also be a system, such as a reputation system.
- **Normative Trust.** Normative trust is trust that originates from the system environment norm. The depender is then depending on the environment norm. There can be only one environment norm.
- **External Trust.** External trust is trust that originates from sources outside of the system environment. These for example can be government bodies. The depender is then depending on an external source of trust. There can be more than one external sources of trust.

**Trust Relationship.** Trust relationship is defined as a relationship that exists between the trustor and the trustee and resolves a dependency based on trust. There are two types of trust relationship, i.e., direct and indirect. Direct trust relationship is the trust relationship that exists between the two actors of a dependency and it is not implied by any other trust relationship. Indirect trust relationships are trust relationships that are implied by direct trust relationships or control relationships and need to exist in order to support them. **Control.** Control is the power that one actor has over another actor. It helps to build confidence in another actor. Control specifies the ability of an actor to gather information about another actor in order to decide whether to execute an action. In addition, control specifies the action that is required for the dependee to

behave in an expected way. So, to achieve control, an actor needs to ensure observation and deterrence capabilities. **Entailment.** Entailment is a condition of trust that is required to be valid for having confidence in the dependency from which it is required. For example, if there is a reported trust resolution then it requires the entailment "the reporter is trusted" to be valid. Also, if there is a control resolution then it requires the entailment "the controller is trusted" to be valid. Such assumptions of conditions of trust require evidence in order to be justified.

### 3.2 Process Extension

An extension has been applied also on the framework's process. In particular, during activity 3 "Selection of cloud service providers", new steps have been added to identify resolutions and entailments, and examine the validity of the entailments. Figure 1 shows the updated activity, using the Software & Systems Process Engineering Metamodel Specification (SPEM). In particular, for each candidate cloud provider a resolution and entailment diagram needs to be constructed. The diagrams enable the identification of indirect trust relationships and the reasoning of them. A resolution can be trust or control and if it is trust resolution, then it can have single or multiple types of trust. Depending on the type of trust, new dependencies may be introduced. So, a reported-based trust resolution creates a new dependency that needs to be resolved. The new dependency is on the reporter. While the other three types of trust resolution, i.e., experiential, normative and external, do not introduce new dependency. If the resolution is control-based then it introduces a new dependency in a similar way as the reported trust resolution. But, this time the dependency is on the controller. Therefore, whenever there are new dependencies created by a reported trust resolution or a control resolution the activity has to be applied again in order to identify resolutions for the new dependencies. At the end there is a list of resolutions that show why the cloud adopter is confident in the fulfilment of the security and privacy mechanisms and a resolution diagram that graphically shows the resolutions in order to allow better understanding and analysis.

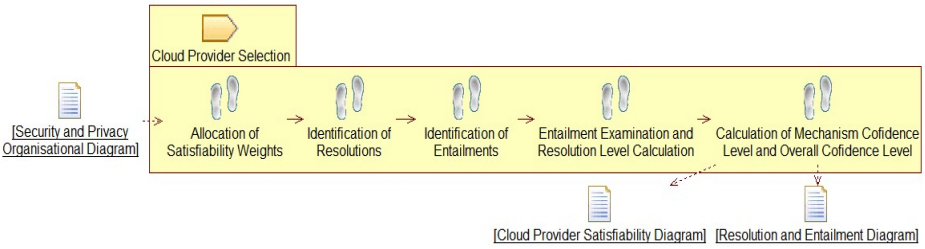



**Fig. 1.** Extended process definition in SPEM

The next step identifies and analyses entailments, which are the trust conditions that need to be in place to justify trust relationships. This step starts by identifying entailments based on the resolutions identified from the previous step. Therefore resolution diagrams are necessary to identify the entailments. Entailments can be

identified based on the following five cases and graphically represented in an entailments diagram that shows from which resolution they originated: i) Control based resolution requires an entailment that the controller is trusted; ii) Experiential trust requires an entailment that the trustor can trust himself; iii) Reported trust requires an entailment that the reporter is trusted; iv) Normative trust requires an entailment that the environment norm is trusted; v) External trust requires an entailment that the external source of trust is trusted.

At this stage, evidence is collected in order to validate the entailments. However, not all entailments may be valid due to lack of evidence or conflicting evidence. Then the resolution level of a dependency on a cloud provider is calculated by dividing the number of valid dependency entailments with the number of all identified dependency entailments.

$$Dependency\ Resolution\ Level = \frac{Number\ of\ Valid\ Entailments}{Total\ Number\ of\ Entailements}$$

Then summing up all Resolution levels $RL$ multiplied with the Satisfiability levels $SL$ and dividing that sum with the overall number of security and privacy mechanisms $m$ calculate the overall score of a single cloud provider. At the end the provider with the highest overall score level is selected.

$$Overall\ Score = \frac{\sum_{x=1}^{m} RL_x \times SL_x}{m}$$

### 3.3 Tool Extension

The tool was extended to support the creation and analysis of diagrams related to the trust analysis (Figure 2). In particular, the following diagrams are now supported by the tool: **Resolution diagram**. This diagram graphically shows the resolutions of the dependencies on the cloud providers for the provision of the identified mechanisms. Also, it shows the indirect trust relationships that are implied from the existence of direct trust relationships. **Entailment diagram**. This diagram graphically shows the entailments and from which resolutions originate. Also, it contains a list of valid entailments, which contains the conditions of trust that are true and a list of invalid entailments, which contains the conditions of trust that are not true and as a result further actions are required if the particular cloud provider is selected.

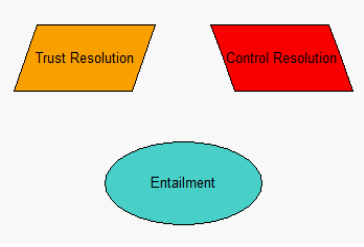



**Fig. 2.** Trust and control graphical notation

## 4 Case Study

In our previous work [1] the framework was applied on a real-world case study based on the development of a cloud based solution for the domain of Electronic-Point-Of-Sale (EPOS). The case study reported on a project that took place between the School of Architecture, Computing and Engineering at the University of East London and a company specialising at the provision of EPOS solutions[1]. EPOS Ltd depends on the Cloud Provider to *Provide EPOS Software as Service*, *Manage EPOS Software Licencing* and *Provide Cloud Services*. Figure 3 illustrates the partial result of the analysis that took place as part of that project. In particular, Figure 3 focuses on one of the EPOS Solutions goals, i.e. *Provide EPOS Software as Service* and on two security constraints (Ensure Availability of Software, Ensure Data Confidentiality) and one privacy constraint (Ensure Data Residency) related to that goal. For each one of these constraints, relevant security and privacy measures and mechanisms were identified as shown in the diagram.

Based on the set of security and privacy mechanisms identified, the next activity aims to evaluate how specific service providers satisfy the security and privacy mechanisms identified in the previous step. In the rest of the case study the focus is on five of the security and privacy mechanisms. During the project discussed above, our analysis consisted of the evaluation of three cloud providers[2]. The outcome was the Satisfiability diagram shown in Figure 4.

Following the new activities and language extensions described in the previous section, we have enhanced the analysis of the case study to consider trust relationships during the selection of the cloud provider. For each of the three cloud providers a resolution and entailment diagram is constructed. To keep the length of the paper to a minimum, we have combined in our illustration the resolution and entailment diagrams for each of the cloud providers.

The combined diagram for Cloud provider 1 (CP1) is shown in Figure 5. There are a number of dependencies on the cloud provider to provide the five mechanisms. In particular, the dependencies for *Log Data*, for *Pseudonymisation*, and for *ACID* are resolved by *Reported Trust*. The reporter is the *University Partner 1*, who reports that CP1 is trusted for the provision of *Log Data*, *Pseudonymisation*, and *ACID* respectively. Nevertheless, as stated in the previous section, reported trust resolutions create new dependencies. The new three dependencies are on the *University Partner 1* who is reporting that CP1 can be trusted for the provision of the three mechanisms respectively. Therefore, new resolutions need to be identified for the new dependencies. The resolutions of the new dependencies are *Experiential Trust* as there is previous direct experience with the *University Partner 1*. However, the remaining two dependencies on CP1 for the provision of *VM Isolation* and *Data Tokenization* could not be resolved.

---

[1] For confidentiality reasons we are not allowed to disclose the name of the company so we use the name "EPOS Ltd" to refer to it throughout the paper.

[2] For confidentiality reasons, we are not able to reveal the true identities of the analysed cloud providers. We report however, the real satisfability scores.

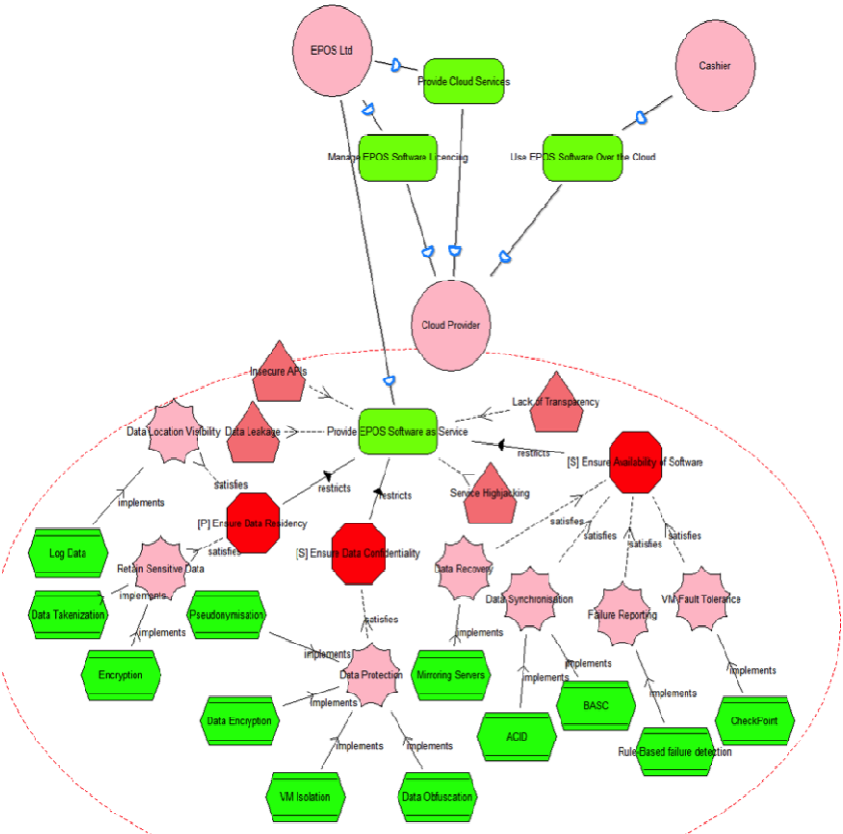



**Fig. 3.** Security and privacy analysis diagram

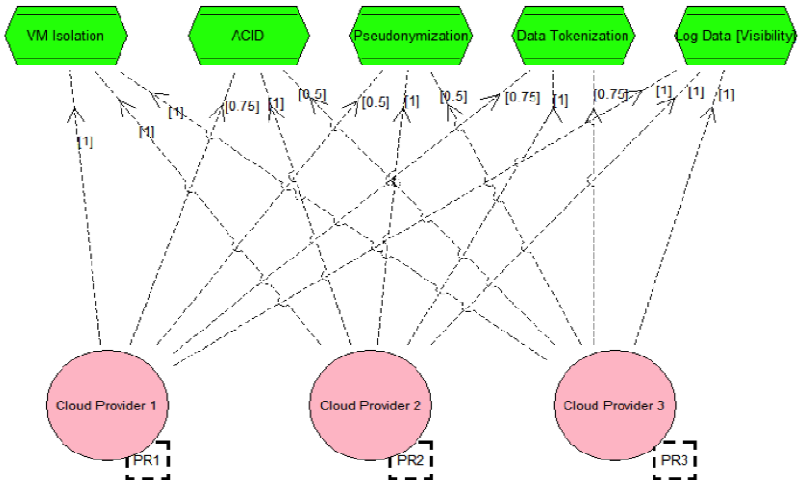



**Fig. 4.** Satisfiability diagram

Each identified resolution requires an entailment. Therefore, the entailments are identified based on the rules described in the previous section. The entailments that we need to trust ourselves for what the *University Partner 1* is reporting are valid, as there is a long history of collaboration, which make the specific dependencies on the CP1 resolved. Since, there is only one resolution for each of these dependencies with valid entailments then their resolution level is 1. On the other hand, the resolution level of the dependencies without resolutions, and therefore without valid entailments, is 0.

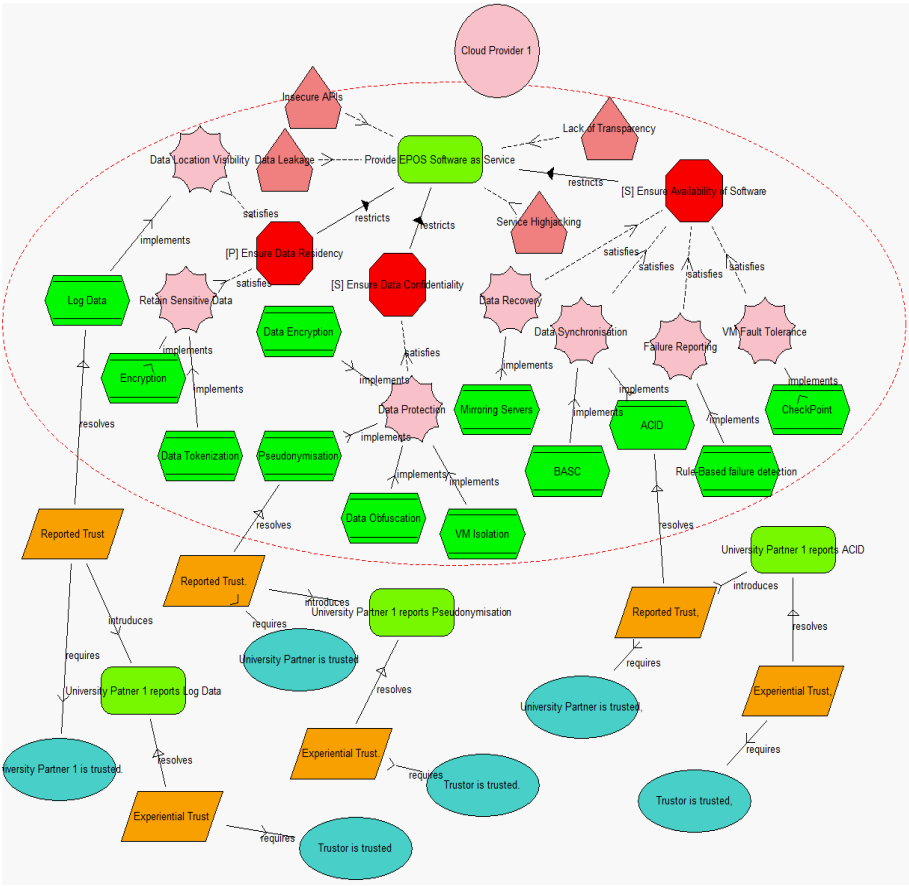



**Fig. 5.** Resolution and entailment diagram for cloud provider 1

Next, the analysis of the resolutions and entailments of the second cloud provider is carried out and it is shown in figure 6. CP2 is being audited regularly and audits are a form of control on cloud providers as they monitor their performances and services provision. Therefore, all five dependencies on CP2 are resolved through *Audit Control*. Control resolutions though, introduce new dependencies on the controller to successfully audit the cloud provider. The controller in this case is a third party who is performing the audits, and the dependencies on it introduce new uncertainty. These

new dependencies though could not be resolved. However, there is another source of trust to resolve the initial dependencies, which is *Normative Trust*.

Normative trust requires an entailment that the environment norm is trusted. In fact, CP2 is a company of high reputation with a large list of clients. Therefore, the entailments that the norms are trusted are valid. Since, there is one valid and one invalid entailment for the resolutions of each of the dependencies on CP2, the resolution level of those dependencies is 0.5 for each one of them.

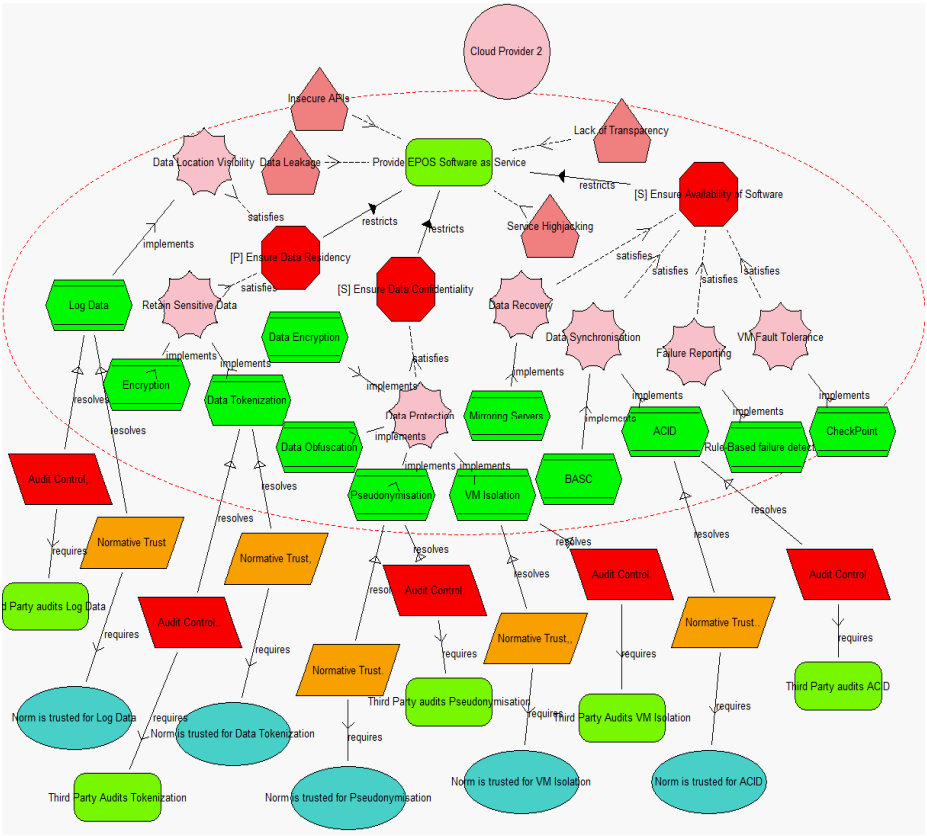



**Fig. 6.** Resolution and entailment diagram for cloud provider 2

Similarly the resolution and entailment diagram for Cloud Provider 3 (CP3) is constructed and shown in figure 7. CP3 is also under audit checks, but these checks are limited to the provision of *Data Tokenization*, *Pseudonymisation*, and *VM Isolation*. The existence of control resolutions has as result the introduction of new dependencies. We depend on a third party, as the controller, to control the cloud provider, i.e. perform audits checks. Again, though, no resolutions of the dependencies on the third party could be found. However, reported trust resolutions of the dependencies on CP3 for the provision of *Pseudonymisaton* and *VM Isolation* were identified. The reporter is another university partner, named for the purpose of

our explanation as *University Partner 2*. As said before, reported trust resolutions create new dependencies on the *University Partner 2*, who is reporting that the CP3 can be trusted for the provision of *Pseudonymisation* and *VM Isolation*. *Experiential Trust* is then identified to resolve these new dependencies.

The reported trust resolutions require an entailment that the *University Partner 2* is trusted for what is reporting, while the experiential trust resolutions require entailments that we can trust ourselves. Again, there is a long history of collaboration with the *University Partner 2* to trust our judgment, which makes the corresponding entailments valid. As a result, the entailments the University Partner is trusted for reporting that CP3 provides *Pseudonymisation* and *VM Isolation* are valid. The resolution level of the dependencies without resolutions or with resolutions but with out valid entailments is 0, while the resolution level of the dependencies for the provision of *Pseudonymisation* and *VM Isolation* is 0.5 as each one of them has one resolution with valid entailment and one with invalid entailment.

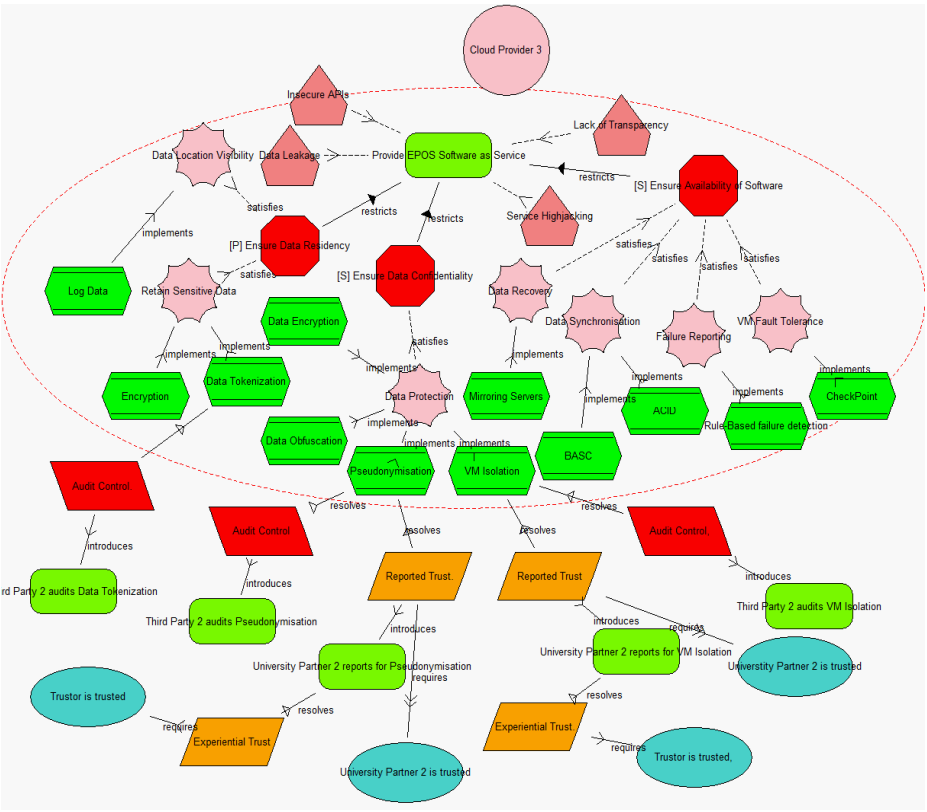



**Fig. 7.** Resolution and entailment diagram for cloud provider 3

Once the resolution levels of the dependencies on the cloud providers and their Satisfiability levels have been calculated then we can follow the steps of the selection activity, as described in the previous section. For each of the providers we take each

dependency for the provision of a mechanism and multiply its Resolution Level with its Satisfiability Level. Then their sum is divided with the number of mechanisms in order to produce an overall score for each cloud provider as shown below:

$$CP1 = (0*1+1*0.75+1*0.5+0*0.75+1*1) / 5 = 0.45$$
$$CP2 = (0.5*1+0.5*1+0.5*1+0.5*1+0.5*1+0.5*1) / 5 = 0.5$$
$$CP3 = (0.5*1+0*0.5+0.5*0.5+0*0.75+0*1) / 5 = 0.15$$

The provider with the highest score, and therefore preferred, is provider 2.

## 5 Related Work

The literature has examples of works that focus on security requirements analysis and/or privacy requirements analysis. For example, methods, such as Secure Tropos [3], SQUARE [11], and SecReq [10,12], focus explicitly on security issues, while others, such as PriS [4] and LINDDUN [13], focus on privacy issues. Most of the works related to security focus on the requirements stage. However, none of these works considers their analysis within the context of cloud computing. On the other hand, there are works [14-17] that have been developed based on the idea of cloud computing, but these mostly focus on implementation concerns related to security and privacy in the cloud, and they do not provide a methodology to support the elicitation and analysis of security and privacy requirements and the selection of an appropriate cloud provider based on such requirements. Also, the literature provides examples of works [18-20] in trust analysis on the cloud but, again, these works focus on implementation concerns and trust is considered as a narrow concept that is limited only to security or accountability among others, excluding issues such as shared interests and goodwill.

The work presented here differs from these approaches in that the proposed framework provides explicit support for elicitation and analysis of security and privacy requirements within the context of cloud computing and a systematic process to analyse trust as part of the cloud provider selection process. Assumptions about trust relationships are explicitly identified along with their underlying trust relationships. There is a systematic approach towards better understanding of why there is trust, or there is no trust, in a specific cloud provider.

## 6 Conclusion

The adoption of cloud computing imposes an unavoidable release of control over valuable assets. As a result trust in the cloud provider is required for a confident adoption of cloud computing and full utilization of its benefits. In this paper we extended previous work to incorporate a new activity that enables to identify direct and indirect trust relationships and to analyse the respective trust assumptions during the selection of a cloud provider.

By applying the extended framework on the case study we have illustrated the applicability and the benefits of our approach. In particular, we identified trust

assumptions that are underlying the successful provision of five specific security and privacy requirements by three potential cloud providers and reasoned about them. However, it does not guarantee that the requirements will be met but that there is confidence in their fulfillment and that the selection of the cloud provider has been justified. If these had been left unexamined then the selection of the cloud provider could have been wrong, as the cloud provider would not have met the security and privacy requirements that we focused on in the case study.

Future work will focus on methods that will further support the process of the validation of entailments. For instance what kind and how much evidence is required for entailments to be valid. We also plan to formalise the work and to enhance the tool to better support our framework.

## References


1. Kalloniatis, C., Mouratidis, H., Islam, S.: Evaluating Cloud Deployment Scenarios Based on Security and Privacy Requirements. Requirements Engineering Journal, REJ (2013), http://dx.doi.org/10.1007/s00766-013-0166-7
2. Mouratidis, H., Islam, S., Kalloniatis, C., Gritzalis, S.: A framework to support selection of cloud providers based on security and privacy requirements. To appear in Journal of Systems and Software (2013)
3. Mouratidis, H., Giorgini, P.: Secure Tropos: A Security-Oriented Extension of the Tropos Methodology. International Journal of Software Engineering and Knowledge Engineering 17(2), 285–309 (2007)
4. Kalloniatis, C., Kavakli, E., Gritzalis, S.: Addressing privacy requirements in system design: The PriS method. Requirements Engineering Journal 13(3), 241–255 (2008)
5. Giorgini, P., Mylopoulos, J., Nicchiarelli, E., Sebastiani, R.: Reasoning with Goal Models. In: Spaccapietra, S., March, S.T., Kambayashi, Y. (eds.) ER 2002. LNCS, vol. 2503, pp. 167–181. Springer, Heidelberg (2002)
6. Castelfranchi, C., Falcone, R.: Trust Is Much More than Subjective Probability: Mental Components and Sources of Trust. In: 33rd International Conference on System Sciences, Hawaii (2000)
7. Pavlidis, M., Islam, S., Mouratidis, H., Kearney, P.: Modeling Trust Relationships for Developing Trustworthy Information Systems. International Journal of Information Systems Modelling and Design 5(1) (2014)
8. Pavlidis, M., Mouratidis, H., Islam, S.: Dealing with Trust and Control: A Meta-Model for Trustworthy Information Systems Development. In: Sixth IEEE International Conference on Research Challenges in Information Science, Valencia, Spain (2012)
9. Mollering, G.: The Trust/Control Duality: An Integrative Perspective on Positive Expectations of Others. International Sociology 20(3), 283–305 (2005)
10. Schneider, K., Knauss, E., Houmb, S.H., Islam, S., Jürjens, J.: Enhancing Security Requirements Engineering by Organisational Learning. Requirements Engineering Journal (REJ) 17(1), 35–36 (2012)
11. Mead, N.R., Steheny, T.: Security Quality Requirements Engineering (SQUARE) methodology. SIGSOFT Software Engineering Notes 30(4), 1–7 (2005)
12. Houmb, S.H., Islam, S., Knauss, E., Jürjens, J., Schneider, K.: Eliciting Security Requirements and Tracing them to Design: An Integration of Common Criteria, Heuristics, and UMLsec. Requirements Engineering Journal 15(1), 63–93 (2010)

13. Deng, M., Wuyts, K., Scandariato, R., Preneel, B., Joosen, W.: A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requirements Engineering Journal 16(1), 3–32 (2011)
14. Smith Gillam, L., Li, B., O'Loughlin, J.: Adding Cloud Performance To Service Level Agreements. In: 2nd International Conference on Cloud Computing and Services Science (CLOSER), Portugal (2012)
15. Islam, S., Mouratidis, H., Weippl, E.: A Goal-driven Risk Management Approach to Support Security and Privacy Analysis of Cloud-based System. In: Security Engineering for Cloud Computing: Approaches and Tools. IGI Global Publication (2012)
16. Wenzel, S., Wessel, C., Humberg, T., Jürjens, J.: Securing Processes for Outsourcing into the Cloud. In: 2nd International Conference on Cloud Computing and Services Science. SciTe Press (2012)
17. Khajeh-Hosseini, A., Sommerville, I., Bogaerts, J., Teregowda, P.: Decision Support Tools for Cloud Migration in the Enterprise. In: 4th International Conference on Cloud Computing. IEEE Computer Society (2011)
18. Ko, R., Jagadprama, P.: TrustCloud: A Framework for Accountability and Trust in Cloud Computing. In: World Congress on Services (2011)
19. Peterson, G.: Don't Trust. And Verify: Security Architecture Stack for the Cloud. IEEE Security and Privacy (September/October 2010)
20. Pearson, S., Benameur, A.: Privacy, Security and Trust Issues Arising from Cloud Computing. In: 2nd IEEE International Conference on Cloud Computing Technology and Science (2010)

# Author Index