Privacy and Security Risks Analysis of Identity Management Systems

Ebenezer Paintsil

Thesis submitted to Gjøvik University College
for the degree of Doctor of Philosophy in Information Security

2013

Privacy and Security Risks Analysis of Identity Management Systems

Faculty of Computer Science and Media Technology
Gjøvik University College

Privacy and Security Risks Analysis of Identity Management Systems / Ebenezer Paintsil
Doctoral Dissertations at Gjøvik University College 1-2013
ISBN:978-82-93269-24-3
ISSN:1893-1227

Acknowledgments

Without my family, the last three years would have been unbearable. I thank my wife Alberta Selawose Paintsil, my children, Annabella and Adumadze, and my mother, Comfort Paintsil for their moral support and prayers. A part from my family, there are special people who are behind the success of this work and deserve my appreciation and recognition. Special thanks go to my first advisor Lothar Fritsch. Your valuable pieces of advice helped me to complete this work. You are patient and have incredible knowledge in the research area. I also want to thank Einar Snekkenes, my second supervisor for his valuable and constructive feedbacks. I wish to acknowledge my colleague at Norwegian Computing Center (Norsk Regnesentral) and the leadership of DART for their wonderful support. I thank all my brothers, sisters and friends especially Richard Glavee-Geo who thoroughly edited this thesis. Finally, I wish to express my gratitude to Research Council of Norway (NFR) for making the funds available for this research.

Declaration of Authorship

I, Ebenezer Paintsil, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:
(Ebenezer Paintsil)
Date: 20.08.2013

Abstract

This thesis develops a risk model and model-based risk analysis method for privacy and security risks analysis of identity management systems (IDMSs) in order to reduce cost and provide scientific support for the choice of identity management approaches. In order to analyze a system, we need a clear understanding of the system as well as what can go wrong in it. Risk assessors often rely on system specifications and stakeholders (end-users and system owners) to understand a targeted system. Similarly, system stakeholders may rely on risk assessors to understand the risk analysis process.

Model-based risk analysis methods use graphical models to assist system stakeholders to understand the risk analysis process. The graphical risk models communicate what can go wrong in a system and assist in the security risk analysis. They facilitate participation, risk communication and documentation. However, current model-based risk analysis methods provide general support for security risk analysis but pay little attention to privacy. Privacy requirements complement that of security but conflicts can arise in their implementation. Identifying and understanding such conflicts are a prerequisite for developing adequate and a balanced risk analysis method. Furthermore, due to lack of data on past events, model-based risk analysis methods either rely on subjective intuitions of risk assessors and system stakeholders, or complex mathematical validation techniques to determine a system's risk. Subjective intuitions lead to high uncertainties in risk analysis. Moreover, complex mathematical risk modeling and validation techniques are expensive, difficult to learn and can impede risk communication among system stakeholders.

This thesis develops a balanced approach to risk analysis where systems' characteristics and tools that are relatively easy to learn are relied upon to analyze privacy and security risks in IDMSs. It provides new knowledge on how to develop a privacy and security risks model for IDMSs from the characteristics of information that flow in them. Furthermore, it develops an executable model-based risk analysis method (EM-BRAM) to improve risk communication, automation, participation as well as documentation in IDMSs. The EM-BRAM relies on system behaviors or characteristics rather than data on past events or intuitions of a risk assessor to analyze privacy and security risks in IDMSs. Consequently, the method can reduce subjectivity and uncertainty in risk analysis of IDMSs. EM-BRAM identifies risk factors inherent in IDMSs and uses them as inputs for the privacy and security risks analysis. The risk factors are categorized into external and internal misuse cases. The external misuse cases consist of risk factors that may be outside the control of IDMSs while the opposite is true for internal misuse cases. The internal misuse cases are used for the privacy and security risks analysis.

In order to determine a system's risk, the EM-BRAM uses Colored Petri Nets tools and queries to model and analyze the characteristics of information flow in the target IDMS. The method has been applied to analyze the security and privacy risks of popular IDMSs such as OpenID and SAML single sign-on services for Google Apps. The results show that the EM-BRAM is effective in analyzing privacy and security of IDMSs if it is applied to low level system specifications.

Sammendrag

Denne avhandlingen lager en risikomodell og modellbasert risikoanalysemetode for privacyanalyse og sikkerhetsrisikoanalyse av identity management-systemer (IDMS-er), med det formål å redusere kostnader og gi vitenskapelig støtte for valg av identity managementtilnrminger. Modellbaserte risikoanalysemetoder kan hjelpe systemeiere med å forstå en risikoanalyseprosess gjennom metodens effektive bruk av grafiske modeller for å understtte deltakelse, kommunikasjon av risiko og dokumentasjon. Disse grafiske risikomodellene illustrerer hva som kan gå galt i systemet og hjelper med estimering av sikkerhetsrisiko.

Dagens metoder for modellbasert risikoanalyse støtter generell sikkerhetsrisikoanalyse, men tar ikke for seg privacy i nevneverdig grad. Videre, på grunn av manglende data om tidligere hendelser er disse metodene avhengige av subjektive intuisjoner fra de som analyserer risiko eller fra systemeiere eller de er avhengig av komplekse matematiske valideringsteknikker for å bestemme risiko for et system. Subjektive intuisjoner medfrer høy usikkerhet i risikoanalyse. Videre er komplekse matematiske teknikker for risikomodellering og -validering kostbare, vanskelige å lære og kan vanskeliggjøre kommunikasjon av risiko mellom systemeiere. Denne avhandlingen utvikler en balansert tilnrming til risikoanalyse der systemkarakteristikker og verktøy som gjemmer kompleks matematikk brukes for å analysere privacy- og sikkerhetsrisiko i IDMS-er. Den gir ny kunnskap om hvordan lage risikomodeller for IDMS-er fra karakteristisk informasjonsflyt i systemene. Videre lager den en utførbar modellbasert risikoanalysemetode (EM-BRAM) for å forbedre kommunikasjon, automatisering, deltakelse og dokumentasjon i risikoanalyse av i IDMS-er. For å analysere privacy- og sikkerhetsrisiko gjør EM-BRAM bruk av utførbare modeller som er relativt enkle å forstå. EM-BRAM gjør bruk av systemoppførsel og -karakteristikker i stedet for data om tidligere hendelser eller intuisjonen til de som analyserer risiko. Dette betyr at metoden kan redusere subjektivitet og usikkerhet i risikoanalyse av i IDMS-er.

Contents

List of Figures

List of Tables

Chapter 1

Part I

1.1 Introduction

Identity management is the processes and policies involved in managing the lifecycle of an identity [63]. Identity lifecycle describes the possible states of an identity. The state of an identity is "unknown" if it has not been enrolled or deleted; "established", if it is enrolled or restored from archive state; and "active" if it is in used. An identity is in the "suspended" state if it is temporarily not available for use and in the "maintenance" state, if it is being updated.

Systems that create and manage these identities are known as identity management systems (IDMSs) [69]. The traditional role of IDMSs is to issue end-users with credentials after an account is created. The end-users can then use these credentials to authenticate and request access to services or resources. IDMSs 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 for end-users after their identities have been verified. SPs may be referred to as relying parties (RPs). Examples of IDMSs are OpenID [129], Facebook [147] and MinID [35].

IDMSs are crucial to protecting privacy and security online because they collect and manage personal data. Poor identity management could contribute to the rising privacy violations and identity theft crimes online [105], [155], [96]. Currently, a number of organizations including the Norwegian center for information security (NorSIS) [104] and Australian Privacy Commission [105] have noticed the rising trend of privacy violations and identity theft crimes online. In Spring 2007, NorSIS established the ID-tyveri unit to reduce identity theft and identity related crimes in Norway. In the same year, the office of the Australian Privacy Commissioner (OAPC) [105] commissioned a survey to study identity related crimes in Australia. The survey revealed that 50% of Australians were concerned about their privacy online. 9% of Australians claim to have been victims of identity theft and 60% are concerned about becoming a victim. 45% believes identity theft is likely to occur as a result of using the Internet.

We can reduce privacy risk and its related crimes online with the help of privacy enhancing IDMSs [169], [57]. Privacy enhancing mechanisms of current IDMSs can reduce the number of credentials end-users have to remember. This in turn can reduce the end-users' mental load, rampant loss of credentials, high recovery cost and privacy risks as they subscribe to several services online [69]. In addition, they can enable end-users to act under pseudonyms; and to control the use and release of their digital identities. Current privacy enhancing IDMSs can facilitate successful service delivery in both organizations and government institutions [96]. Organizations can gain competitive advantage and reduce financial losses if they can enhance identity management. Similarly, governments can gain the trust of their citizens if they can do the same.

However, current IDMSs greatly differ in their privacy enhancing or security mechanisms [96], [57]. They prescribe different security mechanisms and focus on different problems areas. These inconsistent and complex mechanisms make system evaluation and selection a daunting task for system stakeholders – end-users and system owners.

Privacy and security risks analysis can provide a means of evaluating such complex and

inconsistent systems [91]. In addition, it can assist system stakeholders to choose a privacy enhancing IDMS based on their privacy preferences and create awareness of their privacy and security risks as they use a particular IDMS.

Moreover, privacy and security risks analysis of IDMSs can be mandated by law especially when the IDMSs manage sensitive information such as health data and financial records [81]. Above all, security and privacy of IDMSs can be enhanced if their privacy and security risks are known.

Risk analysis consists of risk identification and estimation [62]. Risk estimation involves assignment of risk values. A risk analysis process rather than the assignment of risk values is the most important part of risk analysis [25]. Model-based risk analysis methods have been proposed as one of the ways of improving risk analysis process [91], [58], [71]. They rely on graphical models to facilitate participation, risk communication and documentation.

However, current model-based risk analysis methods provide general support for security risk analysis but pay little attention to privacy. The adequacy and suitability of a risk analysis method depend on the objectives and purposes of the analysis [91]. In addition, privacy requirements complement that of security but conflicts can arise in their implementation [177]. Identifying and understanding such conflicts are a prerequisite for developing adequate and a balanced risk analysis method.

Besides, the cost of a security risk analysis method depends on the expertise required to implement the method [161]. A method designed for system stakeholders to operate could be cost effective while those requiring expertise are costly. For example, CORAS [91] is expensive because it involves experts in getting information from information technology (IT) and business unit staff members to model and analyze risk [161]. Moreover, CORAS requires over 150 man-hours of risk analysis to be cost effective [91].

Furthermore, due to lack of data on past events, model-based risk analysis methods either rely on subjective intuitions of risk assessors and system stakeholders, or complex mathematical validation techniques to determine a system's risk [13]. Subjective intuitions lead to relative risk estimation and high uncertainties in risk analysis [25], [13]. Notwithstanding, complex mathematical model-based risk analysis methods such as UMLSec [71] are expensive because of their learning curve and the expertise required. Additionally, they can impede risk communication among stakeholders [52]. A relatively easy to understand privacy and security risks analysis method could reduce cost involved in the analysis.

1.2 Problem Statement

Current risk analysis methods for IDMSs are mainly qualitative, rely on manual inspections and incomplete because the stakeholders' interests are ignored [22], [166]. For example, the methods proposed by Delft, and Oostdijk [34], Gajek et al. [43], and Gross [46] are ad-hoc and rely on manual inspections. Moreover, the metric-based framework proposed by Cabarcos et al. [22] provides no support for system stakeholders to understand the risk analysis process. In spite of their limitations, model-based risk analysis methods can be improved to address the current risk analysis challenges in IDMSs better than the existing methods. This thesis investigates a less subjectivity and relatively easy to learn model-based risk analysis method for privacy and security risks analysis in IDMSs.

1.3 Objective of the Thesis

The overall objective of this thesis is to investigate how a privacy and security risks model and method can be developed for IDMSs in order to reduce cost and provide scientific support for the choice of identity management approaches. In order to achieve this objective, this thesis investigates privacy and security risks analysis requirements and applies them

to create artifacts for IDMSs. An artifact is a term for describing something that is artificial, or constructed by humans, as opposed to something that occurs naturally [137].

1.4 The Scope of the Thesis

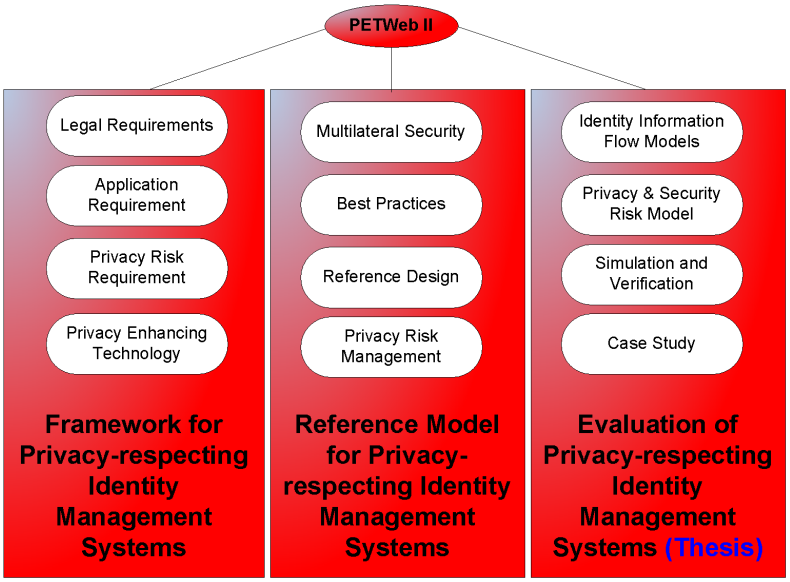

Figure 1.1: The Petweb II Project

This thesis is part of the PETWeb II project. PETWeb II project aims at providing scientific support for the choice of identity management approaches, in particular by supporting the analysis of specific technical and regulatory risks relating to the choice of an identity management approach. The choice of an identity management approach has severe consequences on the way the information system can use, store, combine and misuse personal data.

Figure 1.1 represents the three research areas of the PETWeb II project. The framework and reference model research areas investigate and develop a framework and reference model for privacy-respecting IDMSs respectively. The evaluation of the privacy-respecting IDMSs research area investigates and develops a method for risk analysis in IDMSs.

As part of the bigger PETWeb II project, this thesis focuses on the evaluation of the privacy-respecting IDMSs. The evaluation is intended to rely on the analysis of information that flow in IDMSs. Information that flow in IDMSs are organized and communicated in a form of tokens [170]. Tokens are technical artifacts providing assurance about identities [117], [127]. They can communicate authentication entitlement and attribute information in IDMSs [170]. They are personal data sources and gateways to end-users' personal data [103]. A token may consist of an attribute such as username, a claim such as a password, an assertion such as SAML token, a credential such as a X.509 certificate or a combination of these.

1.5 Research Questions

This thesis investigates the following research questions:

1. **Question I: What are the challenges and trade-offs in the information security risk assessment approaches?**

   The relevant requirements and methods for addressing the challenges in a research area can facilitate the understanding of the research problems in question [51]. This question investigates the challenges in security risk assessment discipline and methods for addressing them. In addition, it investigates the trade-offs that result from the applications of these methods. The outcome of the investigation is applied to scope the thesis.

2. **Question II: Can characteristics of information that flow in identity management systems contribute to privacy and security risks?**

   Risk analysis consists of risk identification and estimation [62]. Risk identification finds, lists and characterizes elements of risk or risk factors. The risk factors can then be used for risk modeling and estimation [102]. Information that flow in IDMSs can provide meaningful insights into the overall system security and privacy. Yet the contributions of the characteristics of information that flow in IDMSs to privacy and security risks have not been the main focus of research. This question investigates whether the characteristics of information flow can contribute to privacy or security risk in IDMSs and whether they can be used to create a risk model for IDMSs.

3. **Question III: Given the characteristics of information that flow in identity management systems, in what way can a risk analysis model be developed for them?**

   Risk analysis models are very important in security analysis [40]. They support the risk analysis and guide the subsequent iteration of the entire risk analysis process. Risk models "define the key terms used in risk assessment including the risk factors to be assessed and the relationships among those factors" [102]. They give a clear idea of what we want to mitigate and the technical capabilities of adversaries.

   Currently, executable privacy and security risks analysis models for IDMSs are missing in literature. The existing risk models are tabulated [62], [156]. A table of risk factors is easy to create but it is often imprecise and hides important information among irrelevant details. This impedes early discovery of risk elements [156], [167]. A graphical risk model can produce a precise documentation; in which information is structured and presented at an appropriate level of abstraction for communication among all system stakeholders [91].

   This question investigates a modeling technique for privacy and security risks analysis of IDMSs. In addition, it investigates how to apply the technique to develop a risk model for IDMSs.

4. **Question IV: To what extent can we assess privacy and security risks of identity management systems with the characteristics of information that flow in them?**

   The extent to which information flow can be used to automate privacy and security risks analysis in IDMSs has not been the main focus of risk analysis research. Current risk analysis methods for IDMSs rely on manual inspections to analyze a system's risk [11], [96], [166]. Moreover, the metric-based framework developed by Cabarcos [11] defines a set of metrics for risk quantification in IDMSs but it does not target

information flow [11]. This question investigates how characteristics of information flow can be used for privacy and security risks automation in IDMSs.

1.6 Research Questions and their Relationships

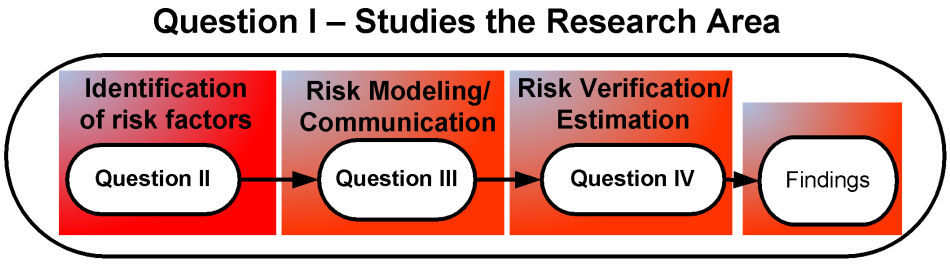

Figure 1.2: Research Questions and their Relationships

Figure 1.2 shows the research questions and their relationships, and the task for each question. Question I is the literature review of the research area. It encompasses other research questions. Question II and Question III focus on risk identification [62], and risk modeling respectively. Finally, Question IV focuses on risk analysis of IDMSs. The outcomes of the research are the findings. Each question builds on the previous question.

1.7 List of Publications for the Thesis

1. **Article I** [114]: Ebenezer Paintsil, **Taxonomy of Security Risk Assessment Approaches for Researchers**, in Proceedings of 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN). Article I answers Question I **(See Section 2.4 for details)**.

2. **Article II** [117]: Ebenezer Paintsil and Lothar Fritsch, **A Taxonomy of Privacy and Security Risks Contributing Factors**, in Privacy and Identity Management for Life, vol. 352 of IFIP Advances in Information and Communication Technology. Springer Boston, 2011, pp. 52-63. Article II answers Question II **(See Section 2.5 for details)**.

3. **Article III** [115]: Ebenezer Paintsil, **Evaluation of Privacy and Security Risks Analysis Construct for Identity Management Systems**, in IEEE Systems Journal, VOL. 7, Issue 2, (2013). Article III answers Question II and evaluates the taxonomy in Article II **(See Section 2.5.1 for details)**.

4. **Article IV** [113]: Ebenezer Paintsil, **A Model for Privacy and Security Risks Analysis**, in New Technologies, Mobility and Security (NTMS), 2012 5th International Conference, May (2012). Article IV answers Question III **(See Section 2.6 for details)**.

5. **Article V** [118]: Ebenezer Paintsil and Lothar Fritsch, **Executable Model-Based Risk Analysis Method for Identity Management Systems: Using Hierarchical Colored Petri Nets**, in Springer LNCS 2013 Proceedings of the 10th international conference on Trust, privacy and security in digital business. Article V answers Question IV **(See Section 2.7 for details)**.

6. **Article VI** [116]: Ebenezer Paintsil, **Towards Automation of Risk Analysis in Identity Management Systems**, in Trust, Security and Privacy in Computing and Communications (TrustCom2013) IEEE CPS 12th International Conference. Article VI answers Question IV **(See Section 2.7 for details)**.

1.8 List of Additional Publications

This section lits additional work in the research area but the results are outside the principal scope of this thesis.

- **Article VII** [111]: Ebenezer Paintsil, **Towards Legal Privacy Risk Assessment and Specification**, in Proceedings of the 8th international conference on Trust, privacy and security in digital business (Berlin, Heidelberg, 2011), Springer-Verlag, pp. 174-185.

- **Article VIII** [112]: Ebenezer Paintsil, **Executable Model-Based Risk Assessment Method for Identity Management Systems**, in proceedings of Norwegian Information Security Conference (NISK 2012), Tapir Akademisk Forlag publishers, ISBN 978-82-321-0012-5.

- **Article IX** [119]: Ebenezer Paintsil and Lothar Fritsch, **Executable Model-Based Risk Assessment Method for Identity Management Systems**, in proceedings of IFIP Advances in Information and Communication Technology, Springer 2013 (Position Paper).

1.9 Summary of Contributions

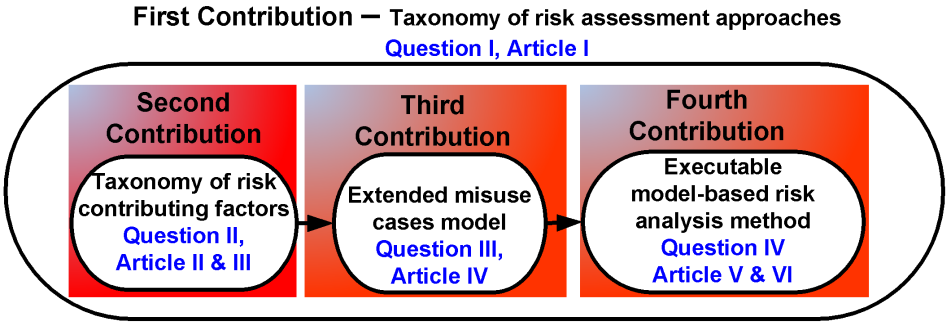

Figure 1.3: Research Contributions and their Relationships

This section discusses the contributions of this thesis. Contributions are based on the development of the three main artifacts – construct, model and method. Figure 1.3 depicts how the contributions build on each other. The first contribution encompasses all the contributions and it is based on research question I and Article I [114]. The second contribution is based on research question II, Article II [117], and III [115]. The third contribution builds on the second contribution and it is based on question III and Article IV [113]. Finally, the fourth contribution builds on the third contribution and it is based on question IV, Article V [118] and VI [116]. The summary of the contributions are explained as follows:

- **Question I: What are the challenges and trade-offs in the information security risk assessment approaches?** (Addressed in Section 2.4)

  - The first contribution is the development of a taxonomy of risk assessment approaches that can guide researchers to create innovative artifacts.

    An artifact is innovative if it is better or relevant than existing ones [90], [51]. The literature survey of risk assessment approaches and their challenges revealed that no critical analysis that focuses on the research needs has been performed

yet. The taxonomy in Article I [114] focuses on challenges of risk assessment approaches and how to guide a security risk assessment research. It clearly identifies the challenges of the various risk assessment approaches and proposes a criterion that can guide researchers to develop innovative artifacts for security and privacy risks assessment. It shows the possible trade-off in the risk assessment approaches and how to classify risk assessment methods with the taxonomy.

- **Question II: Can characteristics of information that flow in identity management systems contribute to privacy and security risks?** (Addressed in Section 2.5)

  – The second contribution is the development of a taxonomy of risk contributing factors for IDMSs. The risk contributing factors can be employed to create risk model for IDMSs.

  Information that flow in IDMSs can provide meaningful insights into the overall system security and privacy. Knowledge on how a risk model can be developed from the characteristics of information flow in IDMSs are missing in literature. This thesis contributes by providing new knowledge on how the characteristics of information flow can contribute to privacy and security risks in IDMSs. It then develops a taxonomy of risk contributing factors based on these characteristics. The taxonomy can be employed to create a risk model for IDMSs.

- **Question III: Given the characteristics of information that flow in identity management systems, in what way can a risk analysis model be developed for them?** (Addressed in Section 2.6)

  – The third contribution is the extension of the misuse cases modeling approach to create privacy and security risks model for IDMSs. In addition, the model is refined and implemented with colored Petri nets (CPNs) [77] in order to represent or communicate dynamic system behaviors to all system stakeholders.

  Use and misuse cases (UMCs) modeling approach has a substantial advantage over other modeling approaches because it is simple, user-friendly and has substantial industrial application. In addition, UMCs models are best at capturing and analyzing security problems even though they do not explicitly specify them [14]. However, UMCs models lack privacy and security risks analysis constructs and they are informal. In order to model privacy and security risks of IDMSs, an extended misuse cases (EMCs) model is developed. The EMCs model extends the UMCs model with privacy and security concepts. The model is then refined and implemented with CPNs. The EMCs modeling approach can enable communication of both static and dynamic behaviors of threat agents to all system stakeholders.

- **Question IV: To what extent can we assess privacy and security risks of identity management systems with the characteristics of information that flow in them?** (Addressed in Section 2.7)

  – Finally, an executable model-based risk analysis method is developed for privacy and security risks analysis of IDMSs.

  The extent to which model-based risk analysis methods can enhance privacy and security in IDMSs has not been the main focus of research. The available risk analysis methods for IDMSs do not rely on executable system model to analyze risk. They rely on manual inspections to determine risk in IDMSs. The executable model-based privacy and security risks analysis method developed in this thesis creates an executable system model of a target IDMS and validates its correctness before risk analysis. The executable model enable simulation,

debugging of the target IDMS and provides tool support for the risk analysis in IDMSs. The tool support or the automation can reduce cost, time spent in the analysis, increase accuracy and above all enhance security and privacy in IDMSs. The method is relatively simple, reduces subjectivity, cost and can enable system stakeholders to understand the risk analysis process.

1.10 Structure of the Thesis

This thesis consists of eleven chapters. Chapter 1 is the main introduction for the thesis consisting of background information, problem statement, the research objective and scope of this thesis. It also presents the research questions and their relationships, the publication list and the summary of contributions. Chapter 2 and Chapter 3 discuss the research methodology with the main results and the state of the arts respectively. The rest of the chapters are the articles which constitute the main study of this thesis.

Chapter 2

Research Methodology

This section discusses the important research results and explains the design science research (DSR) methodology [122] applied in this thesis. The DSR methodology seeks to extend the boundaries of human and organizational capabilities by creating innovative artifacts. An artifact is a term for describing something that is artificial, or constructed by humans, as opposed to something that occurs naturally [137]. An artifact is innovative if it is better or relevant than existing ones [90], [51]. The DSR methodology bridges the gap between engineering and behavioral science.

The fundamental principle of DSR is that "knowledge and understanding of a design problem and its solution are acquired in the building and application of an artifact" [133]. Design is often a complex process, and designing artifacts is further challenged by the need to be innovative in an environment or domain where existing theories are often insufficient [51]. Scientific progress is achieved in design science when existing technologies are replaced by more effective and relevant ones [90], [51].

Research Activities / **Research Outputs**	Build	Evaluate	Theorize	Justify
Constructs				
Model				
Method				
Instantiation				

Table 2.1: Design Science Paradigm [90]

Table 2.1 illustrates the two DSR paradigms as proposed by March and Smith [90]. In Table 2.1, the products of DSR are the research outputs. They are constructs, models, methods and instantiations. The constructs are concepts, terms or vocabularies for describing and specifying solutions for a problem within a discipline [90]. They form the specialized language and shared knowledge of a discipline or sub-discipline. Models use constructs to represent a real world situation. They are abstractions and representations. "Models aid problem and solution understanding and frequently represent the connection between problem and solution components enabling exploration of the effects of design decisions and changes in the real world" [51]. Methods are algorithms, guidelines or practices for performing a task. An instantiation is the implementation of an artifact in its environment [90]. It operationalizes constructs, models, and methods. An instantiation can be a prototype or a specific working system or some kind of tool. These research outputs are known as artifacts.

The research activities include build, evaluate, theorize and justify. The purpose of the build and evaluate paradigm is to create innovative artifacts that serve human purposes [90]. The build and evaluate belong to the engineering paradigm while the "theorize" and "justify" belong to the behavioral science research paradigm [51]. The purpose of the "theorize" and "justify" is to understand the reality that would aid the building of innovative artifacts for a particular environment. The reality can be physical, biological, social and behavioral. The environment refers to people, organization and technology [51].

Evaluation of artifacts is an important requirement of the DSR methodology. Evaluation provides evidence that a new technology developed in DSR achieves the purpose for which

it was designed [109]. Evaluation substantiates that a developed artifact will be useful for solving problems or making improvements. Similar to the justification of theories in the natural science, evaluation determines the utility of an artifact. Thus, how well an artifact performs or how much progress we have made [90].

2.1 Justification for Using Design Science Research Methodology

Natural science research is a formal activity that builds and tests theories [134]. Theories are knowledge obtained by directly confronting the natural world. Science "makes progress because of its systematic method, and because that method allows the natural world to play a role in the evaluation of theories" [134]. The purpose of scientific method is to ensure consistency in scientific research. It enables different scientist to perform an experiment similarly, agree on important questions and considerations, and accept or reject the same hypotheses when considering the same evidence.

The activities that underline scientific methods are systematic observation, measurement, experiment, as well as the formulation, testing, and modification of hypotheses [110]. These activities may be guided by research methodologies. A research methodology is "a system of principles, practices, and procedures applied to a specific branch of knowledge" [37]. An example of research methodology is action research [150]. Action research prescribes the principles and procedures for doing a research.

There are two main research methods or methodologies–quantitative and qualitative [99], [32]. Qualitative research methods are designed to help researchers understand social and cultural context within which people live, what they say and do [99] while quantitative research methods are designed for researchers to understand natural phenomena.

The general belief in qualitative research is that the goal of understanding a phenomenon from the point of view of the participants and its particular social and institutional context is largely lost when textual data are quantified [99]. While qualitative research is useful for developing theories about human the quantitative research focuses on developing theories about natural world or phenomena.

However, information systems (IS) research is an applied discipline which relies on theories from economics, natural sciences, computer sciences, and the social sciences, to solve problems at the intersection of IT and organizations [122]. In addition, it embraces engineering discipline because of its applied nature. Engineering disciplines "accept design as a valid and valuable research methodology because the engineering research culture places explicit value on incrementally effective applicable problem solutions" [122].

The main advantage of the DSR methodology over other research methodologies is its ability to combine the engineering, human and natural research methods to solve problems at the intersection of IT and organizations. Furthermore, DSR has become one of the popular research methodology in the IS research [59], [55]. Since its introduction, the DSR has seen an increasing interest within the IS research community especially after Henver et al.'s work [51] in 2004. Since 2004, prestigious IS journals and conferences have launched special issues on DSR to analyze and suggest how to improve the DSR [59]. DSR has been applied in many successful IS research such as legal risk management [88], business model ontology [108] among others.

2.2 Research Method

Figure 2.1 depicts the research method for this thesis. The first phase is the research preparation phase. It investigates privacy and security risks analysis approaches and their implications for risk analysis in IDMSs. The main activity in this phase is literature review.

The second phase is the risk identification phase. This phase investigates the constructs for risk analysis in IDMSs.

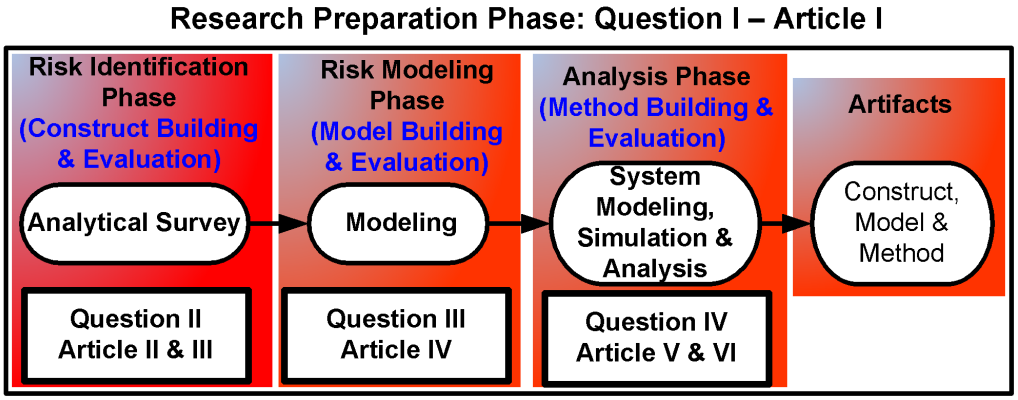

Figure 2.1: Research Phases

The third phase is the risk modeling phase. It focuses on creating a risk analysis model for IDMSs.

The fourth phase is the analysis phase. This phase investigates a method for analyzing privacy and security risks in IDMSs. The main activities involved in this phase are system modeling, simulation and analysis. The output of each phase is the input for the next phase.

Each phase is mapped to the DSR activities. The risk identification, risk modeling and the analysis phases map to the construct, model and method building and evaluation respectively in the DSR methodology.

Question I corresponds to the first phase in Figure 2.1 and it is investigated in Article I [114]. Article I analyzes the challenges in the information security risk assessment discipline and the trade-offs in the methods for addressing them.

Question II is investigated in the risk identification phase and its corresponding articles are Article II [117], and III [115]. Article II is a literature survey of factors that can contribute to privacy and security risks in IDMSs. The result is a taxonomy of privacy and security risks contributing factors. Article III [115] validates and evaluates the set of constructs in Article II [117] with Delphi study.

Question III is investigated in the risk modeling phase and it corresponds to Article IV [113]. Article IV develops an extended misuse cases (EMCs) technique for privacy and security risks modeling. The technique is applied to analyze the results in Article III in order to introduce a risk analysis model for IDMSs.

Question IV corresponds to the analysis phase in Figure 2.1 and it is investigated in Article V [118] and Article VI [116]. Article V and VI introduce the executable model-based risk analysis method for IDMSs.

Finally, the contributions of the thesis are artifacts which are the constructs, a risk model and method for privacy and security risks analysis in IDMSs.

2.3 Introduction to the Main Results

The following sections present the summary of results obtained in this thesis. They are organized according to the research phases in Figure 2.1. In accordance with DSR methodology [122], each section states a research question, its objective and briefly justifies the value of answering the question. In addition, each section introduces the artifact created to solve the problem posed by the research question. The demonstration and evaluation of the main artifacts are discussed in Section 2.8.

2.4 Research Preparation Phase

Question I: What are the challenges and trade-offs in the information security risk assessment approaches?

The relevant requirements and methods for addressing the challenges in a discipline can facilitate the understanding of a research problem and its relevance. This question investigates the challenges in the information security risk assessment discipline and the methods for addressing them. In addition, the question investigates the trade-offs that result from the application of these methods. The question is investigated in Article I [114] and the main result is the taxonomy in Figure 2.2.

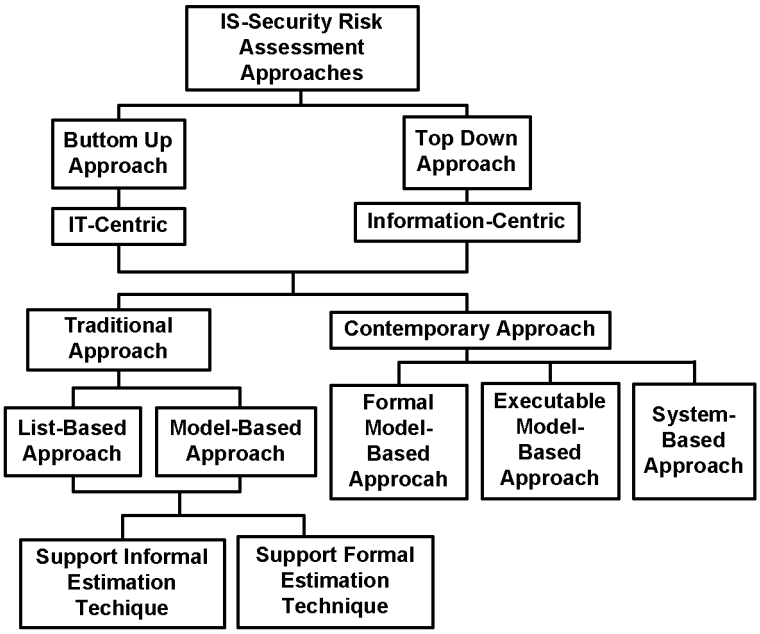

Figure 2.2: Taxonomy of Risk Assessment Approaches – Article I [114]

In Figure 2.2, the top down approach determines security requirements and verifies the violations of those requirements [44]. The security risk is requirement-oriented and may be defined as a violation of a security property [166].

On the other hand, the Bottom Up approach is based on the traditional risk assessment approaches where asset protection rather than protection goals is the main focus. Security risk is defined as a function of the likelihood of an unwanted incident and its consequence for an asset or a group of assets [91]. This asset-oriented definition has no explicit relationship with security requirements or properties.

The Top Down approach directly relates to Information-Centric risk assessment approach while the Bottom Up approach directly relates to IT-Centric risk assessment approach. IT-Centric refers to physical asset protection while information-centric refers to information security.

The IT-Security or Information-Centric approaches are subdivided into "Contemporary" and "Traditional" approaches. The traditional IT-Security risk assessment approaches use three main estimation methodologies (qualitative, quantitative and semi-quantitative) [62], [13] to estimate risk of a targeted system. Due to lack of exhaustive public and internal data on past events [149], [146], the traditional risk assessment approaches rely on the intuitions of risk assessors or system stakeholders to estimate risk [13]. This makes the

estimated risk relative and highly uncertain [25]. On the hand, the "Contemporary" approaches depend on the characteristics of system elements or system behaviors to estimate risk. Data is obtained by modeling, testing or formally analyzing the security properties of the targeted system. This makes the risk analysis less subjective.

The "Traditional" approach is subdivided into two categories – list-based and model-based categories. The list-based approach relies on standards, checklists, best practices, compliance checks or recommendations [60], [152] to estimate the security risk of a targeted system. The recommendations, standards or checklists guide the risk assessor through the risk assessment process. The risk assessor estimates the risk by comparing the targeted system and/or procedures with the given standard, recommendations or checklist. The risk assessment process involves no explicit creation of conceptual system model and may rely on the intuitions of the risk assessors. Examples of such approaches are the AICPA/CICA Privacy Risk Assessment Tool [3] and PIA [60].

The list-based approaches are ambiguous and hide important information among irrelevant details thereby impeding early discovery of risk elements [156], [167]. This makes the risk assessment process costly and error prone. On the contrary, the model-based approaches require explicit risk models. The models enhance risk identification and communication among system stakeholders [58]. They employ semi-formal graphical models to conceptually model the targeted system before risk estimation and evaluation. An example of model-based methods is CORAS [58].

The "Traditional" risk assessment approaches may be prone to errors because of the possible intuitive inputs in the risk estimation process. To minimize the uncertainty in the estimated risk as a result of the subjective inputs, we employ formal estimation techniques or methods such as Bayesian network model (BNM) [28], Game Theory and Monte Carlo simulation [30]. In this case, the "Traditional" approach is said to support formal estimation technique. The "Traditional" approaches that support none of these formal techniques or methods are said to support informal estimation techniques. Such methods may involve simple mathematics that does not minimize the uncertainty in the estimated risk. In addition, they may rely on an expected value matrix for the risk estimation [161]. An example is the Risk IT [64] risk assessment method.

The "Contemporary" approach is subdivided into three categories – formal model-based, executable model-based and system-based approaches. The formal model-based approach stresses on understanding of the targeted system [125]. It involves abstract system modeling and verification. The system's performance as a result of the verifications determines the risk but not the subjective knowledge of a risk assessor. An example of formal model-based approach is the UMLsec [71] risk verification approach.

The formal model-based approach involves formal modeling and verification. The formal modeling is mathematically based. However, mathematically based formal methods are expensive, have difficult learning curves, and can hamper effective risk communication among non-technical stakeholders [77]. Hence, in contrast to the formal model-based approaches, the executable model-based approaches that rely on easy to learn modeling techniques may be preferred. They hide the complex mathematics from non-technical stakeholders in order to enhance communication among all system stakeholders. An example of the executable model-based approach is the colored petri nets (CPNs) approach by Stephenson [148].

The executable model-based approaches may be able to address the system fidelity problem in the formal model-based approach because we can formally and visually check the behavior correctness of the targeted system model with the help of the system stakeholders [77]. The executable model-based approaches can also simulate the dynamic system environment in order to understand the dynamic behavior of the targeted system.

The formal modeling approaches rely on abstract system models for the risk estimation. It is difficult to ensure the fidelity of a system with abstract models [125]. Hence, the system-based approach focuses on how to understand the real system without any

abstraction. In other words, it verifies or tests a concrete system rather than an abstract model of it. It requires a prototype or concrete system implementation. An example of the system-based approach is emulation techniques [141]. Emulation techniques can be used for monitoring a system behavior without modifying it. It is preferred to simulation when evaluating real world system but complex, less scalable and time consuming.

The traditional approaches are highly subjective and may affect the accuracy of risk analysis in IDMSs. Notwithstanding, the contemporary approaches may be complex and expensive because of their learning curve and the expertise required. Additionally, they can impede risk communication among stakeholders [52]. However, the executable model-based approaches may be relatively easy to understand and could reduce cost involved in risk analysis of IDMSs. Hence, based on this taxonomy, this thesis investigates how privacy and security risks in IDMSs can be analyzed with the executable model-based risk analysis approach.

2.5 Risk Identification Phase

Question II: Can characteristics of information that flow in identity management systems contribute to privacy and security risks?

This phase maps to the Construct Building and Evaluation in the DSR [90]. It identifies factors that can contribute to privacy and security risks of IDMSs. Information that flow in IDMSs can provide meaningful insights into the overall system security and privacy. However, the contributions of the characteristics of information that flow in IDMSs to privacy and security risks have not been the main focus of current research. Consequently, this section analyzes how the characteristics of information that flow in IDMSs contribute to privacy and security risks. Detailed answers can be found in Article II [117] and Article III [115].

Information that flow in IDMSs are organized and communicated in a form of tokens [170]. Tokens are technical artifacts providing assurance about identities [117], [127]. They can communicate authentication entitlement and attribute information in IDMSs [170]. They are personal data sources and gateways to end-users' personal data [103]. A token may consist of an attribute such as username, a claim such as a password, an assertion such as SAML token, a credential such as a X.509 certificate or a combination of these.

Category of Factors	Risk Factors
Token Mobility	*copyable, remotely usable, concurrently usable, immobile*
Token Value at Risk	*loss, misuse, disclosure, disruption, theft, replacement value*
Token Provisioning	*creation, editing, deletion*
Token Frequency & Duration of Use	*Uses per year, life-time, multiple times, one-time*
Token Use & Purpose	*original, unintended*
Token Assignment & Relationship	*forced, chosen, jointly-established, role, pseudonymity*
Token Obligation & Policy	*absent, present, functionality*
Token Claim Type	*single, multiple*
Token Secrecy	*public, inferable, secret*
Token Security	*origination, identification, validation, authentication, authorization*

Table 2.2: Privacy and Security Risks Analysis Contributing Factors for IDMS – Article II [117]

Article II [117] is a literature survey of factors that can contribute to privacy and security risks in IDMSs. The survey focuses on the characteristics of tokens that can contribute privacy and security risks in IDMSs. The literature survey resulted in a taxonomy of privacy and security risks contributing factors for IDMSs as shown in Table 2.2. Some of the individual risk contributing factors in Table 2.2 exist in the literature. However, these factors have not been categorized or synthesized for risk analysis in IDMSs.

The first column of Table 2.2 represents the categories of risk factors, while the second column represents the risk contributing factors. The risk contributing factors and their categories are discussed as follows:

Token Mobility: This category indicates the degree of mobility of a token. The degree of mobility refers to how easy it is to copy a token or its content, the physical constraints regarding the movement of the token, among others. For example, the content of a low cost RFID tag with no additional security could easily be read by anyone but a high cost RFID tag that comes with additional security may ensure that only authorized readers have access to its content. Various forms of mobility create risk in IDMSs. We assess the contributions of token mobility to privacy and security risk according to the following:

1. *Copyable*: the token can be copied with limited effort.
2. *Remotely usable*: the token can be used for remote identity management.
3. *Concurrently usable*: the token can be used concurrently in many parallel sessions, transactions, or applications.
4. *Immobile*: a token is not 'mobile', if it must be physically present.

Token Value at Risk: Finding assets and the value of the assets at risk is an important part of risk analysis [64]. Tokens are assets and their value at risk can contribute to privacy and security risk. Thus, we can quantify the risk of using tokens by assessing the significance of the token or the value of the token to the operation and security of the IDMS. We classify the value at risk [123] as follows:

1. *Loss*: value at risk when token is lost.
2. *Misuse*: value at risk when token is used in wrongful ways.
3. *Disclosure*: value at risk when token or token-related information gets known by someone else.
4. *Disruption*: value at risk when token doesn't function.
5. *Theft*: value at risk when a token is stolen.
6. *Replacement value*: cost (effort, resources, time) to replace a token.

Token Provisioning: The amount of personal information collected during tokens' creation could contribute to privacy risk [87], [21]. Data minimality principles prohibit excessive personal data collection. In addition, tokens that contain excessive personal information may be used for other unintended purposes. There should be a means by which tokens can be updated in order to ensure data integrity. Further, there should be a means by which a token can be deleted or destroyed when the purpose for its creation is no longer in the interest of the end-user. This would ensure the privacy of the end-user. Therefore, we distinguish the following phases of the token provisioning life cycle: *creation, editing and deletion*.

Token Frequency & Duration of Use: The underlying IDMS information flow protocol could determine if multiple use of the same token is susceptible to privacy and security

risks [57]. Different uses of even specifically constructed tokens remain linkable if the underlining IDMS's information flow protocol is not designed to prevent such privacy risk. IDMSs that allow a service provider (SP) and an identity provider (IdP) to share information or use the same token repeatedly may be susceptible to profiling. A token used in association with life time identification or multiple times causes high risk of secondary use and profiling [41]. We divide tokens into three categories – a token chosen or assigned for life (life-time token), a token that can be used multiple times but not for life (multiple times token) and a token that is valid for a session (one-time token).

Token Use & Purpose: Purpose specification is an important privacy principle. It requires that personal information be collected for a specific purpose and used in an open and transparent manner [49]. Personal information should not be used for purposes other than its original purpose unless informed consent is given. Since tokens may carry personal information and are the gateway to personal data [103], any form of function creep or misuse could contribute to a privacy or security risk. Function creep occurs when a token is used for unintended purposes. The abuse of the purpose of a token or how the purpose of a token is achieved may contribute to privacy and security risks.

Token Assignment & Relationship: The need for user-centric IDMSs emphasizes the importance token assignment. User-centric IDMS offer the end-users the ability to control their personal information and enforce their consent by determining the choice of their identity tokens [19].

The origin of and control over tokens contribute to privacy or security risk. A token can be chosen by a person, jointly established or forced upon by an authority. They can relate to a role or pseudonym rather than a person [57].

The token assignment and relationship risk contributing factors determine the privacy risk if a token is forced on an end-user, chosen by an end-user, or jointly-established with the end-user. It also assesses the end-user's privacy risk if a token is chosen under a pseudonym or a role [57].

Token Obligation & Policy: Some IDMSs give the end-users an opportunity to specify policies regarding how their tokens should be used [10]. In the absence of a policy, functionality such as the use of anonymous credentials and selective disclosure of attributes can mitigate privacy and security risk. Therefore, this category checks if a policy or functionality exists for privacy protection.

Token Claim Type: We can enforce the security of identifiers stored in a token or protect the misuse of identifiers by attaching a secret or a claim to the tokens. The type of claims are generally classified as *"something we know"* (a secret e.g. password), *"something we are"* (something we cannot share with others, e.g. finger print, iris etc.) and *"something we have"* (e.g. possessing a smart card, key card or a token etc.) [87]. Each claim type is a factor. Token claim type considers the effect of the number of factors used for securing a token in IDMSs. Some tokens may require no additional secrets in order to protect its content. We refer to such tokens as single claim tokens since the possession of the token itself is a factor. A token that requires a factor such as a personal identification number (PIN) in order to access its content is referred to as a two-factor or multiple claim token.

The number of authentication factors may determine the complexity and security of the token. For example, current credit cards may require no additional secret PIN when used for online transactions. We regard such a token as a single factor authentication token.

Token Secrecy: Tokens have claim type as discussed above. The secrecy or the constraints of authentication factors such as physical presence can contribute to security or privacy risk in IDMS. The nature of the secret and constraint of the additional authentication factors may determine the "claim type".

A token secret that can easily be guessed is classified as an *inferable secret*. A token's secrecy is *public* if its additional authentication factors or claims are known by a number of people, group and organization or exist in many databases, at the disposal of an unknown number of persons and organizations. We classify the secrecy of a token as *secret* if it is

private or not shared with others because it is linked to a valuable resource such as a bank account [1]. A possible example of private secrets is a private cryptographic key and a PIN.

Token Security: The security of a token can contribute to the security of the IDMSs. In order to ensure token security, there should be a means of ensuring the validity, identity and legitimacy of the token. Similar to Jan and Van [87], we describe the security of tokens as follows:

1. "Origin": the token is issued by the indicated legitimate authority.
2. "Identification": the token identifies the subject-the entity possessing the token.
3. "Validation": the token has not expired; its lifespan is within the validity period or has passed the validity test.
4. "Authentication": the token belongs to the entity presenting it to the IDMS.
5. "Authorization": the token grants relevant permissions to the entity possessing it.

2.5.1 Evaluation of the Risk Analysis Constructs

The taxonomy in Table 2.2 represents a set of constructs for risk analysis in IDMSs [90]. According to the DSR methodology [122], the performance and utility of constructs are determined through evaluation. Evaluation determines constructs' completeness, simplicity, elegance, ease of use and understandability. Evaluation of a construct can be done with the Delphi method [159]. The Delphi method solicits expert opinions on a subject matter in a structured group communication process.

Article III [115] is a three-round Delphi study that validates and evaluates the set of constructs or the taxonomy in Table 2.2. The evaluation focused on the completeness (comprehensiveness), simplicity, ease of use and understandability of the constructs in Table 2.2. The Delphi study involved nine experts during the first two rounds and six in the final round. The experts were drawn from the academia, industry and the research community.

New Protection Goals	Category of Factors	Factors
Confidentiality, Unlinkability	Frequency & duration of use	one-time, multiple times, life time
Integrity, Confidentiality, Intervenability	Provisioning	created, updated, deleted or archived attribute with: limited personal data, overloaded personal data, sensitive personal data
Unlinkability	Purpose of use	application specific, single sign-on, multiple services, context specific, silo
Intervenability, Unlinkability	Assignment & Relationship	forced, self, jointly-established, role, pseudonym
Confidentiality	Secrecy	inferable, public, obfuscated, revocable, recoverable,
Confidentiality	Claim Type	password, crypto key, biometric, challenge-response, single-claim, multiple-claims
Availability, Confidentiality	Mobility	copyable, remotely usable, concurrently usable, immobile
Availability	Value at Risk	loss, misuse, disclosure, disruption, theft, replacement value
Transparency	Obligation & Policy	policy absence, policies present

Table 2.3: Mapping of the Risk Factors and the New Protection Goals

Table 2.3 is the result of the Delphi study. It is a taxonomy of risk contributing factors and validated constructs for privacy and security risks analysis in IDMSs.

In Table 2.3, the third and second columns show the risk contributing factors and their categories respectively. The first column of the table maps the risk factors to the new protection goals (NPGs) [131]. The mapping aligns the risk contributing factors and their categories to the NPGs in order to show that the taxonomy is relatively comprehensive or complete. Furthermore, the mapping can aid communication among legal and technical experts involved in privacy and security risks analysis [177].

The NPGs provide technically convertible principles that cover both security and privacy protections [131],[177]. The NPGs extend or complement the classical security goals – integrity, availability and confidentiality, by adding central privacy concepts which are transparency, unlinkability, and ability to intervene (Intervenability).

Transparency requires that the purpose of data processing is comprehensible by all stakeholders [131], [177]. Unlinkability verifies if personal data collected for a particular purpose is being used for another purpose or personal data is unlinkable to any other set of privacy-relevant data outside a domain or context. The ability to intervene (intervenability) gives the data subjects or parties the ability to control or intervene in the processing of their personal data.

The NPGs complement each other; but sometimes conflicts can arise in their implementation [177]. Identifying and understanding such conflicts are a prerequisite for developing adequate and a balanced risk analysis model.

The Delphi study concluded that characteristics of tokens can contribute to privacy and security risks in IDMSs. In addition, the study identified "frequency & duration of use", "secrecy" and "claim type" as the key risk indicators (KRIs). KRIs are highly relevant metrics with a high probability of predicting or indicating important risk [64]. They are sensitive, having high impact, easy to measure, easy to understand and relevant to both technical and non-technical stakeholders.

2.6 Risk Modeling Phase

Question III: Given the characteristics of information that flow in identity management systems, in what way can a risk analysis model be developed for them?

The risk modeling phase is meant to answer the research Question III. Question III investigates a risk modeling approach for IDMSs. A precise risk model for IDMSs could prevent important information from hiding among irrelevant details and thereby enhance risk identification [167]. The risk modeling phase maps to Model Building and Evaluation in DSR [90].

Article IV [113] answers Question III. It analyzes various security modeling approaches. The analysis reveals that current modeling approaches are generally complex and security-oriented. Article IV [113] addresses some of these challenges with the extended misuse cases (EMCs) modeling approach. Use and misuse cases (UMCs) modeling is simple and user-friendly [138]. It makes human judgments more informed and systematic [5], [38]. UMCs risk modeling appeals to the industry because of the substantial connection of use cases to the existing system development process [106].

However, UMCs models lack quality goals, their operationalization and effect [38]. In addition, UMCs models cannot depict the dynamic behavior of system agents. Consequently, Article IV [113] introduces the extended misuse cases (EMCs) modeling technique. The EMCs extend misuse case diagram with "asset", "goals" and a privacy construct ("right") in order to model privacy and security risks in IDMSs. In addition, Article IV [113] refines and implements the informal EMCs model using colored petri nets (CPNs) [77] as shown in Figure 2.5. The CPNs' model is executable (can be simulated) and can

communicate or represent the dynamic behaviors of actors and adversaries to all system stakeholders.

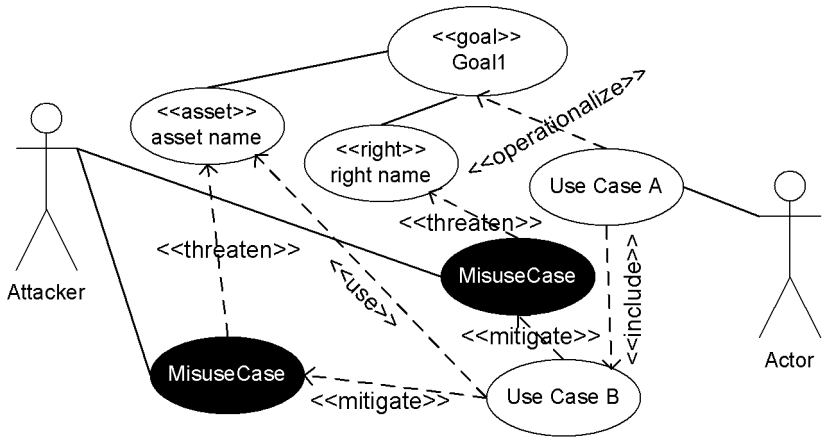

Figure 2.3: Extended Misuse Cases Modeling – Article IV[113]

Figure 2.3 and Figure 2.5 depict the EMC risk model for IDMSs [113]. An EMC model identifies the use and misuse cases and how they align with security goals or requirements. It models risk categories as the assets or the privacy rights being protected. The risk categories that enforce privacy are modeled as "rights" while that of security are "assets". Risk factors represent the misuse cases. They are factors that can contribute to privacy or security risk in IDMSs. Similarly, use cases represent controls or risk mitigation measures. The factors that protect system goals are modeled as use cases or countermeasures while those that threaten system goals are the misuse cases.

Figure 2.3 is a generic model of the EMCs model. In Figure 2.3, asset is what the system is protecting or something that may advance a system goal and is represented by its name and the stereotype $<< asset >>$. A use case may use an asset as a data object or a resource to accomplish a task [106]. This is represented by the dotted line and the stereotype $<< use >>$. The second extension is the "right". It is represented by its name and the stereotype $<< right >>$. "Right" incorporates privacy concepts into the model. While security modeling focuses on assets or protection of security goals, privacy on the other hand focuses on both security and right protection [158], [111]. Hence, the stereotype $<< right >>$ distinguishes the model from the traditional security models.

A misuse case may threaten a right or an asset [138]. In order to simplify the model, we do not make explicit distinction between vulnerabilities and threats. Both are assumed to be misuse cases. A use case may mitigate a misuse case and may help operationalize a system goal. Goal operationalization is represented by the stereotype $<< operationalize >>$. A system goal is operationalized if all its known misuse cases are mitigated with appropriate use cases or countermeasures. A goal represents the reasons why we need to protect an asset or a right from misuses. A goal is represented by a name and the stereotype $<< goal >>$. It may consist of sub-goals. If a goal is associated with an "asset" or "right" then it means that the goal is intended to protect the asset or the right from misuses.

Figure 2.4 is an example EMCs model for the "Value at Risk" and the "Mobility" categories in Table 2.3. The model has two goals – "Availability" and "Confidentiality" goals. The "Value at Risk" category is aligned with the "Availability" security goal while the "Mobility" category aligns with the "Availability" and "Confidentiality" goals. The "Copyable" misuse case threatens the "Mobility" of tokens (asset) which in turn affects confidentiality of tokens. The "remotely usable" and the "concurrently usable" mitigate the "Immobile" misuse case and operationalize the "Availability" goal.

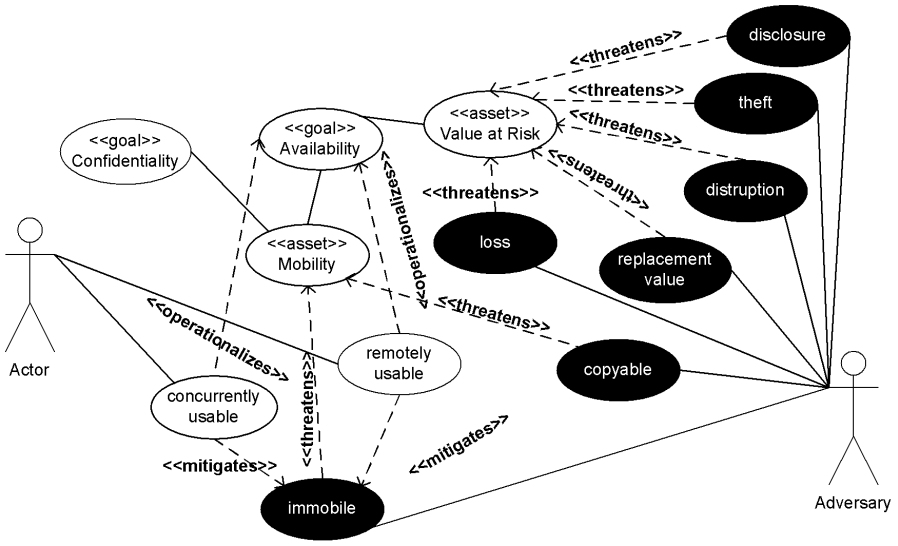

Figure 2.4: Executable Extended Misuse Cases Model for Mobility and Value at Risk Category

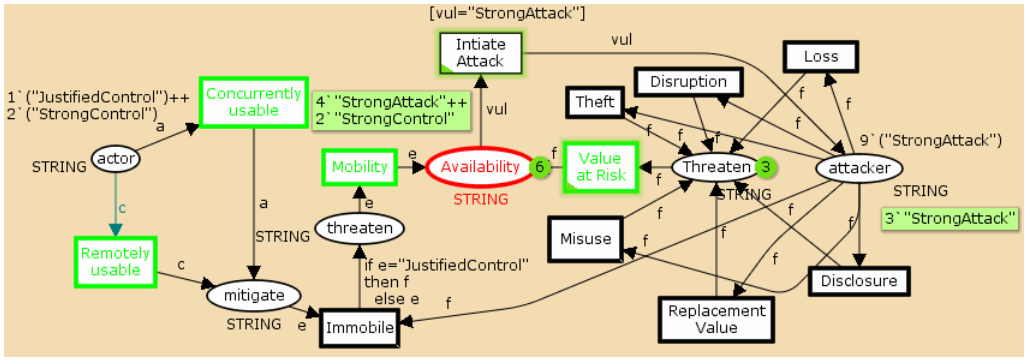

Figure 2.5: Colored Petri Nets Implementation of Executable Extended Misuse Cases Model

The EMC modeling can be refined and implemented with CPNs in order to precisely communicate the dynamic system behavior of system agents to stakeholders. Figure 2.5 is an example of the refined and precise version of an EMC model in Figure 2.4 with CPNs. The CPNs implementation converts the static EMC model to an executable risk model capable of precise and dynamic risk communication [77]. The CPNs model in Figure 2.5 implements the "Mobility" and "Value at Risk" risk categories explained above and focuses on the availability security goal.

In Figure 2.5, actors and adversaries are represented by places. The countermeasures for the actor are the use cases or the CPNs transitions (colored green). The misuse cses for the attacker or adversary are the black colored transitions. The "include", "mitigate", "threaten" and goals such as availability are modeled as places. Some of the misuse cases have countermeasures others are not. For example, token "loss", "theft", "disclosure", "misuse" and "replacement" misuses have no corresponding use cases or countermeasures. An adversary can always violate a security goal through these unmitigated misuse cases. The transition "Initiate Attack" is to enable or illustrate the continual attack by the adversary through the unmitigated misuse cases.

The initial markings "JustifiedControl" and "StrongControl" correspond to the perceived strength or capabilities of the actor's action. On the other hand, the initial mark-

ing "StrongAttack" corresponds to the perceived strength or capabilities of the attacker's actions. The token movement corresponds to the progress of an attack or the effect of the actor's countermeasures [95]. If an attacker's token reaches a goal (place) then the actor's countermeasure was ineffective.

The execution of the risk model in Figure 2.5 can enable stakeholders to observe the dynamic behavior of the actors and adversaries. The risk model can be extended with additional concepts and animation tier to communicate risk more effectively [77].

The EMCs and its CPNs implementation have not been tested on system stakeholders to determine their effectiveness. In addition, this thesis did not address the complexity of misuse cases modeling in relation to goal-oriented risk modeling.

2.7 Analysis Phase

Question IV: To what extent can we assess privacy and security risks of identity management systems with the characteristics of information that flow in them?

This phase maps to Method Building and Evaluation in DSR [90]. It investigates an automated, relatively simple and less subjective model-based risk analysis method for IDMSs. The method is built on the risk factors identified in Article III [115] and the extended misuse cases models in Article IV [113]. The misuse cases serve as inputs to a CPNs' model of a targeted IDMS to analyze the system's risk. Automation can reduce cost and enhance privacy and security in IDMSs.

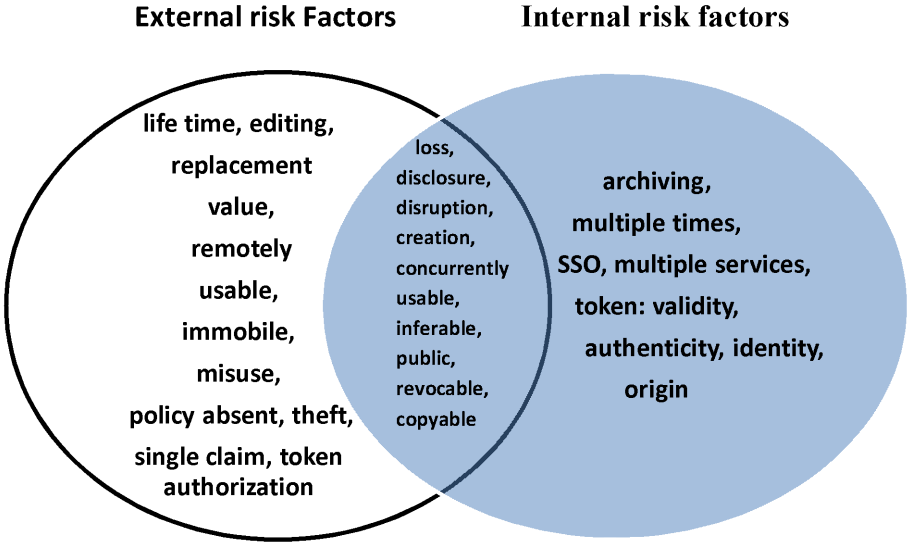

Figure 2.6: External and Internal Risk Factors – Article IV [113]

The main activities in this phase is system modeling, validation and risk verification with CPNs. Firstly, the misuse cases identified in Article IV [113] are categorized into external and internal risk contributing factors or misuse cases as shown in Figure 2.6. The external misuse cases consist of risk contributing factors that may be outside the control of IDMSs or factors that could not be easily verified. For example, risk contributing factors that affect privacy policy or access control may lie outside the control of IDMSs. Internal factors are those under the control of IDMSs and they are relied upon to automate the privacy and security risks analysis.

The main objective of risk analysis is to identify and assess all risks in order to suggest a set of controls that will reduce these risks to an acceptable level [44]. Traditionally, risk

assessment requires estimation of likelihood of a threat manifesting and its impact. In cases where there is little data to validate the likelihood of a threat manifesting, risk assessors rely on their experience and subjective intuitions to estimate the likelihood. This way of analyzing risk is useful for management decision but may be less useful for technical risk reduction [25], [44].

The requirement analysis approach may provide adequate risk analysis results than the traditional risk analysis approach [44]. It determines security requirements of a system or an organization and deduces the most suitable set of security controls from these requirements. The executable model-based risk analysis method (EM-BRAM) developed in this thesis relies on the EMC risk analysis model in Article IV [113] to perform a requirement analysis (from technical perspectives) on a given IDMS in order to determine the security controls needed to secure the system. The analysis focuses on internal misuse cases that can be automated. The external misuse cases may be relied upon for policy formulation for IDMSs.

After the analysis, the EMC model in Article IV [113] is relied upon to determine the appropriate controls to mitigate the risk posed by the misuse cases in the given IDMS.

The EM-BRAM models IDMSs and verifies the internal risk using CPNs [77]. CPNs is less mathematical and has a high degree of automation making it relatively easy to use. The automation can potentially reduce cost involved in the risk analysis process. The given IDMS is modeled and validated using CPNTools [77] before the risk analysis. Validation loosely refers to "the process of determining that a design is correct" [36]. According to Holzmann [53], the three important general properties for a correct model are, absence of cycle, deadlock and improper termination. Firstly, the CPNs simulation tools are used to validate the correctness of the system model. Secondly, the three correctness properties stated by Holzmann [53] are verified with the state space report generated automatically by the CPNTools [77].

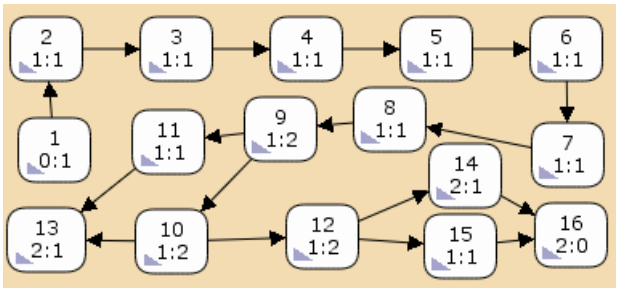

Figure 2.7: State Space for SAML Single Sign-on Service for Google Apps and an Intruder SP [12]

In addition, the CPNTools [77] generate state space (graph) for the risk analysis. For example, Figure 2.7 is the state space graph for SAML Single Sign-on Service for Google Apps and an Intruder SP [12] generated automatically by the CPNTools. CPNs queries and ML predicate functions are then applied to analyze the state space (graph) for privacy or security risks. The state space represents all the possible states of a given IDMS. The queries search through each system state to verify if a given risk condition exist in a state. For example, the following CPNs query can be used to verify whether a token is used for multiple services or single sign-on (SSO). Although SSO reduces human error, it leads to sharing of valuable information across services or domains.

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

The example query above verifies if the user-name "bob" can be found in the "GoogleSP"

sub-model (see [118]). The "GoogleSP" sub-model is outside the sub-models for the end-user "bob".

Factors	Risk Value Low Level GApps	Risk Value High Level OpenID	Meaning
multiple times	Yes	Yes	Tokens can be linked or profiled by a SP
single sign-on/ multiple services	Yes	Yes	Tokens may be linked or profiled
creation	Yes No	- No	Token has insufficient number of attributes Token has no sensitive attributes
archiving	Yes/No	 Yes	Token can be archived but it is not sensitive Token can be archived
public	No	No	Token secret is kept private between end-users and IdP
copyable	No	No	Tokens cannot be copied
concurrently usable	Yes	Yes	Token can be used concurrently
loss	No	No	Tokens cannot be lost in the IDMS
disclosure	No	No	Tokens cannot be disclosed in the IDMS
disruption	No	No	Tokens are not disrupted by conflict or deadlock in the IDMS
origination	No	 -	Tokens' originators are included in the information flow Unable to model
authentication	Yes	 No	Tokens' authenticity test failed Tokens' authenticity test did not fail
identification	No	No	Tokens include the identity of the end-user
validation	Yes -	 -	Token validation failed Unable to model

Table 2.4: Risk Analysis Report

The various security and privacy risks analyzed are tabulated as in Table 2.4. Table 2.4 depicts the results for low-level and high-level IDMSs' specifications. The Low Level GApps represents the low-level specification of the SAML SSO service for Google Apps [45], [12] while the High Level OpenID represents the high level specification of OpenID [129]. The "Yes" or "No" in the table means risk was identified while "-" means the method was unable to verify the risk because of the specification.

Article V [118] and Article VI [116] introduce the EM-BRAM for IDMSs described above. The EM-BRAM is applied in analyzing the low-level specification of single sign-on service (SSOS) for Google Apps [45] and high-level specification for OpenID [129]. All the risks were verified in the low-level specification. However, the method is unable to verify some of the risks in the high-level specification. This means the EM-BRAM works better with low-level specification.

The EM-BRAM is limited by the expertise required to convert a system specification to a CPNs model. A considerable amount of work and technical expertise are required to do so. However, the reliance on a predefined risk model could significantly reduce the cost involved in the application of the method. The quality of the Delphi survey results may affect the quality of the results produced by EM-BRAM. The Delphi evaluation of the risk contributing factors used between 6 and 9 respondents which is less than the recommended number of respondents [159].

2.8 Evaluation of Artifacts

This section discusses how the main artifacts developed in this thesis were evaluated in accordance with the DSR methodology. In Table 2.1, we have four research activities and four research outputs. The engineering research activities are "Build" and "Evaluate" while the "Theorize" and "Justify" are for natural science [90], [51]. This thesis created three main artifacts for IDMSs – risk analysis constructs, risk model and risk analysis method.

Table 2.5 states the three main artifacts created in this thesis and the methods of evaluation, the metrics for the evaluation and the goals of each artifact.

Construct: Article II [117] introduces the construct for privacy and security risks analysis of IDMSs. The evaluation of the construct was done with a Delphi study [159] in Article III [115].

Artifacts	Goal	Build Outcome	Evaluation Metrics	Evaluation Method
Construct	Identify factors that contribute to privacy and security risks in IDMSs and use them to develop a model for risk analysis of IDMSs	Taxonomy of privacy and security risks contributing factors (Article II) [117]	Completeness, Understandability, Ease of use, Sensitive, Ease of measurement and relevance	Expert validation and Evaluation (Article III)[115]
Model	To develop an explicit risk model for IDMSs	A Model for privacy and security risk analysis (Article IV) [113]	Completeness	Descriptive method and Simulation
Method	To develop an innovative method for risk analysis of IDMS	Executable Model-based risk analysis method (Article V [118] & VI [116])	Operationality	Descriptive method and Case study

Table 2.5: Evaluation of Artifacts

The study evaluated simplicity, understandability, ease of use and completeness of the construct. The expert study evaluated how easy it is for system stakeholders to understand the risk factors identified in the taxonomy in Table 2.3. In addition, the study evaluated the ease of use of the construct by asking the experts to assess how easy or difficult it is to quantify the risk factor. The completeness of the construct [126] was evaluated by aligning the outcome of the study to the new protection goals [131]. The alignment shows that the construct is relatively complete.

Model: Article IV [113] introduces an extended misuse cases (EMCs) model for risks modeling. The evaluation of the model was done by comparing and positioning the model to the existing literature in the field of conceptual risk modeling (descriptive method) [90], [51]. The existing use and misuse cases models lack quality goals, their operationalization and effect [38] and therefore incomplete for privacy and security risks modeling. Furthermore, they are not executable and therefore cannot automatically communicate dynamic behavior of threat agents. Unlike the use and misuse cases models, the introduction of quality goals, asset and right stereotypes makes the model relatively complete for privacy and security risks modeling. An erroneous model confuses matters, possibly providing a false sense of security [125].

The implementation of the EMCs model by means of CPNs enables automatic communication of dynamic behavior of actors and threat agents. In addition, the implementation enables further analysis of the model. The state space analysis of the CPNs implementation shows that the model behaves as expected.

Method: Article V [118] and Article VI [116] develop a method for privacy and security risks analysis of IDMSs. The "operationality" of the method is evaluated with case studies. In addition, the method is compared to the existing risk analysis methods for IDMSs. Unlike the current risk analysis methods for IDMSs, the executable model-based risk analysis method in Article V [118] and Article VI [116] is partially automated, relatively simple, less subjective and can facilitate risk communication among system stakeholders. An automated and less subjective model-based privacy and security risks analysis method can reduce cost, time spent in the analysis and above all enhance security and privacy of IDMSs.

2.9 Conclusion

In order to define the research problem and frame the research activities that will address the need for innovative risk analysis artifacts for identity management systems (IDMSs), the first question of the thesis is meant to understand the problems or challenges in the information security risk assessment discipline. This goal is achieved by reviewing the

literatures in the research area (Article I [114]). The literature review led to the creation of a taxonomy of risk assessment approaches that can be used to guide a security risk assessment research. The taxonomy identifies the challenges of the various risk assessment approaches and proposes a criterion that can guide researchers to develop innovative artifacts for security and privacy risks analysis.

Information flow can provide meaningful insights into the overall system security and privacy of IDMSs. However, the contributions of the characteristics of information that flow in IDMSs to privacy and security risks have not been analyzed in detail. The second question of the thesis is meant to investigate whether risk assessment constructs for IDMSs can be created from the characteristics of information that flow in IDMSs. The constructs are concepts, terms or vocabularies for describing and specifying solutions for a problem within a discipline. Risk analysis constructs can be employed to create risk models for privacy and security risks in IDMSs. The creation of the taxonomy of privacy and security risks contributing factors for IDMSs in Article II [117] and the identification of key risk indicators from the characteristics of tokens in Article III [115] show that characteristics of tokens can contribute to privacy and security risks in IDMSs. Consequently, they can be used for risk modeling and analysis in IDMSs.

Risk modeling is one of the important requirements in information security risk assessment because it gives a clear idea of what we want to mitigate and the capabilities of adversaries. It supports the risk analysis and guides the subsequent iteration of the entire risk analysis process. The third question of the thesis is meant to create a privacy and security risks model for IDMSs. The extended misuse case (EMC) modeling technique in Article IV [113] answers this question. The EMC extends the user-friendly use and misuse cases modeling technique. Use and misuse cases modeling techniques lack quality goals, their operationalization and effect. Hence, the EMC modeling technique extends them with quality goals, and security and privacy constructs. It is then applied to create a privacy and security risks model for IDMSs. This leads to the conclusion that use and misuse cases modeling technique can be extended for the creation of a privacy and security risks model for IDMSs. The model can be refined and implemented with colored petri nets (CPNs) in order to enhance precision and communicate dynamic system behavior of actors and adversaries. The dynamic risk model rather than static risk model is developed.

Model-based risk analysis method for IDMSs is not the main focus of current research. Model-based risk analysis methods enhance risk analysis process by facilitating participation, risk communication and documentation. Current model-based security risk analysis methods do not target IDMSs. They are either too complex or highly subjective and lack automatic validation tools. The last question of this thesis is meant to investigate whether the characteristics of information that flow in IDMSs can be employed to develop a model-based risk analysis method for IDMSs. The development of the executable model-based risk analysis method (EM-BRAM) for IDMSs in Article V [118] and Article VI [116] shows that it is possible to develop risk analysis method for IDMSs with characteristics of information that flow in them. The EM-BRAM relies on system behaviors or characteristics to analyze privacy and security risks in IDMSs rather than data on past events or intuitions of a risk assessor. Consequently, it can reduce subjectivity and uncertainty in the risk analysis of IDMSs. EM-BRAM relies on CPNTools to determine the correctnesses or validity of a system model before risk analysis.

2.10 Suggestions for Future Research

This section makes some suggestions for future research.

Converting a system specification to a CPNs model requires a considerable amount of work and technical expertise. An automated or semi-automated process can reduce the amount of work involved in converting an IDMS specification to CPNs model. This can improve the ease of use of the EM-BRAM. Notwithstanding this, the cost of using EM-

BRAM may be minimized by the help of the predetermined privacy and security risks model (EMC).

The EM-BRAM developed in this thesis depends on the risk contributing factors. The quality of the Delphi survey results may be affected by the limited number of experts. The Delphi evaluation of the risk contributing factors used between 6 and 9 respondents which is less than the recommended number of respondents. A comprehensive Delphi study can be conducted to improve the quality of the risk contributing factors. The external risk contributing factors identified in the study can be developed as technical or non-technical policies in order to enhance privacy of IDMSs. An ontology of risk contributing factors could provide a holistic view of how the factors relate to each other. Furthermore, the internal risk contributing factors can be formalized with CPNs formalisms for technical stakeholders in order to mitigate ambiguities in the meaning of the factors.

CPNs modeling cannot communicate risk effectively to non-technical stakeholders without additional communication concepts and animation tier. CPNs models can be extended with visual communication tier capable of explaining the modeling concepts to non-experts system stakeholders [68]. Effective risk communication concepts and animation tier are not considered and thus left for future research.

The extended misuse modeling approach and its CPNs implementation developed in this thesis have not been tested on system stakeholders to determine their effectiveness. In addition, the complexity of misuse cases modeling in relation to goal-oriented risk modeling was not evaluated. Future work may consider how SEQUAL (quality framework for evaluating modeling languages) [58] can be extended to evaluate the privacy and security risks model.

The case studies in this thesis considered attacks involving system stakeholders. Further studies can focus on attacks involving non system stakeholders. In addition, standard attack models and queries can be developed to facilitate the risk analysis process in IDMSs.

The risk analysis method developed in this thesis supports one-time tokens. However, the usability implications of one-time tokens are not clearly addressed. The number of partial identities and the corresponding credentials increase when end-users use different tokens for different services. The excessive mental and manual load involved in managing the numerous partial identities and the corresponding credentials raises the psychological burden of end-users [69].

The ML queries used to search through the state space of the CPNs model can be improved with more fine grained queries that target privacy requirements. This can be achieved by combining computation tree logic formulas with the ML queries.

The possibilities of using graph theory to analyze the state space of CPNs models can be considered for privacy and security risk analysis in IDMSs.

Chapter 3

State of the Art

This chapter discusses how security risk is mitigated in the current identity management systems (IDMSs). Furthermore, it discusses the existing security models and analysis methods for IDMSs.

3.1 Architecture Based Security for Identity Management Systems

IDMSs create identities and manage their lifecycle [69]. Security in IDMSs can be enhanced from the system architecture, behavior and policy dimensions. How the system architectures enhance security in IDMSs have been extensively analyzed [70], [27], [169]. This has led to various classifications including models, types, classes and paradigms.

Models of IDMSs include isolated, centralized and federated. In the isolated models, the SP and the IdP are one entity who creates and manages the identity lifecycle of the end-users. The identity can be used for authentication and authorization when it is in the active state.

The isolated model is simple but may be inconvenient for system end-users because of the explosive growth of online services. The model allows end-users to manage more and more identities as they subscribe to different services. This raises the mental and manual load of the end-users [69] which in turn leads to rampant loss of credentials, high recovery cost and security risks [27].

The centralized model separates the task of the SPs from that of IdPs. All SPs use the global unique IdP and the identities of every SP are stored in IdP. When an SP needs to authenticate an end-user, it will send the user information to the IdP to finish the process [27], [17]. The requirement that only one IdP stores all the identities creates single point of failure and high privacy risk. Thus, it can give the system owner a significant power to misuse the personal information it collects. Examples of centralized IDMSs are Mircosoft.Net passport [97] and Kerberos [92].

The federated model is simply an aggregation of two or more centralized model. The aggregation is regulated by policies and protocols that allow end-users in one domain to access resources from another domain [27]. The domains establish agreements among each other so that identities from an identity domain are recognized across all domains [69]. Protocols, standards and systems for federated model include Security Assertion Markup Language (SAML) [26], WS-Federation [97], and the Liberty Alliance framework [61].

The federated model requires sharing of valuable information across domains using loosely coupled network protocols [89]. It requires all parties to secure their communication channels against replay attacks, man-in-the-middle attacks, session hijacking as well as other attacks that lead to malicious use of identity information or resources.

The future of identity in the information society (FIDIS) categorizes IDMSs into types and classes [169]. The Type 1 IDMSs support account management. They implement authentication, authorization and accounting management services. The Type 2 systems support profiling of end-users' data. They keep detailed log files or data warehouses that support personalized services or the analysis of customer behavior. The Type 3 IDMSs are user-centric. User-centric IDMSs support user-controlled context-dependent role and pseudonym management [169].

The Type 1 and 2 IDMSs allow end-users' personal data to be managed by central administrators or operators [169]. The end-users have no control over their personal data. Thus, they cannot control context-dependent role and pseudonym management. The Type 1 and 2 IDMSs may be subjected to similar privacy and security risks as the centralized IDMS models. On the other hand, the Type 3 IDMSs can allow end-users to act under pseudonyms and control the use and release of their partial identities. However, the user-centric architectures may suffer from a number of security risks including profiling, phishing, cross-site scripting and identification risks [43], [34].

In addition to types of IDMSs, FIDIS provides product oriented classification of IDMSs [169]. The classification is based on the extent to which an application or a system supports IDMSs' functionalities. The functionalities are the creation and management of the life-cycle of end-users' identities. The class 1 applications or systems are those that support IDMSs' functionalities in full while the class 2 systems are those that support some IDMS functionalities. The Class 3 systems are those that offer at least some identity management functionality, such as add-ons.

Cao and Yang [27] proposes three paradigms of IDMSs – network centric, service centric and user-centric paradigms. The network paradigm occurs in the early development stage of an IDMS technology. It separates identity management, creation and deletion from access control. The IDMS is established and operated by a single entity for a fixed user and resource community. An example is Microsoft Windows domain. The network paradigm is neither subjected to the restrictions of the applying scenarios nor considers the identity of application scenarios. An entity may have several identities; every identity may have several identifiers or attributes. The network paradigm is prone to privacy risks since excessive data is collected for different applications [87], [21].

The service centric paradigm allows service providers to manage services for end-users [27]. They are composed of services from different providers across multiple domains. The end-user can dynamically and explicitly delegate his access rights of a service to a service provider. Examples of the service centric IDMSs are Google Calendar and Microsoft Exchange. In this paradigm, the end-users have no control over their identities but rely on the service providers for privacy protection.

The user centric paradigm allows end-users to control their identities [27]. They shift the control of digital identities from service providers to end-users. The end-user is in the middle of transactions between identity providers and relying parties. This enables end-users to control the use and release of their partial identities. The user-centric paradigms may suffer from a number of security risks including profiling, phishing, cross-site scripting and identification risks [43], [34].

The layered architecture of IDMSs is introduced by Suriadi et al. [153]. It provides the functional view of identity meta-systems. The architecture consists of the following layers:

Presentation Layer: Provides an interface for end-user to interact with an IDMS.

Provisioning Layer: This layer provides functionalities for end-users to create, store, delete and update tokens. It also provides functionality for an IdP to revoke an identity.

Token Layer: This layer is used to communicate end-user's claims, entitlement and attributes or tokens to system stakeholders or agents. The tokens include security assertion markup language (SAML) assertions and X.509 certificate.

Protocol Layer: IDMSs have various protocols that allow end-users to identify, authenticate and access services or resources. These include authentication protocol for OpenID [129], CardSpace [97], SAML [107] and Idemix [57]. Furthermore, this layer consists of protocols that specify how a token should be secured, requested, issued and validated [170]. Examples of such protocols are WS-Trust and WS-Security protocols [170]. The SSO, CardSpace, OpenID and WS-Trust are protocols for specific IDMSs.

Service Layer: IDMSs are usually developed to support a particular service. This layer defines the services available for a particular type or class of IDMSs. The services include accounting services, profiling and logging services [169]. For example, type 1 IDMSs sup-

port accounting services while type 2 supports profiling [170].

Management Layer: This layer defines how tokens are managed in IDMSs. The layer also manages the services provided by the IDMSs or the service layer. The management functionalities include session management, access management and registration management [169].

Policy and Functionality Layer: This layer is responsible for handling the policies that are associated with the use of tokens. Policy refers to the regular privacy respecting policies that are used to solicit end-users' consent. Functionality refers to mechanisms such as sticky policy [10], selective disclosure and anonymous tokens [23] used to enhance privacy in IDMSs.

Architecture security is outside the scope of this thesis. This thesis focuses on some aspect of behavior security in IDMSs.

3.2 Behavior and Policy Based Security for Identity Management Systems

Recent research tends to focus on system behavior for security enhancement. U-Prove [20] and identity mix (Idemix) [57] credential systems are typical examples of such research approaches. Credential systems rely on system behaviors to enhance security in IDMSs. For example, the Idemix scheme [57], employs anonymous credentials to enhance privacy. It supports zero knowledge proof system. The zero knowledge proof ensures that the SP and IdP know the end-user by different pseudonyms such that the two pseudonyms cannot be linked. It enables an end-user to convince a SP that he possesses a signature generated by an IdP on a statement containing attributes and pseudonym (N), and he knows the master secret key that protects N.

It has three main parties. The end-user who obtains a pseudonym in a form of an anonymous token from the identity issuer. The verifier who verifies the credentials or tokens. The end-users can authenticate with a verifier without revealing their pseudonyms. This means end-users' activities cannot be profiled or linked by service providers or the verifier.

Idemix uses both password or passphrase, master secret (private key) and a pseudonym (public key) to secure end-user's identity. Cryptographic tokens are sophisticated and not easy to infer. Obfuscated or sophisticated token can enhance security and privacy because they may be difficult to infer. Recovering an Idemix master secret will be difficult because they are tightly coupled with credentials [57]. In addition, Idemix can revoke a credential upon misuse without notifying the owner of the credential [18]. So long as a credential is misused, found in a revocation list or certain functionality decides its revocation, the anonymity of that credential can be revoked [18], [57]. This makes it difficult for Idemix to detect a revocation attack.

The behavior based security mechanisms may rely on effective control and manipulation of information flow or tokens in order to enhance security and privacy in IDMSs. However, the extent to which characteristics of tokens employed in such systems contribute to privacy and security risks has not been the main focus of research. This thesis focuses on the contributions of the characteristics of token to privacy and security risks in IDMSs.

Finally, the policy based research approaches rely on privacy policies to enhance privacy in IDMSs. However, the privacy enhancement depends on the extent to which the policies are enforced. Examples of policy based research approaches are PrimeLife policy language for privacy enforcement [10] and the Platform for Privacy Preferences (P3P) [84]. The P3P is a protocol for expressing privacy policy in both a machine and human readable way using a standard XML schema [84]. The standard schema allows the service provider to use a set of predefined terms to describe their privacy policies. The privacy policies may specify the kind of data the web site collects, dispute resolution procedure, how long data will be retained and how the personal data will be used. Furthermore, the

World Wide Web Consortium (W3C) designed a Platform for Privacy Preference Exchange Language (APPEL) to enable individuals to express their privacy preferences, to query the data represented by P3P, and to make decisions accordingly [84].

Furthermore, Ardagna et al. introduced the PrimeLife policy language for privacy enforcement [10]. The language uses modalities such as temporal constraints, pre-obligations, conditional obligations, and repeating obligations to model different types of obligations. It uses authorization modality to specify data transfer competence. Rules have two modalities, permit or deny. The language uses the concept of trusted credentials and specifies the agreement between a data controller and a data subject as a promise. Policy-based security is outside the scope of this thesis.

3.3 Risk Analysis Approaches for Identity Management Systems

Current IDMSs such as OpenID [129], and Microsoft CardSpace [43] are meant to enhance privacy online. They are meant to reduce the mental load of end-users as they subscribe to services online [69]. In addition, current IDMSs are meant to reduce the rampant loss of credentials, high recovery cost and security risks. Many companies including Facebook, Google and Yahoo already support current IDMSs [144]. Thorough risk analysis of current IDMSs can lead to their revision for better security and privacy of system stakeholders.

Kormann and Rubin analyzed the security risk of the Microsoft.Net Passport IDMS [75]. The manual analysis focused on the security flaw in the interaction of Microsoft.Net Passport and Netscape browsers that leaves an end-user logged in while informing him that he has successfully logged out. They described how an adversary may steal the security token concealed in a ticket granting ticket cookie by mounting DNS attacks. The analysis assumes that an average end-user may not properly understand the certificate model and the user interface of the secure socket layer protocol. They identified risks caused by improper cookie management, key management risk caused by insecure key generation and single point of failure caused by centralized identity management.

Gajek et al. [43] analyzed 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 described 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 CarsdSpace 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. The SOP is a security policy universally supported in browsers. SOP policy enforces that the access of scripts is limited to objects originating from the same source.

Gross's security analysis of the security assertion markup language (SAML) single sign-on protocol discovered flaws that allow attacks on the protocol. The attacks described include man-in-the-middle attack, replay attack, attack by information leakage, and impersonation [46]. He identified that the end-user's browser can connect to inter-site and transfer URL and other data without authentication. The end-user's browser cannot necessarily verify the IdP's certificate in order to identify it. This lack of verification is a major cause of man-in-the-middle attacks on the communication between browser and IdP's site. The protocol has six main steps. Gross analyzed the attacks in each of the steps. For example, in step one, the initial message or URL from the end-user's browser to the IdP contains the resource the end-user wants to access. However, the protocol neither specifies further elements of the URL nor prohibits the inclusion of more elements. This can allow an adversary to accumulate and mount replay attacks.

Similarly, an improved version of the SAML protocol was again analyzed by Gross and Pfitzmann [47]. Their new findings suggest further improvements.

Bart and Martijn introduce a classification of security issues of the OpenID IDMS [34]. The classification presents the security issues of OpenID in a uniform way in order to raise awareness about security issues surrounding the IDMS. Four classes of issues identified are single sign on, open to anyone, OpenID specific issues and use of web standards. The security attacks of these classes are presented and risk levels are qualitatively estimated. Finally, countermeasures are proposed to mitigate the attacks.

Security analysis of OpenID by Sovis et al. [144] focuses on how identity information of the IDMS can be manipulated. They described possible manipulation of assertions caused by improper verification and lack of integrity protection for authentication requests. Their analysis focused on transport security, parameter injection and parameter forgery.

The above risk analysis methods rely on manual inspections and incomplete because the stakeholders' interests are ignored [22], [166]. Manual inspection is time consuming and therefore expansive.

Cabarcos et al. [22] proposed a metric-based risk analysis framework for IDMSs. The framework requires data collection and extraction of significant numerical values, the metrics, that could be used together with statistics and probability theory, to create a risk model for IDMSs. The two phases of the framework are pre-federation and post-federation. The pre-federation encompasses the establishment of a relationship and the exchange of all the required information to engage in cooperation while the post-federation contemplates the transactions between two federated entities. Although interests of stakeholders were the main motivations for the proposed framework, it is not clear how they were addressed in the framework.

Moreover, the metric-based framework proposed by Cabarcos et al. [22] is based on the assumption that subjectivity or uncertainty is removed in risk assessment by combining probability of past event with stochastic processes to analysis. However, this hybrid approach could only minimize the uncertainty in the subjective risk estimation [113].

Suriadi et al. [154] evaluated two security and privacy goals of IDMSs. Using colored petri nets, they showed that end-users could maintain anonymity throughout multiple single sign-on (SSO) 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. Relatively comprehensive models that consider both privacy and security goals including transparency and intervention [177] could provide a better analysis.

The reviews above show that current risk analysis methods for IDMSs are ad-hoc, incomplete or less comprehensive and based on manual inspection. This thesis investigates more comprehensive and automated method for privacy and security risks analysis in IDMSs.

3.4 Risk Analysis Approaches

Privacy Impact Assessment (PIA) is a framework for assessing the impact of personal data processing on privacy before an information system is implemented [60]. The PIA is a compliance check of personal data processing with respect to privacy laws, policies, or regulations. It specifies the requirement for privacy risk assessment without any explicit assessment technique or method. Similarly, Solove introduces taxonomy of privacy [143]. He introduces high-level privacy requirements which include data processing, data collection, data transfer and invasion.

Abu-Nimeh and Mead introduced privacy and security risk assessment technique that combines the existing security risk assessment techniques in SQAURE with the PIA technique and the Health Insurance Portability and Accountability Act (HIPAA) to address the full spectrum of security and privacy risks [2]. The technique is based on the traditional risk

assessment approaches. The traditional risk assessment approaches have a very high subjective influence because of lack of exhaustive publicly available data on occurred events. They lack automated search capabilities, rely on intuition of risk assessors and therefore costly and error prone [166], [146].

CORAS [58] is an example of contemporary model-based risk assessment method. It provides a better approach to IT security risk assessment than the classical risk assessment methods that depend on text and tables. However, CORAS is prone to errors because risk estimation depends on subjective intuitions of risk assessors. Moreover, CORAS does not model the dynamic system behavior. It provides specialized support for risk modeling with little support for the modeling of the actual system under investigation [91]. In addition, it addresses privacy issues in very limited manner. Moreover, CORAS is prone to errors because it provides no formal means of validating the correctness of a target system model before risk analysis.

Another model-based IT security validation technique is UMLSec [71]. UMLSec provides a formal model of a targeted system before risk verification. The UMLSec's formal specifications are mathematically based. Mathematically based formal methods are expensive, have difficult learning curves, and cannot communicate risk to non-technical stakeholders.

Model checking refers to a set of mechanized techniques for automatically discovering scenarios in which the actual system behavior and that of the stakeholders' model diverge from one another [74]. These scenarios identify potential failures and pinpoint areas where design changes or revisions should be considered. The communicating sequential process (CSP) is among the most popular model checking approaches. The CSP describes a system and verifies how it performs by the Failure Divergence Refinement (FDR) model checker. The properties to be checked are expressed as CSP processes specification. The FDR checks are then used to verify if the set of behaviors generated by the system is a subset of those generated by the specification.

Unlike CSP formal method, colored petri nets (CPNs) [77] hide large portions of complex mathematics and have a high degree of automation making it relatively easy to use. They have tools supporting model building, interactive simulation and formal analysis. The hierarchical modules of CPNs enable modeling of large and complex systems manageable and compositional. CPNs can be extended with time for performance analysis. In addition, CPNs can conveniently express the best upper and lower bounds of places [77].

Furthermore, CPNs can be extended with animation tier to enhance communication among technical and non-technical stakeholders [77]. CPNs have a concise mathematical definition which contains very few but powerful primitives making it easy to learn, use and to develop strong analysis methods by which properties of system models can be proved [164]. It is flexible in terms of "token" definition and manipulation. Various elements such as use cases, messages and task can be represented by different types of tokens. CPNs "Tokens" are a variable with data type and value.

3.5 Risk Analysis Models

The ENISA's position paper on electronic identity tokens (eID) represents expert opinion on risk contributing factors of eIDs [103]. The paper discusses the significance of tokens and shows how their security characteristics can be used to compare eIDs offered by European Union (EU) member states. Tokens or eIDs are gateways to personal information and unwanted disclosure of personal information as a result of the issuance or use of the eIDs constitutes a privacy violation. The paper identified 14 threats and how to address them. These include falsification of content, eavesdropping, physical attacks, man-in-the-middle attacks and side-channel attacks.

Finding assets and the value of the asset at risk is an important part of risk assessment [64]. Similarly, tokens are assets of IDMS and their value at risk can contribute to privacy

and security risks. Thus, we can quantify the risk of using tokens by assessing the significance of the token or the value of the token to the operation and security of the IDMS and the privacy of the end-users. Peterson introduced factors for asset value computation and stresses the importance of asset in risk calculation [123]. We can use such factors to quantify the contributions of tokens to privacy and security risks. Peterson derived the asset value of tokens from their loss, misuse, disclosure, disruption, replacement value, or theft. The six risk factors for IDMS have been found in analogy to classic risk assessment focusing on asset value at risk. However, asset value at risk does not cover all aspects of privacy and security of tokens.

ISO27005 [62] classifies eight threat types for information security risk assessments. They include physical damage, natural events, and loss of essential services. The others are disturbance due to thermal radiation, compromise of information, technical failure, unauthorized action and comprise of functions. Similarly, the RiskIT [64] classifies five risk contributing factors including external environment, internal environment, risk management capability, IT capability and IT related business capability. However, risk models need to be explicit and should define the risk contributing factors to be assessed and the relationships among those factors [102]. Risk models can increase the reproducibility of organization and repeatability of risk assessments process. The traditional risk assessment approaches have no risk models for IDMSs.

Chapter 4

Part II
Table of Content for the Main Publications

Chapter 5

Article I
Taxonomy of Security Risk Assessment Approaches for Researchers,

in IEEE Proceedings of 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)

Ebenezer Paintsil

Abstract

This article introduces a taxonomy of security risk assessment approaches. The taxonomy is based on the challenges in the information system security (IS-Security) risk assessment discipline. Traditionally, classification schemes for IS-Security risk assessment approaches are motivated by business needs. They aim at offering management an effective tool for selecting methods that meet their needs rather than meeting research needs. Researchers may value new ideas, how to improve the approaches in the existing paradigms, and how to create a new paradigm to solve the unsolved problems of the existing paradigms more than business interests. The taxonomy proposed in this article aims at guiding researchers to choose research areas, and to discover new ideas and paradigms in the IS-Security risk assessment discipline.

5.1 Introduction

The early 90s marked the beginning of information technology (IT) era, where IT departments became an integral part of organizations [44]. Up to this era, IT was mainly used for automating business operations. However, how to effectively utilize information became the prime focus in the information systems' (ISs) era. This led to a shift from IT-Security to IS-Security.

IT-Security targets physical asset protection where ensuring continuous availability of computer systems is the paramount interest [44]. It is asset-oriented because it focuses on protecting physical assets. On the other hand, IS-Security aims at ensuring confidentiality, integrity and availability of information systems. It focuses on protecting both tangible and information assets.

The IT era steered the development of a number of IT-Security risk assessment approaches or methods. Currently, we have over 200 IT-Security risk management methods [94]. These methods "greatly differ in approach, complexity of usage, level of detail and applicability to organizations of different sizes and business models" [142]. They cover different phases of risk assessment as well as concentrate on different aspects, problems or business areas [146]. The inconsistent requirements make method selection a daunting task for organizations, hence the need for classification schemes.

A number of classification schemes are meant to offer management an effective tool for selecting suitable methods [161], [39]. This motivation sometimes leads to extensive focus on the utility of the risk assessment methods and business requirements in the classification schemes thereby limiting their technical appeal or value. A classification scheme that focuses on new ideas, the problems in the existing IS-Security risk assessment research

paradigms as well as how to create a new paradigm to solve the unsolved problems of the existing paradigms may be more valuable for researchers than business interests.

This article introduces a taxonomy of risk assessment approaches for researchers in order to assist them in identifying research areas in the IS-Security risk assessment discipline.

The rest of the article is organized as follows: Section 5.2 discusses the related work focusing on the classification schemes for risk assessment approaches. Section 5.3 discusses the challenges in the IS-Security risk assessment discipline. The discussion forms the basis for the taxonomy. Section 5.4 introduces the taxonomy for risk assessment methods. Section 5.5 applies the taxonomy to classify risk assessment approaches and explains how the taxonomy can guide IS-Security risk assessment research. It evaluates the taxonomy using a set of satisfactory criterion. Finally, Section 5.7 concludes and states the applications of the taxonomy.

5.2 Related Work

We can describe a number of current classification schemes for IS-Security risk assessment methods or approaches as organization-oriented because they focus on meeting business needs rather than research needs. Campbell et al. [125] introduced an organization-oriented classification scheme based on the traditional IT-Security risk assessment methods. The scheme identified three main approaches – temporal, functional and comparative. The temporal approach requires formal system modeling and testing in order to estimate the risk. The comparative approach requires system owners to compare their system and/or procedures with an explicit standard. The functional approach combines some aspects of the comparative and temporal approaches. However, the scheme lacks a classification regarding the elements of the risk assessment approach [146] and provides no explicit direction for research.

Another organization-oriented scheme is introduced by the ENISA ad hoc Working Group on Risk Assessment/Risk Management [39]. Their scheme can generate processes that can enable an organization to consider which information risk assessment and management methodologies are suitable. Risk exposure and risk impact determine the most appropriate risk assessment methodology for an organization. A similar organization-oriented scheme is introduced by Vorster and Labuschagne [161]. The framework focuses on the needs of organizations and provides no explicit direction for research.

A classification scheme for IS-Security risk methods was introduced by Siponen [139]. The organization-oriented scheme identifies five classes – checklists, IS security standards, IS security maturity criteria, risk management, and formal methods. However, the five classes are not subdivided missing any further distinguishing characteristics for decision support [152].

5.3 Challenges of IS-Security Risk Assessment

This section highlights some of the challenges or problems of the current IS-Security risk assessment approaches that will form the basis for the subsequent taxonomy.

The challenges or problems of IS-Security risk assessment approaches include extensive focus on physical asset protection, and lack of data on past events for risk estimation. Additional challenges are how to handle uncertainty in risk estimation, and lack of appropriate risk communication and identification methods.

IT-Security is asset-oriented because it focuses on physical asset protection [44]. Physical asset protection was significant in ensuring continuous availability of computer systems in the information technology era. The choice of security controls were influenced by the availability requirements for the computer system.

On the other hand, the requirement for IS-Security are confidentiality, integrity and availability. Hence, the asset-oriented IT-Security risk assessment approaches may not be

suitable for IS-Security. An appropriate risk assessment approach for IS-Security is the top-to-bottom or requirement-oriented approach [44]. The requirement-oriented approach analyzes security requirements of a targeted system in order to determine the 'amount' of security needed to protect an information asset.

The traditional IT-Security risk assessment approaches use three main estimation methodologies [62], [13] (qualitative, quantitative and semi-quantitative) to estimate risk of a targeted system. How to obtain exhaustive public and internal data on past events for risk assessment is always a challenge [149], [146]. The traditional way of obtaining such data is to rely on the intuitions of risk assessors or system stakeholders [13]. This makes the risk estimation error prone.

To ensure the reliability of the risk assessment process, a method that has low or no subjective influence should be used for the risk estimation. However, even such methods could be characterized by considerable degrees of uncertainty. For example, correct or absolute probabilities are unknown because of lack of data on new risks or information security weaknesses, future events etc. [62], [13]. Hence, there could be a large difference between an estimated or relative probability and absolute probability. This difference is the uncertainty in the estimated risk. How to compute this uncertainty is a challenge in IS-Security risk assessment.

Risk assessment approaches often document risk findings in text and tables [58]. I refer to the text and tables as structured lists. Structured lists are easy to create but difficult to comprehend because they take a long period of time to communicate security risk findings [167], [58]. To enhance communication, model-based approaches are recommended. Unlike structured lists, model-based approaches do not hide important information among irrelevant details thereby facilitating early discovery of risk elements [167].

However, current model-based approaches use static and informal modeling techniques to model the risk of a targeted system. Informal and static models are imprecise, ambiguous and cannot effectively communicate the dynamic behaviors of the system, actors and adversaries to system stakeholders. Furthermore, we cannot formally validate the behavioral correctness of the targeted system model because of their informal nature. The behavioral correctness of the targeted system model is validated through the subjective judgment of the system stakeholders. This makes the risk assessment process error prone.

Another model-based IS-Security verification technique is UMLSec [71]. Unlike semi-formal model-based approach such as CORAS [58], UMLSec provides formal model of a targeted system before risk verification. However, the formal specification is mathematically based. Mathematically based formal methods are expensive, have difficult learning curves, and cannot effectively communicate risk to non-technical stakeholders [78]. An executable approach (such as Colored Petri nets (CPNs) [78]) that hides the complex mathematics from non-technical stakeholders may be less expensive and communication enhancing.

Notably, formal model-based approach such as UMLSec and its variants may not be regarded as mainstream security risk analysis approaches [152]. They may be classified as secure IS approaches. However, we cannot rule them out because of the importance of privacy, and the influence of privacy by design principles [66] in information security today. The privacy by design principles advocate for privacy concerns to be identified and addressed from the first principle. In other words, privacy requirements should be verified at the system design phase and should be treated as functional requirements. In addition, the design stage of a project is the best stage to complete privacy impact assessment (PIA) [60].

Risk identification is about activities to identify assets, the assets' value, assets' vulnerabilities as well as threat to the assets and controls that can mitigate the threats [62], [44]. IS-Security risk identification is difficult because assets are located in distributed environment and may be intangible thereby making it difficult to value them accurately [44]. Moreover, security risk occurs in a dynamic and distributed environment where new laws,

regulations, culture, attacks and conditions keep on changing. Determining the correctness and completeness of identified events, assets, vulnerabilities, attackers etc. in such environment is a great challenge. Furthermore, risk identification processes may consider events in isolation. The isolation of interconnected events to simplify risk estimation may negatively affect the reliability of the results [146]. However, how to develop effective methods for combining the isolated events is a challenge.

5.4 Taxonomy of IS-Security Risk Approaches

This section introduces a taxonomy of IS-Security risk assessment approaches. The taxonomy is based on some of the challenges or the problems of current IS-Security risk assessment approaches discussed in Section 5.3.

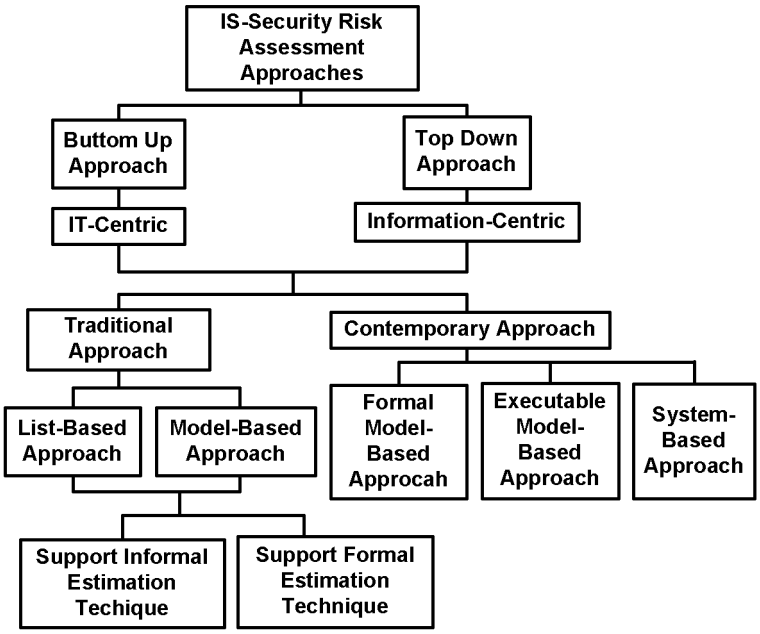

Figure 5.1: Taxonomy of Risk Assessment Approaches

Figure 5.1 is a taxonomy of IS-Security risk assessment approaches. The Top Down approach refers to the top-to-bottom approach proposed by Gerber and von Solms [44] while the Bottom Up approach refers to the traditional IT-Security risk approaches. The IT-Security risk assessment approaches are asset-oriented. They assess risk by identifying an asset and its value, and the likelihood of an unwanted incident and its consequence for the asset [91], [44]. A set of controls that will reduce the risk to the asset to an acceptable level is suggested.

In contrast, the Top Down approach is requirement-oriented. It focuses on establishing the security needs of a system or an organization. The security needs are based on the requirements for confidentiality, integrity and availability as well as risk to infrastructure, legal, and regulatory or statutory requirements [44]. The security needs determine the level of security required by an organization or system. For example, a bank will require high security level while a small retail business may require medium security. To satisfy the security level of an organization or a targeted system, a security requirement analysis is undertaken to determine and select the suitable controls for the organization or the targeted system.

The Top Down approach directly relates to Information-Centric risk assessment approach while the Bottom Up approach directly relates to IT-Centric risk assessment approach.

In Figure 5.1, the IT-Security or Information-Centric approaches are subdivided into Contemporary and Traditional approaches. Public and internal data on past events is the primary data source for the Traditional risk assessment approaches. However, exhaustive public and internal data on past events for risk assessment is not always available [149], [146]. Hence, the intuitions or opinions of system stakeholders or risk assessors are solicited for the risk estimation [162]. Consequently, Traditional risk assessment approaches have very high subjective influence because they may rely on the intuitions of risk assessors or system stakeholders [166].

In contrast, the Contemporary risk assessment approaches rely on the system characteristics or behavior to estimate risk. Data is obtained by modeling, testing or formally verifying the security properties of a targeted system. In other words, subjectivity does not influence the risk estimation process. Repeated risk assessment of a targeted system by different risk assessors may yield the same results (assuming there is no real change in the targeted system). On the other hand, repeated risk assessment by different risk assessors may yield different results in the Traditional risk assessment approaches since they may rely on the intuitions of a risk assessor.

In Figure 5.1, the Traditional approach is subdivided into two categories – list-based and model-based categories. The list-based approach relies on standards, best practices, compliance checks, checklists or recommendations [60], [152] to estimate the security risk of a targeted system. The recommendations, standards or checklists guide the risk assessor through the risk assessment process. The risk assessor estimates the risk by comparing the targeted system and/or procedures with the given standard, recommendations or checklist. The risk assessment process involves no explicit creation of conceptual system model and may rely on the intuitions of the risk assessors. Examples of such approaches are the Risk IT [64] and PIA [60].

The list-based approaches are ambiguous and hide important information among irrelevant details thereby impeding early discovery of risk elements [156], [167]. It makes the risk assessment process costly and error prone. On the contrary, the model-based approaches enhance risk identification and communication among system stakeholders [58]. The system modeling assists the risk assessor and system stakeholders to understand the system. They employ explicit graphical or formal models to conceptually model the targeted system before risk estimation and evaluation. The conceptual models are informal because we cannot formally validate their behavioral correctness or fidelity. In other words, the intuitions of the system stakeholders are relied upon to validate the system model. An example of model-based methods is the CORAS [58] risk analysis method.

The Traditional risk assessment approaches are prone to errors because of the possible intuitive inputs in the risk estimation process. To minimize the uncertainty in the estimated risk as a result of the subjective inputs, we employ formal estimation techniques or methods such as Bayesian network model (BNM) [28], Game Theory and Monte Carlo simulation [30]. In this case, the Traditional approach is said to support formal estimation technique. The Traditional approaches that support none of these formal techniques or methods are said to support informal estimation. Such method may involve simple mathematics that does not minimize the uncertainty in the estimated risk in anyway or may rely on an expected value matrix for the risk estimation [161]. An example is the RiskIT risk approach [64].

The Contemporary approach is subdivided into three categories – formal model-based, executable model-based and system-based approaches. The formal model-based approach stresses on understanding of the targeted system [125]. It involves abstract system modeling and verification. The system's performance as a result of the verifications determines the risk but not the subjective knowledge of a risk assessor. An example of formal model-

based approach is the UMLsec [71] approach.

The formal model-based approach involves formal modeling and verification. The formal modeling is mathematically based. However, mathematically based formal methods are expensive, have difficult learning curves, and can hamper effective risk communication among non-technical stakeholders [78]. Hence, in contrast to the formal model-based approaches, the executable model-based approaches that rely on easy to learn modeling techniques may be preferred. They hide the complex mathematics from non-technical stakeholders in order to enhance communication among all system stakeholders. An example of the executable model-based approach is the colored petri nets (CPNs) approach by Stephenson [148].

The executable model-based approach may be able to address the system fidelity problem in the formal model-based approach because we can formally and visually check the correctness of the targeted system model with the help of the system stakeholders [77]. The executable model-based approaches can also simulate the dynamic system environment in order to adapt the risk analysis to suit the changing environment.

The formal modeling approaches rely on abstract system models for the risk estimation. It is difficult to ensure the fidelity of a system with abstract models. The system-based approach focuses on how to understand the real system without any abstraction. In other words, it verifies or tests a concrete system rather than an abstract model of it. It requires a prototype or concrete system implementation. An example of the system-based approach is emulation techniques [141]. Emulation techniques can monitor a system behavior without modifying the system. It is preferred to simulation when evaluating real world system but complex, less scalable and time consuming.

It is possible to combine the classes in the taxonomy in various ways in order to create a hybrid method. For example, Yacoub and Ammar [173] combined both the Contemporary and Traditional approaches to estimate the reliability risk at the early stage of development life cycle.

5.5 Application

This section classifies some of the common risk assessment approaches and demonstrates how the taxonomy in Figure 5.1 can guide IS-Security risk assessment research.

5.5.1 Classification of Risk Assessment Approaches

Table 5.1 is an example classification of some of the common security risk assessment approaches. The classification includes both privacy and security risk assessment approaches. A generic example is given in the system-based class. In general emulation techniques can be used in the system-based approach and their classification will depend on how the technique is applied.

5.5.2 Uses of the Taxonomy

We identify some criteria that can guide IS-Security researches. There is no attempt to be exhaustive in these criteria. The criteria are explained as follows:

Choice of applicable research paradigm: The taxonomy in Figure 5.1 introduces two main research paradigms – IT-Centric and Information-Centric. The Information-centric paradigm is based on the Top-Down or top-to-bottom approach while the IT-Centric is based on the bottom-up approach. A researcher may choose the IT-Centric research paradigm if he is interested in asset protection or choose Information-Centric paradigm if he is interested in information protection.

Trade-off between correctness and completeness: The researcher needs to choose between correctness and completeness. The Contemporary risk assessment approach leads to formal modeling. It is not possible to model all possible attacks [125], therefore the

Table 5.1: Example Classification

	Bottom Up/IT-Centric	**Top Down/Info-Centric**
Traditional Approaches		
List-based with formal support	Game theoretic approach [132], CRAMM [136], IS business model [151], ISRAM [73]	EBIOS [130]
List-based w/o formal support	ISO27005 [62], OCTAVE [4], RiskIT [64], Delphi approach [162]	
Model-based with formal support	Attack graph-based [163], Dynamic incentives method [7], CORAS [58]	LBRT [30]
Model-based w/o formal support	FTA [58]	Goal-oriented [38], STRAP [120]
Contemporary Approaches		
Formal model-based	Architecture-level approach [173]	Property-oriented approach [166], UMLsec [71]
Executable model-based	CPNs approaches [148], [33]	CPNs approach [154]
System-based	Emulation techniques	Emulation techniques

Contemporary approach may lead to incomplete results. The opposite may be true in the Traditional approach. This means if the researcher wants to capture and estimate all possible risk, then Traditional approach may be more adequate than Contemporary approach. However, the Contemporary approach provides reliable results if the fidelity of the system model can be guaranteed.

Trade-off between fidelity and rigor: Rigor refers to effective use of mathematical formalisms to justify or measure the performance of an artifact or a targeted system [51]. The executable model-based approach may rely on visual aids, simulation tools and automatic validation techniques to validate a targeted system model. In addition, complex mathematics is handled by tools which assist in ensuring the correctness of the model. However, the freedom to construct complex mathematical proofs is constrained by the capabilities of the modeling tools. The researcher has to choose between freedom to operate (rigor) and the use of tools to ensure correctness of the system model (fidelity).

Trade-off between effective communication and rigor: The executable model-based approaches can communicate risk better than the formal model-based approaches because they have visual aids and tools. For example, a CPNs' model can be extended with animations layer for effective communication [77]. The researcher has to choose between the benefit of rigorous mathematics and effective communication.

Trade-off between informal and formal communication techniques: Although the executable model-based approaches may provide effective communication, it may require more effort to construct the formal model and the communication tier [77]. The model-based approach involves the construction of informal diagrams or models [91]. Although the informal models are easier to create than the executable models, we cannot guarantee their correctness.

5.6 Evaluation of the Taxonomy

The evaluation of the taxonomy is based on some of the criteria used by Alvarez and Petrovic [8] and Hernandez-Ardieta [50].

Accepted: A taxonomy should be logical and intuitive so that its categories could become generally approved. The taxonomy in Figure 5.1 is based on the challenges in IS-Security risk assessment discipline. These challenges have already been acknowledged in many IS-Security risk assessment literature including [13], [146], [149] and [44]. The rationale behind each category in the taxonomy is clearly explained to facilitate its approval and acceptance.

Exhaustive: The categories of taxonomy should include all possibilities. The proposed taxonomy considers both IT-Security and Information security risk assessment approaches. Currently, the two categories are the main approaches in the IS-Security risk assessment discipline. Furthermore, the instantiation in Table 5.1 shows that a number of risk approaches could be classified using the taxonomy.

Mutually exclusive: The categories in the taxonomy should not overlap. The examples in Table 5.1 do not overlap, however it is impossible to guarantee if the taxonomy is mutually exclusive.

Comprehensible: A taxonomy should be understood by experts, as well as those who only have an interest in it. The terminologies used in the proposed taxonomy are derived from IS literature and explained from basic principle, therefore it should be possible for experts and those who are interest in the field to understand.

Determinism: The procedure of classifying must be clearly defined. The proposed taxonomy is based on the challenges in the IS-Security assessment discipline.

Repeatable: Repeated applications should result in the same classification, regardless of who is classifying. The proposed taxonomy is based on the challenges in the IS-Security assessment discipline. Repeated classification may yield the same results.

Focused: A taxonomy should be focused in order to be useful and specific to a certain field of knowledge. This taxonomy is focused on IS-Security risk assessment approaches and how to guide research in the discipline.

Useful: A taxonomy could be employed to gain insight into the field of inquiry. The proposed taxonomy can guide new and old IS-Security risk assessment research to choose their research paradigm.

5.7 Conclusion

Classification schemes for information system security (IS-Security) risk assessment approaches are often overshadowed by business interest thereby making them less useful for guiding IS-Security research. In addition, the schemes do not consider the fine details that underline the IS-Security risk research. This article introduces a taxonomy of IS-Security approaches based on the challenges or problems in the current IS-Security risk assessment discipline and shows how the taxonomy can guide IS-Security risk assessment research. The taxonomy makes a clear distinction between IT-Security and information security and focuses on paramount areas of IS-Security risk assessment research. The taxonomy can guide new and old IT-Security researchers in their attempt to create innovative risk assessment artifacts.

Chapter 6

Article II
A Taxonomy of Privacy and Security Risks Contributing Factors,

in Privacy and Identity Management for Life, Springer LNCS Proceedings, vol. 352 of IFIP Advances in Information and Communication Technology
Ebenezer Paintsil and Lothar Fritsch

Abstract

Identity management system(s) (IDMS) do rely on tokens in order to function. Tokens can contribute to privacy or security risk in IDMS. Specifically, the characteristics of tokens contribute greatly to security and privacy risks in IDMS. Our understanding of how the characteristics of token contribute to privacy and security risks will help us manage the privacy and security risks in IDMS. In this article, we introduce a taxonomy of privacy and security risks contributing factors to improve our understanding of how tokens affect privacy and security in IDMS. The taxonomy is based on a survey of IDMS articles. We observed that our taxonomy can form the basis for a risk assessment model.

6.1 Introduction

A token is a technical artifact providing assurance about an identity. Tokens help authenticate and establish the identity of the end-users. They also help the end-user to remember their identifiers and facilitate information flow in identity management system(s) (IDMS). Figure 6.1 depicts an example of a simple identity management system. The identity provider (IdP) is an organization that collects the personal information of the end-user and creates a digital identity or identities for it. An IdP issues or helps the end-user to choose an identifier representing the digital identity. The end-user can then use the identifier(s) to identify or authenticate herself to a service provider (SP) in order to access an online service or resource. Each SP may require a different identifier or has a peculiar identification or authentication scheme. The number of identifiers may grow depending on the number of SPs the end-user interacts with and the kind of services or resources the end-user subscribes. The growth may reach a point where the end-user could no long remember all the numerous identifiers and their corresponding SPs. To solve this identifier management challenge, we either employ a software agent to store and select the correct identifier for a SP as in [97] or a hardware device such as a smart card to store and facilitate the selection of the correct identifier for a SP.

In the physical world, identity tokens consist of identifying information or identifiers stored in a physical device such as credit card, passports, a silicon chip and a magnetic stripe [31]. We also have virtual or software identity tokens such as the Microsoft information card (InfoCard or CardSpace) technology. The InfoCard technology consists of identity metadata stored in a visual icon. The identity metadata point to or associate with a digital identity. The digital identity in this case represents the identity of the end-user. We can also find other kinds of tokens such as user name tokens, binary tokens, nonce, XML based

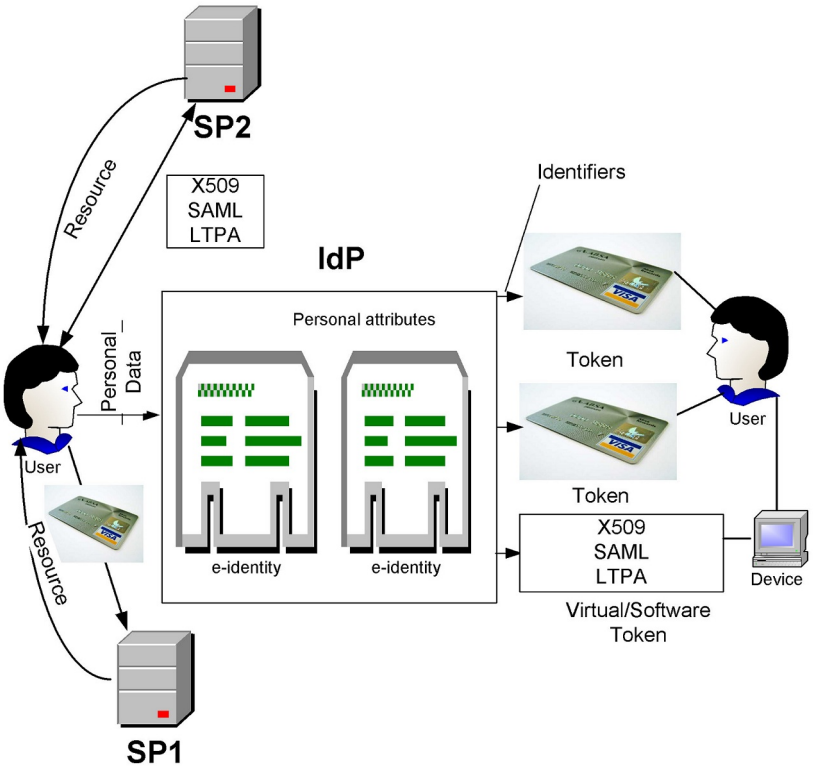

Figure 6.1: A Simple Identity Management System, SP1 means service provider 1, SP2 means service provider 2 and IdP means identity provider

tokens and custom tokens such as Keynote [23], [56]. A token can consist of a piece of data, a mechanism, an algorithm, an assertion or a credential.

The construction and different uses of tokens contribute to privacy and security risks in IDMS [23]. Camenisch and others have suggested anonymous credential systems as a means of enhancing privacy in IDMS. In such systems, different uses of identity tokens by the same user are unlinkable. However, apart from unlinkability, tokens contribute to security and privacy risks in diverse ways. For example, function creep is as a result of using a token for unintended purpose. Anonymous credential systems also lack practical use and may not be compatible with already deployed IDMS [24], [169]. They may rely on a master secret to protect all the tokens, however accidental disclosure of the master secret could lead to identity theft, linkability and eventual privacy or security risk. Thus, the characteristics of tokens can contribute to privacy and security risks even in anonymous credential systems.

In this article, we introduce a taxonomy of privacy and security risks contributing factors in order to understand the impact of the characteristics of tokens on privacy and security in IDMS. In addition, we introduce the applications of our taxonomy.

We organize this article as follows. In Section 2, we introduce existing works on the impact of tokens on IDMS. Section 3 describes our taxonomy of privacy and security risks

A user name token consists of a user name and optionally, password information for basic authentication. A nonce is a random generated tokens used for mitigating replay attacks. Binary tokens are non-XML based security tokens represented by binary octet streams. Examples of binary tokens are X.509 certificates, Kerberos tickets and Lightweight Third Party Authentication (LTPA) tokens. We represent XML based tokens by extensibility markup language. Examples of XML based tokens are, Security Assertion Markup Language (SAML), Services Provisioning Markup Language (SPML) and Extensible rights Markup Language (XrML)

contributing factors in detail. We introduce the applications of our taxonomy in section 4. Section 5 states the conclusion and future work.

6.2 Related Work

There is a large body of literature on identity tokens and how they may contribute to security and privacy risks. However, our work organizes tokens according to their contribution to privacy and security risks.

In [85] Lutz and Del Campo used a custom identity token to bridge the gap between privacy and security by providing a high level of privacy without anonymity. They designed a scheme to prevent replay attack and ensure that personal data is not sent to a foreign domain. They did not focus on the effect of the characteristics of tokens on privacy and security risks within a domain.

In addition, tokens are employed to facilitate single sign-on authentication in federated IDMS [169]. However, this work focuses on a better way of constructing a token but not on the effect of the characteristics of tokens on privacy and security risks.

The identity mix (idemix) scheme proposed in [23], employs anonymous tokens to enhance privacy. The idemix scheme has three main parties. The end-user obtains a pseudonym in a form of an anonymous token from the identity issuer. The verifier verifies the credentials. The users can authenticate with a verifier without revealing their pseudonyms. The characteristics of tokens such as token secrecy, can affect this scheme, as one can misuse a pseudonym if the token's secret is inferred or revealed.

Peterson introduces factors for asset value computation and stresses the importance of asset in risk calculation [123]. We can use such factors to quantify the contributions of tokens to privacy and security risks. Peterson derived the asset value of tokens from their loss, misuse, disclosure, disruption, replacement value, or theft. However, this is just an aspect of the risk contributing factors.

Solove introduces a taxonomy of privacy [143]. He introduces a high-level privacy risk contributing factors. They include data processing, data collection, data transfer and invasion. Nevertheless, the taxonomy is a high-level explanation of privacy principles without paying particular attention to the contributions of tokens to privacy risk.

Privacy Impact Assessment (PIA), a framework for assessing the impact of personal data processing on privacy before an information system is implemented, is introduced in [60]. The PIA is a compliance check of personal data processing to privacy laws, policies, or regulations. It specifies the requirement for privacy risk assessment without any explicit assessment technique or method. Our work differs from PIA in our introduction of an explicit privacy and security risks assessment technique.

The characteristics of tokens affect privacy and security in IDMS. A token characteristic include but not limited to its usages, how it is built, designed or constructed and how it is chosen. Security protocols are often concerned with how to build or construct a security token that can enhance privacy and security in IDMS [23], [85] , [86], [169]. The emphasis may be placed on formal verification of privacy and security risks caused by how tokens are built, designed or constructed. Moreover, tokens contribute in many different ways to security and privacy risks in IDMS. For instance in analogy to [123] tokens can be a good source of security and privacy risk metrics since they can consist of metadata about identity definitions.

6.3 The Taxonomy

This section introduces our taxonomy of privacy and security risks contributing factors. We based our factors on a literature survey of scientific articles in IDMS such as [123], [9], [23], [48], [65], [168] and many more. Our taxonomy, for the time being, avoids the Pfitzmann-Hansen terminology [124] with its definitions of e.g. linkability and observability, as the

terms therein are defined on the background of anonymous communication and information sparsity. We feel that these terms need to be analyzed from our perspective on handling electronic identifiers and their risks. Our future work aims at aligning or re-defining the Pfitzmann-Hansen terminology to be meaningful in our context. Following this, we depict our taxonomy in Table 6.1. The taxonomy is non-exhaustive list of the characterization of tokens according to the manner in which they contribute to privacy and security risks. We explain the meaning of each contributing factor as listed in Table 6.1.

Risk Contributing Factors	Parameters
Token Mobility	*copyable, remotely usable, concurrently usable, immobile*
Token Value at Risk	*loss, misuse, disclosure, disruption, theft, replacement value*
Token Provisioning	*creation, editing, deletion*
Token Frequency & Duration of Use	*Uses per year, life-time, multiple times, one-time*
Token Use & Purpose	*original, unintended*
Token Assignment & Relationship	*forced, chosen, jointly-established, role, pseudonymity*
Token Obligation & Policy	*absence, present, functionality*
Token Claim Type	*single, multiple*
Token Secrecy	*public, inferable, secret*
Token Security	*origination, identification, validation, authentication, authorization*

Table 6.1: Taxonomy

Token Mobility: This category indicates the degree of mobility of a token. The degree of mobility refers to how easy it is to copy a token or its content, the physical constraints regarding the movement of the token, among others. For example, the content of a low cost RFID tag with no additional security could easily be read by anyone, but a high cost RFID tag that comes with additional security may ensure that only authorized readers have access to its content. Various forms of mobility create risk in IDMSs. We assess the contributions of token mobility to privacy and security risk according to the following:

1. *Copyable*: the token can be copied with limited effort.
2. *Remotely usable*: the token can be used for remote identity management.
3. *Concurrently usable*: the token can be used concurrently in many parallel sessions, transactions, or applications.
4. *Immobile*: a token is not 'mobile', if it must be physically present.

Token Value at Risk: Finding assets and the value of the assets at risk is an important part of risk analysis [64]. Tokens are assets and their value at risk can contribute to privacy and security risk. Thus, we can quantify the risk of using tokens by assessing the significance of the token or the value of the token to the operation and security of the IDMS. We classify the value at risk [123] as follows:

1. *Loss*: value at risk when token is lost.
2. *Misuse*: value at risk when token is used in wrongful ways.
3. *Disclosure*: value at risk when token or token-related information gets known by someone else.
4. *Disruption*: value at risk when token doesn't function.

5. *Theft*: value at risk when a token is stolen.

6. *Replacement value*: cost (effort, resources, time) to replace a token.

Token Provisioning: The amount of personal information collected during tokens' creation could contribute to privacy risk [87], [21]. Data minimality principles prohibit excessive personal data collection. In addition, tokens that contain excessive personal information may be used for other unintended purposes. There should be a means by which tokens can be updated in order to ensure data integrity. Further, there should be a means by which a token can be deleted or destroyed when the purpose for its creation is no longer in the interest of the end-user. This would ensure the privacy of the end-user. Therefore, we distinguish the following phases of the token provisioning life cycle: *creation, editing and deletion*.

Token Frequency & Duration of Use: The underlying IDMS information flow protocol could determine if multiple use of the same token is susceptible to privacy and security risks [57]. Different uses of even specifically constructed tokens remain linkable if the underlining IDMS's information flow protocol is not designed to prevent such privacy risk. IDMSs that allow a service provider (SP) and an identity provider (IdP) to share information or use the same token repeatedly may be susceptible to profiling. A token used in association with life time identification or multiple times causes high risk of secondary use and profiling [41]. We divide tokens into three categories – a token chosen or assigned for life (life-time token), a token that can be used multiple times but not for life (multiple times token) and a token that is valid for a session (one-time token).

Token Use & Purpose: Purpose specification is an important privacy principle. It requires that personal information be collected for a specific purpose and used in an open and transparent manner [49]. Personal information should not be used for purposes other than its original purpose unless informed consent is given. Since tokens may carry personal information and are the gateway to personal data [103], any form of function creep or misuse could contribute to a privacy or security risk. Function creep occurs when a token is used for unintended purposes. The abuse of the purpose of a token or how the purpose of a token is achieved may contribute to privacy and security risks.

Token Assignment & Relationship: The need for user-centric IDMSs emphasizes the importance of token assignment. User-centric IDMS offer the end-users the ability to control their personal information and enforce their consent by determining the choice of their identity tokens [19].

The origin of and control over tokens contribute to privacy or security risk. A token can be chosen by a person, jointly established or forced upon by an authority. They can relate to a role or pseudonym rather than a person [57].

The token assignment and relationship risk contributing factors determine the privacy risk if a token is forced on an end-user, chosen by an end-user, or jointly-established with the end-user. It also assesses the end-user's privacy risk if a token is chosen under a pseudonym or a role [57].

Token Obligation & Policy: Some IDMSs give the end-users an opportunity to specify policies regarding how their tokens should be used [10]. In the absence of a policy, functionality such as the use of anonymous credentials and selective disclosure of attributes can mitigate privacy and security risk. Therefore, this category checks if a policy or functionality exists for privacy protection.

Token Claim Type: We can enforce the security of identifiers stored in a token or protect the misuse of identifiers by attaching a secret or a claim to the tokens. The type of claims are generally classified as *"something we know"* (a secret e.g. password), *"something we are"* (something we cannot share with others, e.g. finger print, iris etc.) and *"something we have"* (e.g. possessing a smart card, key card or a token etc.) [87]. Each claim type is a factor. Token claim type considers the effect of the number of factors used for securing a token in IDMSs. Some tokens may require no additional secrets in order to protect its content. We

refer to such tokens as single claim tokens since the possession of the token itself is a factor. A token that requires a factor such as a personal identification number (PIN) in order to access its content is referred to as a two-factor or multiple claim token.

The number of authentication factors may determine the complexity and security of the token. For example, current credit cards may require no additional secret PIN when used for online transactions. We regard such a token as a single factor authentication token.

Token Secrecy: Tokens have claim type as discussed above. The secrecy or the constraints of authentication factors such as physical presence can contribute to security or privacy risk in IDMS. The nature of the secret and constraint of the additional authentication factors may determine the "claim type".

A token secret that can easily be guessed is classified as an *inferable secret*. A token's secrecy is *public* if its additional authentication factors or claims are known by a number of people, group and organization or exist in many databases, at the disposal of an unknown number of persons and organizations. We classify the secrecy of a token as *secret* if it is private or not shared with others because it is linked to a valuable resource such as a bank account [23]. A possible example of private secrets is a private cryptographic key and a PIN.

Token Security: The security of a token can contribute to the security of the IDMSs. In order to ensure token security, there should be a means of ensuring the validity, identity and legitimacy of the token. Similar to Jan and Van [87], we describe the security of tokens as follows:

1. "Origin": the token is issued by the indicated legitimate authority.
2. "Identification": the token identifies the subject-the entity possessing the token.
3. "Validation": the token has not expired; its lifespan is within the validity period or has passed the validity test.
4. "Authentication": the token belongs to the entity presenting it to the IDMS.
5. "Authorization": the token grants relevant permissions to the entity possessing it.

6.4 Applications of the Taxonomy

We can fairly estimate the privacy and security risks of an IDMS based on the characteristics of token as explained in our taxonomy. This section explains the possible application of our taxonomy for privacy and security risks assessment. We state the possibilities of conducting privacy and security risks assessment based on the characteristics of tokens used in the IDMS. We follow the following steps:

Known Security Risks in IDMS: We examine the IDMS in order to determine all the possible system tokens. We assess how the characteristics of the system tokens may affect the achievement of the information security objectives of the IDMS. Thus, which of the characteristics of tokens listed in table 6.1 contribute to the breach of information security such as unavailability, loss of confidentiality, repudiation and lack of integrity [140] and in what way do they contribute to security risk. For example we may assess the impact of multi-factor authentication scheme on the security of the IDMS. We may assess if the IDMS employs a two-factor authentication tokens or a three-factor authentication tokens. A two-factor authentication token is simpler but may provide less secure authentication [87].

We also assess how the security of the token contributes to security risk of the IDMS. In other words, we examine if the IDMS has a method for checking the security of the tokens. That is, we assess if the IDMS has a mechanism to check the legitimacy of the token, validity of the token, identity of the holder of the token, authenticity of the token and level of authorization the token has.

Finally we assign or compute the risk impact of token security, token secrecy etc of the IDMS.

Known Privacy Risks in IDMS: Similar to security risk assessment, we examine the IDMS in order to understand the information flow of the system. We examine how the IDMS tokens are designed and used. We assess if the information flow in the IDMS can contribute to the misuse of personal data, accidental disclosure of personal data etc. We determine if these privacy risks are caused by the construction of the IDMS tokens using the risk contributing factors listed in table 6.1. We then determine the impact and the possibility of the known risk occurring. For example, how will the token mobility contribute to accidental disclosure of personal data and what will be the impact of accidental disclosure on the IDMS?

Stakeholders Risk in IDMS: Usually tokens are created to reflect the needs of the system stakeholders. However, lack of extensive taxonomy of token characteristics to guide this process may contribute to future privacy and security risks of the IDMS. This taxonomy will aid the stakeholders of the IDMS to ask the necessary questions before designing or consenting to use any identity token. Furthermore, the stakeholders can now, before they take an IDMS that was designed for a different purpose into a new application context, thoroughly analyze its properties and the possible risks and consequences. For example, the stakeholders can now assess how an IDMS with high token mobility contribute to function creep. How does function creep affect their interest such as reputation? Furthermore, how does multiple claim authentication token inconvenience the end-user and how does it protect against identity theft in IDMS? We can also examine how token assignment contributes to privacy and security risks from the stakeholders' point of views. For example, user-centric IDMS may allow the end-users to exercise selective disclosure by choosing tokens on their own. This enhances the privacy of the end-user but the other stakeholders may be deprived of information for auditing and accounting [15], [97].

6.5 Conclusion

Identity tokens are constructed without extensive assessment of their effect on privacy and security risks in identity management system (IDMS). The understanding of the contribution of tokens to privacy and security risks can aid in managing privacy and security risks in IDMS. This article defines a token and how tokens affect privacy and security in IDMS. We introduced a taxonomy of risk contributing factors for IDMS based on the characteristics of tokens and the application of the taxonomy for privacy and security risks assessment. We explained how the taxonomy contributes to privacy, security, and the stakeholders' risk in IDMS. Finally, we showed that tokens are rich sources of privacy and security risks metric for IDMS and can serve as a basis for privacy and security risks assessment model.

We intend to investigate the development of a privacy and security risks assessment model based on our taxonomy in our future work.

Chapter 7

Article III
Evaluation of Privacy and Security Risks Analysis Construct for Identity Management Systems,

in IEEE Systems Journal, Vol. 7, Issue 2
Ebenezer Paintsil

Abstract

This paper evaluates a taxonomy of privacy and security risk contributing factors with the Delphi method. The taxonomy was introduced in a previous work, and it is based on characteristics of tokens used in identity management systems (IDMSs). The taxonomy represents a construct for risk analysis in IDMSs. Constructs are concepts, terms, or vocabularies and symbols adopted or developed to describe, conceptualize or define the problems and solutions within a domain. We can determine the performance and utility of a construct through evaluation. Evaluation can determine constructs' completeness, simplicity, elegance, ease of use and understandability. Evaluation of a construct can be done with the Delphi method. The Delphi method solicits expert opinions on a subject matter in a structured group communication process. The Delphi evaluation of the taxonomy led to additional privacy and security risk contributing factors that were not covered in the initial taxonomy. Furthermore, the evaluation identified three key risk indicators and showed that the experts mostly agreed with our initial risk analysis construct for IDMSs.

7.1 Introduction

Identity management systems (IDMSs) create and manage identities of end-users [69]. The traditional role of IDMSs is to issue end-users with credentials and unique identifiers during the initial registration phase. The end-users can then use these identifiers and the credentials to authenticate to a system.

End-users' identifiers or credentials may consist of both pseudonyms and personally identifiable information (PII) with a variety of security features. They may be organized in the form of tokens in order to provide assurance about an identity. We define a token as a technical artifact providing assurance about an identity [117]. A token can be an identifier such as username, a claim such as a password, an assertion such as SAML token, a credential such as a X.509 certificate, or any possible combinations of these.

Our previous work [117] surveys how the characteristics of identifiers contribute to privacy and security risk in IDMSs. The taxonomy is a *construct* [51] for privacy and security risk assessment. A construct is a shared knowledge or a language for communication within a known domain. Constructs are concepts, terms, or vocabularies and symbols adopted or developed to describe, conceptualize or define the problems and solutions within a domain [90]. Constructs may be formal or informal and they may be established through consensus building.

According to the design science research (DSR) method, constructs need to be evaluated for completeness, simplicity, elegance, understandability and ease of use [90]. This paper applies the Delphi method to build and evaluate a security and privacy risk assessment construct. We derive the initial construct from a literature survey of characteristics of tokens used in IDMSs [117]. We then evaluate the construct through consensus building with a Delphi study [159]. A Delphi study solicits expert opinions on a subject matter in a structured group communication process. It is a collective human intelligence process and may be employed to develop and evaluate a construct [159].

This paper is organized as follows: Section 7.2 introduces related work. Section 7.3 gives the background and objectives of the study. Section 7.4 explains the Delphi method and how it applies to this study. We present the result of the study in Section 7.5 and discuss it in Section 7.6. Section 7.7 discusses shortcomings of the study. Section 7.8 aligns the result to a set of security and privacy goals in order to show its completeness. Finally, Section 7.9 concludes and discusses possible future work.

7.2 Related Work

Finding assets and their value at risk is an important part of risk assessment [64]. Tokens used in IDMSs are assets and their value at risk can contribute to privacy and security risk. Peterson introduces asset value computation factors and the importance of asset in risk calculation [123]. Asset value computation factors can quantify the contributions of tokens to privacy and security risk. Peterson derived the asset value of tokens from their loss, misuse, disclosure, disruption, replacement value, or theft. However, these factors are not comprehensive for privacy and security risk assessment.

The taxonomy by Solove [143] classifies privacy risk contributing factors. The factors include data processing, data collection, data transfer, and privacy invasion. However, Solove's taxonomy is quite abstract for thorough privacy and security risk assessments in IDMSs. This paper presents a specific construct for supporting risk assessment in IDMSs.

Yamada et. al. modeled the projected compensation for damages caused by privacy violations [174]. The objective of the model is to "provide organizations with a quantitative understanding of the latent risk involved in handling personal information". The metric estimates the value of leaked personal information based on risk contributing factors such as value of basic information, degree of information sensitivity and degree of ease of identification. This paper introduces a comprehensive construct for privacy and security risk assessment focusing on how tokens contribute to risk in IDMSs.

Similarly, Fritsch and Abie [42] proposed a model called *Return-on-Privacy-Investment* (ROPI) which states the effect that a particular investment has on the privacy-relevant value of an information system. The model is based on the *Return-on-Security-Investment* (ROSI) concept by Berinato [16] analogous to the ROPI. However, the model focuses on only one aspect of privacy and security risk estimation. We introduce a comprehensive construct for privacy and security risk analysis in IDMSs.

7.3 Study Background and Objectives

IDMSs rely on tokens to function [170]. Tokens facilitate identification, authentication and authorization of services and access to resources in IDMSs. Therefore, the privacy and security of IDMSs may depend on the characteristics of tokens. For example, a weak password can lead to a security breach in IDMSs. Consequently, an analysis of privacy and security of IDMSs based on characteristics of tokens is needed.

This paper evaluates the characteristics of tokens that can contribute to privacy and security risk in IDMSs. The initial taxonomy of these characteristics is shown in Table 7.1. The first column of Table 7.1 represents the categories of risk factors, while the second column is a mixture of positive and negative risk factors. We refer to factors that can enhance

privacy or security in IDMSs as positive factors while those that can contribute to privacy or security risk are the negative factors. Individual factors and categories in the table are explained as follows:

Table 7.1: Privacy and Security Risks Contributing Factors

Category of Factors	Risk Factors
Token Mobility	***copyable, remotely usable, concurrently usable, immobile***
Token Value at Risk	***loss, misuse, disclosure, disruption, theft, replacement value***
Token Provisioning	***creation, editing, deletion***
Token Frequency & Duration of Use	***Uses per year, life-time, multiple times, one-time***
Token Use & Purpose	***original, unintended***
Token Assignment & Relationship	***forced, chosen, jointly-established, role, pseudonymity***
Token Obligation & Policy	***absent, present, functionality***
Token Claim Type	***single, multiple***
Token Secrecy	***public, inferable, secret***
Token Security	***origination, identification, validation, authentication, authorization***

Token Mobility: This category indicates the degree of mobility of a token. The degree of mobility refers to how easy it is to copy a token or its content, the physical constraints regarding the movement of the token, among others. For example, the content of a low cost RFID tag with no additional security could easily be read by anyone but a high cost RFID tag that comes with additional security may ensure that only authorized readers have access to its content. Various forms of mobility create risk in IDMSs. We assess the contributions of token mobility to privacy and security risk according to the following:

1. *Copyable*: the token can be copied with limited effort.
2. *Remotely usable*: the token can be used for remote identity management.
3. *Concurrently usable*: the token can be used concurrently in many parallel sessions, transactions, or applications.
4. *Immobile*: a token is not 'mobile', if it must be physically present.

Token Value at Risk: Finding assets and the value of the assets at risk is an important part of risk analysis [64]. Tokens are assets and their value at risk can contribute to privacy and security risk. Thus, we can quantify the risk of using tokens by assessing the significance of the token or the value of the token to the operation and security of the IDMS. We classify the value at risk [123] as follows:

1. *Loss*: value at risk when token is lost.
2. *Misuse*: value at risk when token is used in wrongful ways.
3. *Disclosure*: value at risk when token or token-related information gets known by someone else.
4. *Disruption*: value at risk when token doesn't function.
5. *Theft*: value at risk when a token is stolen.

6. *Replacement value*: cost (effort, resources, time) to replace a token.

Token Provisioning: The amount of personal information collected during tokens' creation could contribute to privacy risk [87], [21]. Data minimality principles prohibit excessive personal data collection. In addition, tokens that contain excessive personal information may be used for other unintended purposes. There should be a means by which tokens can be updated in order to ensure data integrity. Further, there should be a means by which a token can be deleted or destroyed when the purpose for its creation is no longer in the interest of the end-user. This would ensure the privacy of the end-user. Therefore, we distinguish the following phases of the token provisioning life cycle: *creation, editing and deletion*.

Token Frequency & Duration of Use: The underlying IDMS information flow protocol could determine if multiple use of the same token is susceptible to privacy and security risks [57]. Different uses of even specifically constructed tokens remain linkable if the underlining IDMS's information flow protocol is not designed to prevent such privacy risk. IDMSs that allow a service provider (SP) and an identity provider (IdP) to share information or use the same token repeatedly may be susceptible to profiling. A token used in association with life time identification or multiple times causes high risk of secondary use and profiling [41]. We divide tokens into three categories – a token chosen or assigned for life (life-time token), a token that can be used multiple times but not for life (multiple times token) and a token that is valid for a session (one-time token).

Token Use & Purpose: Purpose specification is an important privacy principle. It requires that personal information be collected for a specific purpose and used in an open and transparent manner [49]. Personal information should not be used for purposes other than its original purpose unless informed consent is given. Since tokens may carry personal information and are the gateway to personal data [103], any form of function creep or misuse could contribute to a privacy or security risk. Function creep occurs when a token is used for unintended purposes. The abuse of the purpose of a token or how the purpose of a token is achieved may contribute to privacy and security risks.

Token Assignment & Relationship: The need for user-centric IDMSs emphasizes the importance token assignment. User-centric IDMS offer the end-users the ability to control their personal information and enforce their consent by determining the choice of their identity tokens [19].

The origin of and control over tokens contribute to privacy or security risk. A token can be chosen by a person, jointly established or forced upon by an authority. They can relate to a role or pseudonym rather than a person [57].

The token assignment and relationship risk contributing factors determine the privacy risk if a token is forced on an end-user, chosen by an end-user, or jointly-established with the end-user. It also assesses the end-user's privacy risk if a token is chosen under a pseudonym or a role [57].

Token Obligation & Policy: Some IDMSs give the end-users an opportunity to specify policies regarding how their tokens should be used [10]. In the absence of a policy, functionality such as the use of anonymous credentials and selective disclosure of attributes can mitigate privacy and security risk. Therefore, this category checks if a policy or functionality exists for privacy protection.

Token Claim Type: We can enforce the security of identifiers stored in a token or protect the misuse of identifiers by attaching a secret or a claim to the tokens. The type of claims are generally classified as *"something we know"* (a secret e.g. password), *"something we are"* (something we cannot share with others, e.g. finger print, iris etc.) and *"something we have"* (e.g. possessing a smart card, key card or a token etc.) [87]. Each claim type is a factor. Token claim type considers the effect of the number of factors used for securing a token in IDMSs. Some tokens may require no additional secrets in order to protect its content. We refer to such tokens as single claim tokens since the possession of the token itself is a factor.

A token that requires a factor such as a personal identification number (PIN) in order to access its content is referred to as a two-factor or multiple claim token.

The number of authentication factors may determine the complexity and security of the token. For example, current credit cards may require no additional secret PIN when used for online transactions. We regard such a token as a single factor authentication token.

Token Secrecy: Tokens have claim type as discussed above. The secrecy or the constraints of authentication factors such as physical presence can contribute to security or privacy risk in IDMS. The nature of the secret and constraint of the additional authentication factors may determine the "claim type".

A token secret that can easily be guessed is classified as an *inferable secret*. A token's secrecy is *public* if its additional authentication factors or claims are known by a number of people, group and organization or exist in many databases, at the disposal of an unknown number of persons and organizations. We classify the secrecy of a token as *secret* if it is private or not shared with others because it is linked to a valuable resource such as a bank account [1]. A possible example of private secrets is a private cryptographic key and a PIN.

Token Security: The security of a token can contribute to the security of the IDMSs. In order to ensure token security, there should be a means of ensuring the validity, identity and legitimacy of the token. Similar to Jan and Van [87], we describe the security of tokens as follows:

1. "Origin": the token is issued by the indicated legitimate authority.
2. "Identification": the token identifies the subject-the entity possessing the token.
3. "Validation": the token has not expired; its lifespan is within the validity period or has passed the validity test.
4. "Authentication": the token belongs to the entity presenting it to the IDMS.
5. "Authorization": the token grants relevant permissions to the entity possessing it.

The taxonomy in Table 7.1 represents the initial survey result or construct. The objective of this paper is as follows:

- To evaluate the privacy and security risk analysis construct introduced in Table 7.1.
- To identify the key risk indicators (KRIs) of the construct.

7.4 Method

"Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem" [159]. It collects and establishes an expert consensus on a subject matter. The Delphi method is suitable in situations where concepts are still vague, and the subject of study is not easy to quantify.

The method usually involves two or more rounds of questions and answers eliciting expert opinions [159]. The first round introduces the purpose of the study and the initial round of opening questions. The result of the first round is analyzed for consensus and disagreements. The second round of the study requires that the result of the first round is submitted back to the experts for a second opinion. The purpose of this is to ask the experts to reconsider their initial positions or disagreements. It is an opportunity for an expert to revise their opinions regarding a subject matter. The final round is the consolidation of opinions of the previous rounds. Traditionally, the Delphi method is regarded as a forecasting method. However, we have other application areas, including evaluation and modeling [159].

This work employs the Delphi method to build and evaluate constructs. The study started in December 2010 and was completed in May 2011. We began the study by identifying experts with different backgrounds who are actively working in the area of IT security and privacy. We sent an email to each of them explaining the purpose of the study and requested their participation. Initially, 17 experts agreed to participate in the study. However, only nine experts took part in the first and second rounds of the study. Three experts dropped out in the final round (round three) of the study.

The expert panel consisted of the following: two IT security professors, one from Gjøvik University College, Norway and the other from Radboud University, the Netherlands. A senior IT security expert from Fraunhofer institute, Germany and a PhD fellow from Linkoping University, Sweden. In addition, three researchers, one from $Unabhängiges\ Landeszentrum\ für$ Datenschutz (ULD), Germany and the others are from Technical University of Dresden (TUD), Germany, and IBM Research, Switzerland. Finally, two IT security practitioners, from Todos AB Sweden and A-SIT, Austria. The nine experts came from nine different institutions and six countries. They represented the academia, industry and research community. Two professors from Gjøvik University College, Norway and the Radboud University, Netherlands completed the study. The others who completed the study are the PhD fellow from Linkoping University, Sweden; the senior IT security from Fraunhofer institute, Germany; the researchers from IBM Research, Switzerland and Technical University of Dresden (TUD) Germany. Further details of the experts who agreed to be acknowledged can be found in the acknowledgments.

The experts were consulted through email. The first round of the study involved two tasks. The first task required the experts to list and explain in their own words the properties or the characteristics of tokens that may contribute to privacy and security risk in IDMSs. The list should be independent of any specific application. The second task required the experts to assess if the usages of tokens for identification, authentication or authorization are interchangeable without any consideration of threat and risk assessments. We focus on the result of the first task in this paper.

We categorized the results of the first round into risk contributing factors with a list of factors for each category. We then asked the respondents to provide their opinion on the categorization and the list of factors. We asked them to state whether they agree with the categorization and if not how they would adjust the result. In addition, we asked the respondents if they agreed with the explanations given to each category.

We integrated the initial survey result [117] with the results obtained from the expert panel in the third and final round of the study and asked the experts to give their opinions on the consolidated results. In addition, we asked them to rate the factors according to the following scale:

- Impact – How does the risk factor affect privacy and security in IDMSs.
- Measurement – How easy or difficult is it to measure or quantify the risk factor.
- Sensitivity – Is the risk factor an accurate measure of privacy and security risk of identifiers?
- Stakeholder – Can non-technical stakeholders understand the risk factors?

In order to ensure clarity, each question came with a short and easy to read introduction explaining its rationale. Moreover, the understanding of the questions was tested internally to ensure that possible ambiguities were removed. This was done by asking local experts to review the questions for clarity and understanding. In addition, the experts from the Delphi study who did not understand some of the questions wrote on their answer sheet how they interpreted them. For example, some of the experts interpreted "how to rate the possibilities to measure the risk factors" to mean how to quantify the risk factors. In addition, to the short introductions, the meaning of each risk category is explained in a separate

column during the second and the third rounds of the study (see Table 7.2). This was done to assist the experts in understanding the meaning of the categories. Some of the experts also gave out their phone numbers so that we could call them for further discussions.

Sample questions for the various rounds can be found in the Appendix.

7.5 Results

This section explains the result of the three-round Delphi study.

7.5.1 Round I

We analyzed and categorized the responses to Round I over themes or categories. We then listed the factors under each category as suggested by the experts. The result is shown in Table 7.2. In organizing the responses, we focused on the factors that may lead to consequences such as profiling, identity theft, linkability, availability, and function creep. For each response, we listed and categorized all the consequences and factors that may lead to those consequences.

Table 7.2: The Result for Round I

Rule No.	Category of Factors	Explanation	Factors
1	Frequency & duration of use	Number of times an identifier or its protected mechanism or authenticated data is used and how long it is used	one-time, multiple times, life time, single sign-on, multiple services
2	Purpose of use	When the identifier is designed for a particular purpose, application or context and is switched for another purpose it can lead to privacy and security consequences	application specific, privacy preserving, context specific, silo
3	Provisioning	An identifier constructed, updated or archived, possibly in conjunction with a large set of personal attributes or sensitive personal data such as age, data of birth creates privacy risk	limited personal attributes, overloaded personal attributes, sensitive personal attribute
4	Secrecy	The identifier can be inferred using the arithmetic or other properties of the identifiers	inferable, public, obfuscated recoverable, revocable
5	Claim Type	The nature of the authentication mechanisms used to secure the identifier	password, crypto key, biometric, challenge-response, single-claim, multiple-claims

The most frequently discussed theme or risk contributing category is "Frequency and duration of use". Seven out of nine respondents directly or indirectly identified factors that can be categorized as "Frequency and duration of use". The "Purpose of use" was the second dominant theme or category as four respondents identified it as one of the risk contributing categories. Some of the respondents indirectly identified "Provisioning" as one of the categories.

Other categories identified included "Secrecy" and "Claim type" of identifiers or tokens. Although "Secrecy" and "Claim type" categories and their factors were less frequent, we found them easy to understand and less controversial. Hence, we did not include them in the second round of the study but reintroduced them in the final round. The second round of the study focused on the controversial and most frequent categories such as "Frequency and duration of use" and provisioning (see Table 7.2) .

7.5.2 Round II

During the second round of the study, we submitted the result in the first three rows of Table 7.2 to the experts for further assessment.

We obtained the result in Table 7.3 at the end of Round II. Table 7.3 shows that the respondents did not agree with some of the factors in Rows 1 and 2 of Table 7.2. "Single sign-on" and "multiple services" factors are now categorized under "Purpose of use" instead of "Frequency and duration of use". In general we achieved consensus of opinions on the first two categories. There were divided opinions concerning the "Provisioning" category as some of the respondents did not clearly understand the factors in this category. We clarified this in the subsequent round of the study.

Table 7.3: The Result for Round II

Rule No.	Category of Factors	Explanation	Factors
1	Frequency & duration of use	Number of times an identifier or its protected mechanism or authenticated data is used and how long it is used	one-time, multiple times, life time
2	Purpose of use	When the identifier is designed for a particular purpose, application or context and is switched for another purpose it can lead to privacy and security consequences	application specific, single sign-on, multiple services context specific, silo
3	Provisioning	An identifier constructed, updated or archived, possibly in conjunction with a large set of personal attributes or sensitive personal data such as age, data of birth creates privacy risk	limited personal attributes, overloaded personal attributes, sensitive personal attribute
4	Secrecy	The identifier can be inferred using the arithmetic or other properties of the identifiers	inferable, public, obfuscated recoverable, revocable
5	Claim Type	The nature of the authentication mechanisms used to secure the identifier	password, crypto key, biometric, challenge-response, single-claim, multiple-claims

7.5.3 Round III

Table 7.4 is the result of the Task a of Round III. It combines the Delphi survey results and our initial model [117]. Two respondents considered the term "Mobility" as wrong terminology for the category because mobility usually refers to the positive property of users being able to take identifiers or tokens with them and use the tokens on the go and on multiple devices. Additionally, "Mobility" is closely related to (physical) protection of identifiers or tokens, which may be categorized under another risk contributing category. However, we maintained "Mobility" because the majority of the respondents did not disagree with the terminology.

The factors for "Provisioning" category were not clear according to some of the respondents. Therefore, we changed the factors based on the feedback received in this round. Furthermore, the factor "multiple services" under "Purpose of use" is replaced with "used with multiple services". This means the token or identifier is used with multiple services, but for more than just sign-on.

"Revocable" seems not to fit too well with secrecy because it is orthogonal to its secrecy. "Revocability" is rather a protection mechanism to invalidate secrecy, but can equally be applied to public, inferable, and other kinds of identifiers. However, the majority of the respondents agreed with the placement of "revocable" in the category secrecy. "Revocable" determines whether a token secret can be revoked or how easy or difficult it is to make a secret public.

In addition, we clarified the factors for "Provisioning" based on the experts' feedbacks. "Provisioning" factors consider the effect of creating, updating, deleting or archiving tokens with limited personal attributes, overloaded personal attributes and sensitive personal attributes.

Table 7.4: The Result for Round III

Rule No.	Category of Factors	Explanation	Factors
1	Frequency & duration of use	Number of times an identifier or its protected mechanism or authenticated data is used and how long it is used	one-time, multiple times, life time
2	Purpose of use	When the identifier is designed for a particular purpose, application or context and is switched for another purpose it can lead to privacy and security consequences	application specific, single sign-on, used with multiple services, context specific, silo
3	Provisioning	Identifiers constructed, updated or archived conjunction with a large set of personal attributes or sensitive personal data such as age, data of birth creates privacy risk	created, updated, deleted or archived with: limited personal attributes, overloaded personal attributes, sensitive personal attribute
4	Secrecy	The identifier can be inferred using the arithmetic or other properties of the identifiers	inferable, public, obfuscated, revocable, recoverable,
5	Claim Type	The nature of the authentication mechanisms used to secure the identifier	password, crypto key, biometric, challenge-response, single-claim, multiple-claims
6	Mobility	This refers to how easy to copy the identifier or its content, the physical constraints regarding the movement of the identifier	copyable, remotely usable, concurrently usable, immobile
7	Value at Risk	How much is loss when one or more of the following happens to an identifier. The loss could monetary or reputation	loss, misuse, disclosure, disruption, theft, replacement value
8	Assignment & Relationship	How the identifier is chosen or assigned. Is it by the individual, forced on the individual, assigned to a role or a pseudonym or jointly established.	forced, self, jointly-established, role, pseudonym
9	Obligation & Policy	Are there policy or obligation management mechanisms available in the IDM system?	policy absent, policy present

Task b (Round III) of the survey assesses how we can rate the risk factors in Table 7.4

regarding their effect on privacy and security risk in IDMSs. In addition, the task required the experts to rate the possibilities of quantifying the risk factors and to judge its complexity for non-technical stakeholders. Table 7.5 is the tally of the responses of Task b (Round III). The result in Table 7.5 indicates consensus in many of the ratings. The units M, L, H and N are explained in the Table 7.6. "Undecided" in the table means the respondents were unable to give their opinions on how to evaluate a particular factor.

Table 7.5 depicts the result of the evaluation of the consolidated results in Table 7.4. There is no consensus regarding the effect of "Frequency & duration of use" on privacy and security of IDMSs. 50% of the respondents rated the effect as medium while the others rated it as high. On the average the respondents agreed that "Frequency & duration of use" of a token have significant effect on privacy and security of IDMSs. Furthermore, the respondents agreed that the risk factors of this category are easy to measure or quantify. 50% of the respondents were of the opinion that the sensitivity of the factors is medium or provides accurate measure of privacy and security risk in IDMSs. This represents weak consensus as half of the respondents did not support this rating. Moreover, the majority of the respondents were of the opinion that non-technical system stakeholders can easily understand the factors in this category.

All the respondents agreed that the effect of "Claim type" on privacy and security in IDMSs is high. This represents a very high degree of consensus. Furthermore, the majority of the respondents were of the opinion that the factors of "Claim type" are very easy to measure or quantify and the factors provide accurate measure of privacy and security risk in IDMS. Additionally, the factors can easily be understood by non-technical stakeholders.

The factors for "Purpose of use" have medium effects on the privacy and security of IDMSs and they are difficult to measure. On the average the "Purpose of use" factors may provide accurate measure of privacy and security risk in IDMSs. However, the degree of consensus is low. The results also show that non-technical stakeholders can easily understand the factors.

The majority of the respondents were of the opinion that factors of "Provisioning" have medium effect on the privacy and security of IDMSs. This means "Provisioning" may result in costly loss of tangible or intangible assets or resources. Secondly, there is a low degree of consensus concerning the "measurement" rating. Half of the respondents agreed that the factors of "Provisioning" are easy to quantify. Also, the majority of the respondents agreed that the factors provide accurate measure of "Provisioning". Finally, there is a low degree of consensus on how non-technical stakeholders may understand the factors. Half of the respondents agreed that the factors can easily be understood by non-technical stakeholders.

50% of the experts rated the effect of "Secrecy" on privacy and security in IDMSs as high. This represents a low degree of consensus for the rating. The majority of the respondents were of the opinion that the "Secrecy" factors are easy to measure and they provide an accurate measure of privacy and security risk of tokens. Further, non-technical stakeholders can understand the factors.

The effect of "Value at risk" factors was rated as high by the majority of the respondents. However, there was a low degree of consensus on how difficult it is to quantify the factors. Half of the respondents agreed that it is difficult to measure or quantify the factors. Also, half of the respondents were of the opinion that the factors provide a very accurate measure of security and privacy. This represents a low degree of consensus. Similarly, half of the respondents agreed that the factors are easy to understand by non-technical stakeholders.

The majority of the respondents agreed that the effect of token "Assignment" factors on privacy and security is low. The respondents were split on how easy it is to quantify or measure the factors. Two of the respondents were of the opinion that the factors are easy to measure while others thought they are very easy to measure. The majority of the respondents were of the opinion that the factors provide less accurate measure of privacy and security risk. Finally, half of the respondents agreed that non-technical stakeholders can easily understand the risk factors.

The effect of the "Obligation & Policy" factors is medium. This means the factors have medium effect on privacy and security risk of IDMSs. However, the consensus was weak since only half of the respondents supported this evaluation. Similarly, half of the respondents were of the opinion that the factors are easy to measure or quantify and they provide accurate measure of privacy and security risk in IDMSs. The majority of the respondents were of the opinion that non-technical stakeholders cannot easily understand the factors.

Table 7.5: The Result for Task b of Round III

Rule No.	Category of Factors	Impact Level Tally	Measurement Level Tally	Sensitivity Level Tally	Stakeholder Level Tally
1	Frequency & duration of use	3M, 3H	4M, 2H	3M, 1L, 1H, undecided	4M,2H
2	Claim type	6H	3H, 2M , 1L	3M, 2H, undecided	3M, 2L, 1H
3	Purpose of use	5M, 1H	4L, 1H, 1N	2M, 2H, 1L, 1N	3M,1L,1H,1N
4	Provisioning	4M,2H	3M,1L, 1H, 1N	4M,1L, 1N	3M, 2H, 1N
5	Secrecy	3H, 2L, 1M	5M,1H	3M,1L,1H,1N	3M,1L,1H,1N
6	Mobility	3M, 2H, undecided	2L,2M, 1H, undecided	2M,2L,1H, undecided	3L,2M, undecided
7	Value at Risk	5H, 1L	3L,1M,1H,1N	3H,1L,1M,1N	3M,2H,1L
8	Assignment & Relationship	4L, 2H	2M,2H,1L,1N	4L,1M,1H	3M,1L,1H,1N
9	Obligation & Policy	3M, 1L,1H, undecided	3M,1L,1H, undecided	3M,1L,1H, undecided	4L,1H, undecided

Table 7.6 explains the meaning of the units used in rating the risk factors in Table 7.4

Table 7.6: Rating Unit for the Quantification

Scale	Impact Level	Measurement Level	Sensitivity Level	Stakeholder Level
Low (L)	Low impact	Difficult to measure	Factor(s) provides less accurate measure	Not easy to understand
Medium (M)	Medium impact	Easy to measure	Factor(s) provides accurate measure	Easy to understand
High (H)	High impact	Very easy to measure	Factor(s) provides very accurate measure	Very easy to understand
Not Significant (N)	Not significant impact	Impossible to measure	Factor(s) may provide Inaccurate measure	Rarely understandable

7.6 Discussion

Table 7.1 represents the result of the initial survey of risk contributing factors for IDMSs [117]. The risk contributing factors of the initial survey is similar to that of the consolidated table, Table 7.4. This indicates a fair degree of consensus concerning the evaluation of the initial construct [117].

However, some of the "Factors" have been refined in Table 7.4. The experts opinion regarding "Provisioning" is now more precise than the result obtained in the initial survey. The "Purpose of use" is refined and detailed in Table 7.4 than the initial version in Table 7.1. Furthermore, "Secrecy" and "Claim Type" have additional factors that are not identified in Table 7.1.

The second part of the evaluation requires the respondents to rate the risk contributing factors in Table 7.4. The objective of the rating is to determine whether a set of factors in a category is a key risk indicator (KRI). KRIs are highly relevant indicators possessing a high probability of predicting or indicating important risk [64]. The Risk IT [64] framework characterizes KRI as being sensitive, having high impact, easy to measure and relevant to both technical and non-technical stakeholders.

The rating table, Table 7.5, shows that the respondents rated "Claim Type" factors as having a high effect on privacy and security risk of IDMSs. This means the nature of the authentication mechanisms employ to secure an identifier may have a significant effect in assessing the privacy and security risk of IDMSs. Thus, their nature may result in the highly costly loss of major tangible or intangible assets or resources. It may significantly violate, harm the individual, or even cause death or serious injury to a stakeholder.

How to quantify or measure "Claim type" is skewed towards "High" – it may be very easy to quantify the "Claim type" factors. The factors also provide an accurate measure as sensitivity is skewed towards medium. Non-technical stakeholders may understand the factors as stakeholders understandability is skewed towards medium. We can conclude that "Claim type" is a KRI for privacy and security risk in IDMSs.

Similarly, "Frequency & duration of use" and "Secrecy" may be considered as KRIs for privacy and security risk in IDMSs.

"Purpose of use" is not a KRI as it has medium impact and is difficult to measure. Similarly, "Provisioning" is not a KRI because it may lead to medium privacy and security risk. "Mobility" is not a KRI because its privacy and security impact is skewed towards medium. Moreover, non-technical stakeholders may find it difficult to understand. Furthermore, "Value at Risk" may not be a KRI because it may be difficult to measure or quantify.

"Assignment & Relationship" may not be a KRI because it may lead to low privacy and security risk in IDMSs and may provide less accurate measure. In addition, "Obligation & Policy" may not be a KRI because it may have less significant effect on privacy and security of IDMSs and may be difficult for non-technical stakeholders to understand. The effect may result in insignificant loss of some tangible assets, resources or reputation. The three key risk indicators identified in the study are summarized in Table 7.7.

Table 7.7: The Key Risk Indicators Table

Category of Factors	**Factors**
Frequency & duration of use	one-time, multiple times, life time
Secrecy	inferable, public, obfuscated, revocable, recoverable
Claim Type	password, crypto key, biometric, challenge-response, single-claim, multiple-claims

7.7 Shortcomings

One of the shortfalls observed in this study is the limited number of rounds. The number of rounds for the evaluation of the KRIs should have been more than one. Four rounds could have given some of the respondents an opportunity to rethink their answers. For example, a fourth round could have assisted in reassessing the privacy and security risk of the "Frequency & duration of use" factors since the opinions were divided on certain aspects of the assessment. The number of experts is equally important. Though we have the appropriate number of respondents, we believe that an additional 4 to 20 experts could have enriched the results. We were able to complete the study with six experts instead of nine. However, the effect of the shortfall can be felt on the final round of the study but not the entire study.

7.8 Mapping the Construct with the New Protection Goals

Evaluation of a construct includes completeness among other criteria [90]. We show the completeness of our result by aligning the risk factor with the new protection goals (NPGs). The NPGs turn data protection into a modern, proactive and operational tool. They "fall in line with the approved methods of risk analysis and protective measures such as baseline protection" [131]. They provide technically convertible principles that cover both asset and right protection which is lacking in the privacy by design goals.

The NPGs extend the classical security goals - integrity, availability and confidentiality, by including transparency, unlinkability, and ability to intervene (Intervenability). Transparency relates to the purpose of data processing, necessity, data thriftiness and information needs of the data subjects. Unlinkability operationalizes purpose separation or bindingness; that is, it checks if personal data collected for a particular purpose is being used for another purpose. The ability to intervene (intervenability) gives the data subjects control over the processing of their personal data [131].

The second column of Table 7.8 represents the categories of risk factors while the third represents the risk factors. The first column of Table 7.8 aligns the NPGs [131] with the risk factors. The alignment does not mean the exact match, but rather correspondence, similarity or overlap of concepts [93]. The alignment shows the completeness of the model because each NPG has corresponding risk categories and factors.

New Protection Goals	**Category of Factors**	**Factors**
Confidentiality, Unlinkability	Frequency & duration of use	one-time, multiple times, life time
Integrity, Confidentiality, Intervenability	Provisioning	created, updated, deleted or archived attribute with: limited personal data, overloaded personal data, sensitive personal data
Unlinkability	Purpose of use	application specific, single sign-on, multiple services, context specific, silo
Intervenability, Unlinkability	Assignment & Relationship	forced, self, jointly-established, role, pseudonym
Confidentiality	Secrecy	inferable, public, obfuscated, revocable, recoverable,
Confidentiality	Claim Type	password, crypto key, biometric, challenge-response, single-claim, multiple-claims
Availability, Confidentiality	Mobility	copyable, remotely usable, concurrently usable, immobile
Availability	Value at Risk	loss, misuse, disclosure, disruption, theft, replacement value
Transparency	Obligation & Policy	policy absence, policies present

Table 7.8: Mapping of the Risk Factors and the New Protection Goals

7.9 Conclusion

In this paper, we used the Delphi method to evaluate a construct for privacy and security risk assessment. The study asked an expert panel to list factors that contribute to privacy and security risk of uses of identifiers in identity management systems (IDMSs). The responses obtained from the study were consolidated with a previous model [117] of risk contributing factors for expert evaluation. The study shows that the experts mostly agreed with the initial risk analysis construct for IDMSs. In addition, the study evaluated the consolidated result for key risk indicators (KRIs). The respondents identified "Claim type", "Frequency & duration of use" and "Secrecy" as possible KRIs. Furthermore, the completeness of the result was evaluated by aligning the factors to the new protection goals. The alignment evaluated the completeness of the result obtained in the Delphi study using an established set of privacy and security requirement. The evaluation showed that the result is fairly complete since each category aligns with a security or privacy requirement. The model can be used as a model for privacy and security risk assessment in IDMSs especially the key risk indicators. Future work will show how the factors can be employed to estimate risk in IDMSs.

Appendix

The questions for the three-round Delphi study are as follows:

- Round I
  - We would like you to help us find the various ways by which the properties or the characteristics of electronic identifiers contribute to privacy and security risk in IDMSs. Please list and explain the risk properties of electronic identifiers in your own words. Focus on the electronic identifiers and IDMSs only - the inclusion of application-specific risk are out of the scope of this survey.
- Round II
  - We consolidated the responses received in Round I of this Delphi study into categories of risk contributing properties of electronic identifiers. We found out that some factors require further deliberation. We would like you to share your opinions regarding the factors in the Table 7.2 by answering the questions below. You are free to expand or adjust the table if needed.
    1. Would you consider single sign-on and multiple services risk factors as related to frequency & duration of use (e.g. as an additional dimension in "application space", or rather see them classified into "purpose of use"? (Please see row 1 & 2 in Table 7.2)
    2. Are the factors "application specific", "privacy preserving", "silo" and "context specific" in "purpose of use", in your opinion, point to the same overall concept - or shall they be listed separately? (Please see row 2 in Table 7.2)
    3. Would you agree with how we explain the provisioning of electronic identifier in Table 7.2 row 3? Do you agree with the listed factors that contribute to risk when provisioning electronic identifiers? (Please see row 3 in Table 7.2)
- Round III
  - Table 7.4 consists of the consolidated risk factors of identifiers. We would like you to comment on these risk factors, especially the new additions. The new additions are the rows in bold. They are not direct opinions but inferences we drew from the study and other research works. We encourage any kind of remark (objection, clarification, or acceptance, etc.).
  - This task asks you to rate the risk factors in Table 7.4 concerning their potential impact on privacy and security of IDMSs. In addition, we ask you to rate the possibilities to measure the risk factors, and to judge its complexity for non-technical users. You should fill in Table 7.5 based on the options in Table 7.6.

Chapter 8

Article IV
A Model for Privacy and Security Risks Analysis,

in IEEE Proceedings of 2012 5th International Conference on New Technologies, Mobility and Security (NTMS)

Ebenezer Paintsil

Abstract

This article introduces an extended misuse case (EMC) model for privacy and security risk analysis and formally validates the model by means of colored petri nets (CPNs). The EMC model extends the use and misuse cases (UMCs) model with security and privacy requirements. The proposed EMC model and the CPNs instantiation deal with some of the shortcomings of the traditional UMCs which include lack of quality goals and formal validation techniques. The CPNs instantiation enables automatic detection of possible violation of privacy and security goals and can be extended to communicate risk to both technical and non-technical stakeholders. The CPNs and EMC models are illustrated with privacy and security risk contributing factors for identity management systems (IDMSs).

8.1 Introduction

Security and privacy risk analysis of information systems (ISs) is a multidisciplinary task involving both technical and non-technical stakeholders [64]. Effective communication and risk analysis require a common understanding of threats, vulnerabilities and risks among the different groups of stakeholders participating in the analysis [54].

One of the ways of enhancing IS risk analysis and communication is by means of model-based risk analysis. Unlike the traditional risk analysis methods that document risk in tables and text (making it difficult to comprehend the risk findings), model-based risk analysis methods use system models to facilitate communication and interaction among all stakeholders [58].

Applying a model-based risk analysis method requires a careful consideration of its user-friendliness, learning curve and how well the method automates certain aspects of risk analysis.

Use and misuse cases (UMCs) modeling approach [138] is one of the simple and user-friendly risk analysis approach. It makes human judgments more informed and systematic [5], [38]. UMCs risk modeling appeals to the industry because of the substantial connection of use cases to the existing system development process [106].

However, UMCs models lack quality goals, their operationalization and effects [38]. In addition, UMCs models are informal and therefore cannot automatically detect security and privacy violations.

This article introduces an extended misuse case (EMC) model for both privacy and security risk analysis and formally validates the model by means of colored petri nets (CPNs). The EMC model extends the UMCs model with security and privacy requirements or quality goals such as availability and unlinkability. The proposed EMC model and the CPNs

instantiation deal with some of the shortcomings of UMCs models. The CPNs instantiation enables automatic detection of possible violation of privacy and security goals and can be extended with additional concepts to visually communicate risk to non-technical stakeholders [67].

This article is organized as follows: section 8.2 discusses the related work and section 8.3 presents the proposed EMC model. Section 8.4 illustrates the EMC model with privacy and security risk contributing factors for identity management systems (IDMSs). Section 8.5 instantiates the EMC model with colored petri nets (CPNs) in order to show how privacy and security violations can be detected automatically. Section 8.6 validates the CPNs formal model for correctness, section 8.7 explains the applications of the model and section 8.8 states the conclusion and future work.

8.2 Related Work

The EMC model proposed in this article (see in Figure 8.1) relates to some notable security threat modeling proposals such as i* [82], goal-oriented modeling [38], and Goal-Directed Requirements Acquisition (KAOS) [160]. The goal oriented models provide alternative approaches to system requirement elicitation and analysis. For example, the KAOS method models goals with a precise version of classical AND-OR trees. However, these approaches have complex semantics and learning cost, and lack substantial application in the existing system development process [106]. Use cases have easy to learn semantics and are widely used methodology for system requirement analysis.

The model we propose in Figure 8.1 is similar to the EMC model proposed by Okubo et al. [106]. Their proposal extends the misuse case model [5], [138] with assets and goals. Assets are resources with potentially great value to a system. Asset protection is the main reason for vulnerability or risk analysis. Security goals represent the reasons why the asset must be protected from threats. However, their work is asset driven or security focused and provides no formal technique for validating the model. A misuse case threatens only assets but not privacy right. This work focuses on modeling security and privacy risk which requires both asset and right protection [131], [158] and formalizes the model by means of CPNs.

Bandara et. al. [14] reviewed security modeling approaches and evaluated their capabilities to representing security pattern. Security pattern defines five main characteristics – problem, context, force, solution and consequences. The problem characteristic determines how well a security model captures security goals. The context determines the influence of security environment on the achievement of security goals. It captures security risk of the system environment. Thus, vulnerabilities, assets, possible loss, attackers and their capabilities. The "force" characteristic captures the rationale for selecting a particular security solution from several solutions. "Solution" is the description of the mechanic whose execution can mitigate the security problem. "Consequence" describes the changes that can happen when a security solution is applied [14].

The models analyzed by Bandara et. al. [14] include UMLSec [71], SecureUML [83] and UMCs. Among the three, UMCs model is best at capturing and analyzing security problem [14] even though they do not explicitly specify security problems [93]. Explicit security or quality goal specification in UMCs models would make the modeling approach fully problem oriented. The EMC introduced in this work extends UMCs models with quality or security goals, asset and right protection. The introduction of asset into the UMCs models has the potential of making the EMC model context oriented.

Moreover, UMLSec [71] and SecureUML [83] are security focused and therefore do not include right protection. This work focuses on modeling security and privacy risk which requires both asset and right protection [131], [158] using user friendly EMC model [5], [106]. In addition, UMLSec modeling approach is not user-friendly and targets only tech-

nical stakeholders [72]. This work formalizes a user-friendly EMC model with tools that can communicate risk to both technical and non-technical stakeholders.

UMCs model has been aligned with information system security risk management (ISSRM) reference model [93]. The reference model addresses ISSRM by combing asset, risk and risk treatment. Using the ISSRM reference model, R. Matulevicius et. al. [93] identified that UMCs models lack security controls, risk treatment, information system asset, cause of risk and vulnerability constructs [93]. However, many of these shortcomings may not arise if we use UMCs models for risk analysis instead of risk management. The EMC model proposed in this work focuses on using extended UMCs model for risk analysis.

8.3 Extended Misuse Case Model

This section explains the proposed EMC model for privacy and security risk analysis depicted in Figure 8.1. The EMC model incorporates system goal(s), asset and right into the misuse case model (see Figure 8.1). The asset is what the system is trying to protect or something that may advance a system goal and is represented by its name and the stereotype $<< asset >>$. A use case may use an asset as a data object or a resource to accomplish a task [106]. This is represented by the dotted line and the stereotype $<< use >>$. The second extension is the "right". It is represented by its name and the stereotype $<< right >>$. "Right" incorporates privacy protection measures into the model. While security modeling focuses on asset protection, privacy on the other hand focuses on both security and right protection [131], [158]. The stereotype $<< right >>$ distinguishes the model from security models which are asset driven. The privacy right refers to norms such as obligations, exceptions and prohibition [111].

A threat is denoted by a misuse case which may threaten a right or an asset [138]. In order to simplify the model, we do not make explicit distinction between vulnerabilities and threats. Both are assumed to be misuse cases. A use case may mitigate [138] a threat or a misuse case and may help operationalize a system goal. Goal operationalization is represented by the stereotype $<< operationalize >>$. A system goal is operationalized if all its known misuse cases are mitigated with appropriate countermeasures. A goal represents the reasons why we need to protect an asset or a right from threats. A goal is represented by a name and the stereotype $<< goal >>$. It may consist of sub-goals. If a goal is associated with an asset or right it means that the goal is intended to protect the asset or the right from threats [106].

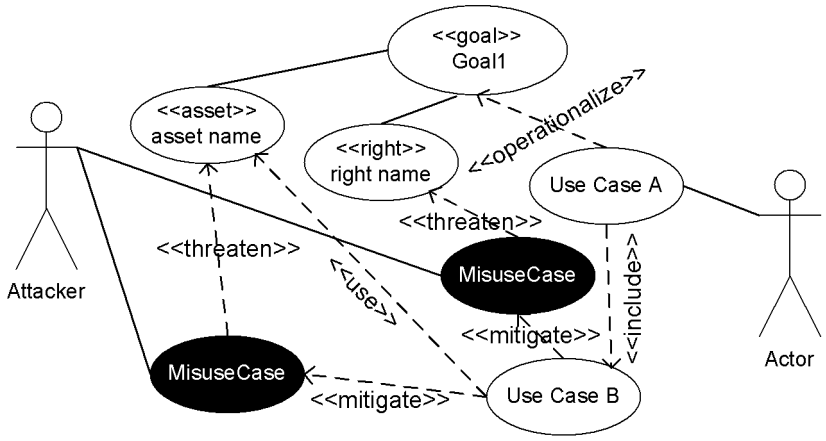

Figure 8.1: The Extended Misuse Case Model

8.4 Application of the EMC Modeling Approach

This section illustrates the EMC model with privacy and security risk contributing factors for IDMSs. IDMSs create and manage identities of end-users. The traditional role of IDMSs is to issue end-users with credentials after an account is created. The end-users can then use the credentials to authenticate and request access to services and resources [69]. The security and privacy of IDMSs depend on the characteristics of credentials or tokens used in the system. Our previous work [117] surveys how the characteristics of identifiers or tokens contribute to privacy and security risk in IDMSs. We conducted a Delphi study to evaluate the initial survey [115] which produced the result in the second and third columns of Table 8.1.

New Protection Goals	Category of Factors	Factors
Confidentiality, Unlinkability	Frequency & duration of use	one-time, multiple times, life time
Integrity, Confidentiality, Intervenability	Provisioning	created, updated, deleted or archived attribute with: limited personal data, overloaded personal data, sensitive personal data
Unlinkability	Purpose of use	application specific, single sign-on, multiple services, context specific, silo
Intervenability, Unlinkability	Assignment & Relationship	forced, self, jointly-established, role, pseudonym
Confidentiality	Secrecy	inferable, public, obfuscated, revocable, recoverable,
Confidentiality	Claim Type	password, crypto key, biometric, challenge-response, single-claim, multiple-claims
Availability, Confidentiality	Mobility	copyable, remotely usable, concurrently usable, immobile
Availability	Value at Risk	loss, misuse, disclosure, disruption, theft, replacement value
Transparency	Obligation & Policy	policy absence, policies present

Table 8.1: Mapping of the Risk Factors and the New Protection Goals

The third and the second columns of Table 8.1 show the risk contributing factors and their categories respectively. The first column of Table 8.1 maps the risk factors with the new protection goals (NPGs) [131]. The mapping aligns the risk contributing factors and their categories to an established privacy and security concept (NPGs). The NPGs turn data protection into a modern, proactive and operational tool. They "fall in line with the approved methods of risk analysis and protective measures such as baseline protection" [131]. They provide technically convertible principles that cover both asset and right protection which is lacking in the privacy by design goals.

The NPGs extend the classical security goals - integrity, availability and confidentiality, by including transparency, unlinkability and ability to intervene (Intervenability). Transparency relates to the purpose of data processing, necessity, data thriftiness and information needs of the data subjects. Unlinkability operationalizes purpose separation or bindingness. That is, it checks if personal data collected for a particular purpose is being used for another purpose. The ability to intervene (intervenability) gives the data subjects control over the processing of their personal data [131].

We apply the EMC technique to model the risk factors in Table 8.1. The categories that enforce privacy are rights while that of security are assets. We model the factors that protect the system goal(s) as use cases. The factors that threaten the system goal(s) are

misuse cases. For brevity, Actors and attackers are omitted in the EMC model. The factors in Table 8.1 are modeled and analyzed as follows:

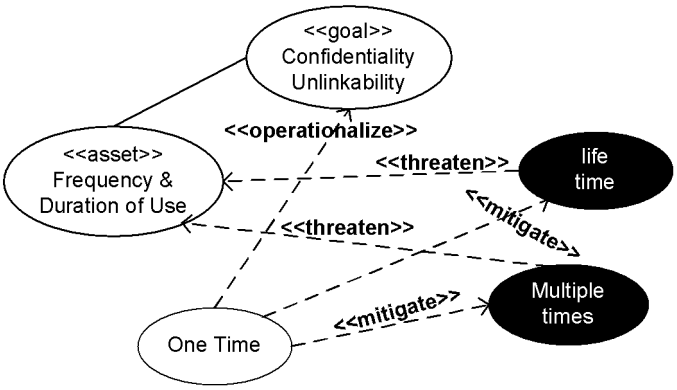

Figure 8.2: EMC Model for Frequency & Duration of Use

Frequency & duration of use: Figure 8.2 depicts the EMC model for the "Frequency & duration of use" risk category. This category relates to the 'Confidentiality' and 'Unlinkability' NPGs. Confidentiality requires that confidential information is shared only among authorized persons or organizations [100]. In other words, an attacker should not be able to deduce anything about legitimate users' activity [135]. An attacker can deduce the behavior of end-users from the profile of their activities. Any attack or unauthorized access of the end-users' profile can lead to breach of confidentiality. The profile of end-users' activities can be generated by a repeated use of the same identifier in an IDMS. In addition, the repeated use of the same token can enable an attacker to link tokens. Therefore, multiple time or life-time tokens can lead to profiling of end-users' activities which in turn may affect confidentiality and linkability. On the other hand, one-time tokens are valid for a login session or a transaction period because of that they can prevent successful profiling of end-users' activities. Any unauthorized intrusion into the IDMS may not reveal substantial information about the activities of the token's users. Hence, one-time tokens are use cases but multiple and life-time tokens are misuse cases.

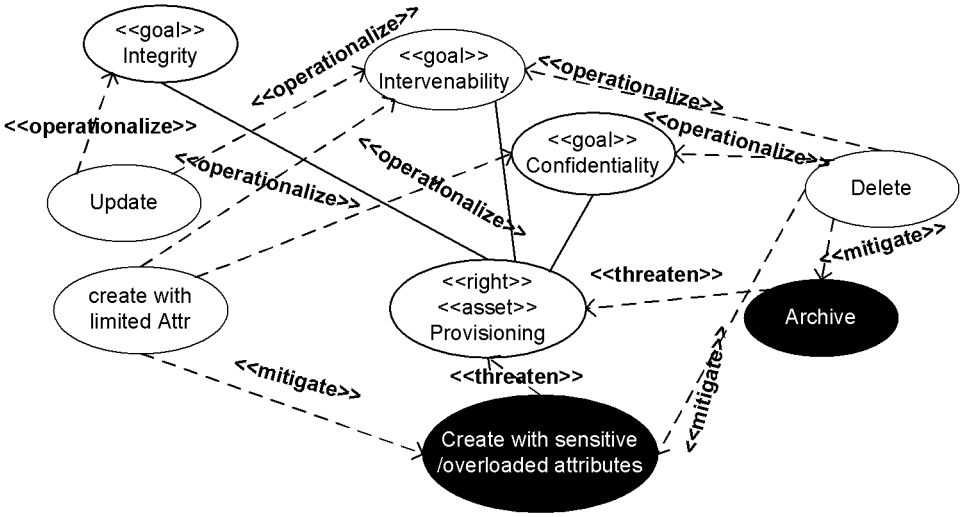

Figure 8.3: EMC Model for Provisioning

Provisioning: Figure 8.3 depicts the EMC model for the "Provisioning" risk category. It relates to the 'Integrity', 'Confidentiality' and 'Intervenability' NPGs. Integrity ensures that the information is authentic, accurate and complete [100]. The 'Integrity' goal is achieved if a token keeps authentic and up-to-date information. This may imply that the token should always store or carry correct information. The ability to update a token allows outdated information to be replaced with authentic and accurate information. We assume that the actor or end-user who updates the token is not malicious. We can mitigate data integrity risk by destroying or deleting the token. Deletion here refers to total destruction of the token and all its associated data and archival information. Such token cannot be subjected to any threat. The deletion of token operationalizes the confidentiality goal while the update operationalizes the integrity goal.

Furthermore, deletion, creation and updating may offer the token owner the ability to intervene if something goes wrong with the token. The confidentiality of a token may depend on the kind of attributes used in its creation. A token created with few but sufficient or less sensitive attribute may enhance confidentiality. Data minimality principle prohibits excessive collection of personal data (see [89], [21]). When we archive or create a token with sensitive or more attributes than necessary ("overloaded attributes") we may risk confidentiality because much personal data can be revealed when a misuse occurs. Therefore, we consider a token created with "limited" or few but sufficient attributes, the ability to delete or destroy a token and the ability to update a token as use cases of the provisioning category.

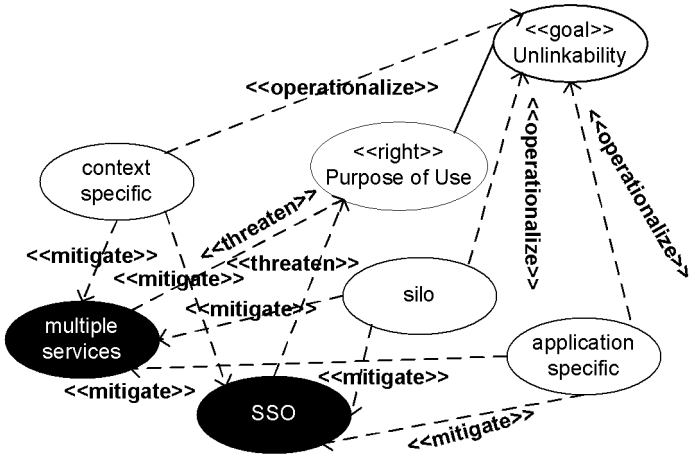

Figure 8.4: EMCs Model for Purpose of Use

Purpose of use: Figure 8.4 depicts the EMC model for the "Purpose of use" risk category. This category relates to the unlinkability NPG. Unlinkability means purpose separation or bindingness [131]. We can minimize privacy risk of a token if it is created and used for a specific purpose or used within a specific context [49],[21]. A token used for multiple purposes or services may be subjected to abuse or illegal processing. IDMSs that support single sign-on (SSO) allow the use of tokens for multiple services and multiple domains upon a single authentication [169]. Although SSO reduces human error, it leads to sharing of valuable information across domains using loosely coupled network protocols [89]. It requires all parties to secure their communication channels against replay attacks, man-in-the-middle attacks, session hijacking, and other attacks that can allow malicious use of identity information or resources. Hence, in relation to "Purpose of use", multiple services and SSO factors are misuse cases. The remaining factors are the use cases.

Secrecy: Figure 8.5 depicts the EMC model for the "Secrecy" risk category. This category relates to the confidentiality NPG. If a token's secret can be inferred, guessed or found

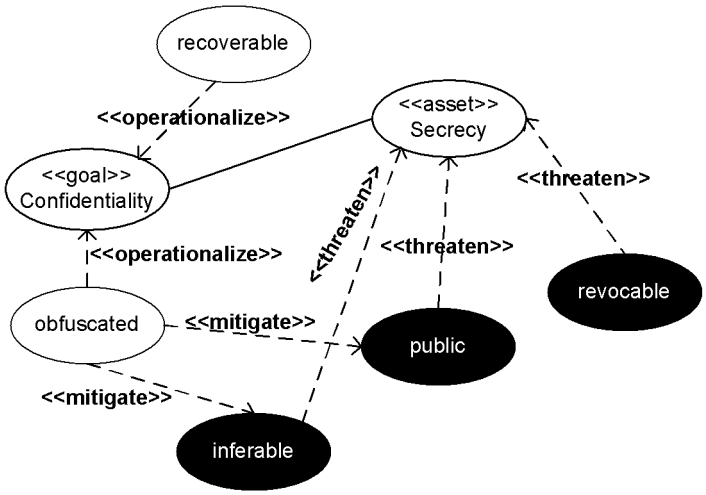

Figure 8.5: EMC Model for Secrecy

in a public domain (public database) or revoked then such token lacks secrecy. When a token's secret is revoked, the end-user of the token could be identified or confidential information may be made available to unauthorized persons. In addition, the behavior of the user could be revealed to an unauthorized persons or entities as a result of the revocation. Hence, inferable, public and revocable tokens are misuse cases. The remaining factors are the use cases. The model in Figure 8.5 reveals that the 'revocable' misuse case has no countermeasure. Therefore, a countermeasure is needed in order to operationalize the goal. The EMC modeling is able to reveal the missing countermeasures in Table 8.1.

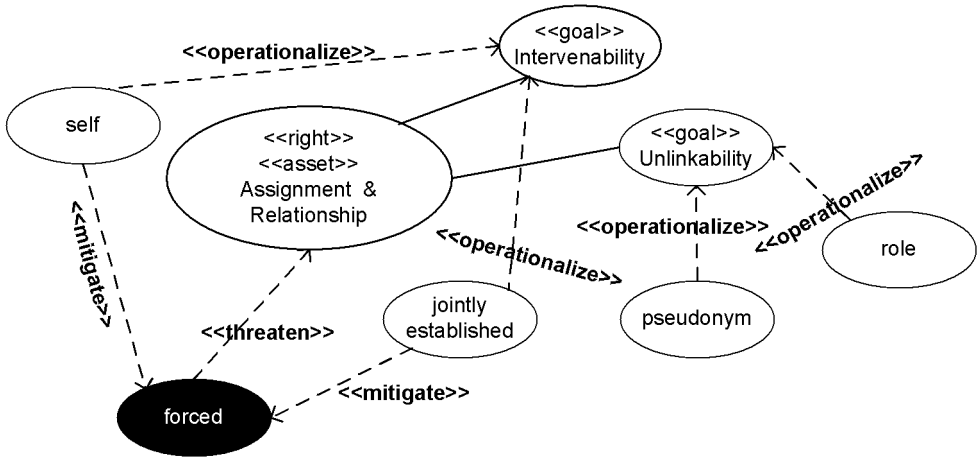

Figure 8.6: EMC Model for Assignment & Relationship

Assignment & Relationship: Figure 8.6 depicts the EMC model for the "Assignment & Relationship" risk category. It relates to the "intervenability" NPG. We can enhance privacy if end-user can intervene in token creation or assignment. When a token is forced on an end-user by a higher authority or an identity provider (IdP), the token's end-user would be deprived of the right to intervene or the right to be engaged in the token creation. In addition, the IdP may not understand the privacy needs of the data subject.

We may mitigate privacy risk when a token is assigned to a role instead of an individual. In this case, the token is created for generic use usually based on functions within

an organization. The usages of such tokens may be unlinkable to a particular individual thereby minimizing the privacy risk of the token [121], [21]. Similarly, a pseudonymous token may enhance privacy [23].

Furthermore, IdPs can involve the end-users in assigning the token or they can allow them to choose the attributes they wish to reveal. Therefore, 'forced' token factor is a misuse case while others are use cases.

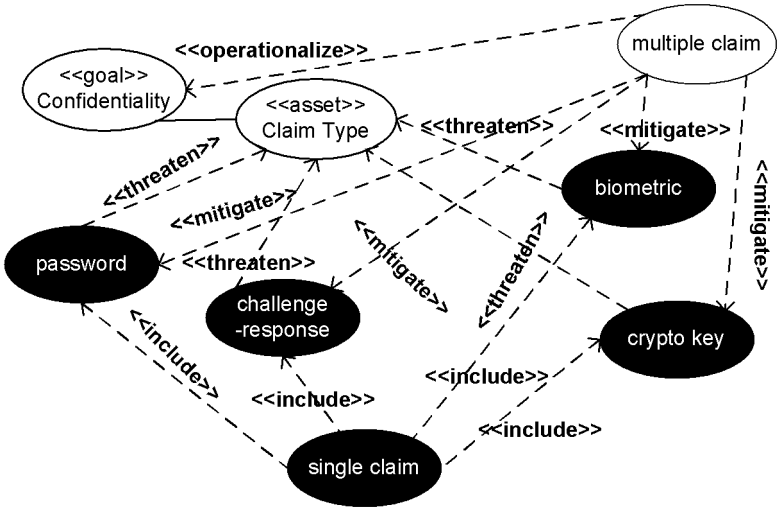

Figure 8.7: EMC Model for Claim Type

Claim Type: Figure 8.7 depicts the EMC model for the "Claim type" risk category. It relates to the confidentiality NPG. The main factors are single and multiple claim types. We have 'password', 'challenge-response', 'biometric' and 'crypto key' included in single claim type. We assume that any combination of these factors may enhance confidentiality or prevent an unauthorized access to resources and services. Therefore, multiple claim type is a use case while the single claims are misuse cases.

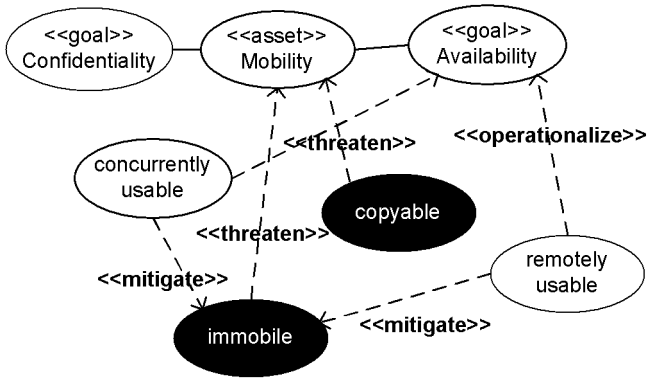

Figure 8.8: EMC Model for Mobility

Mobility: Figure 8.8 depicts the EMC model for the "Mobility" risk category. It relates to the availability and confidentiality goals. A token should be available when needed. Immobile token may affect the availability of tokens because of the physical constraints. For example, a bank calculator token may require the physical device that generates the random numbers. When the physical device is not available the banking services cannot be accessed. Therefore, the immobile constraint may affect availability of tokens. It follows

that, 'immobile' token is a misuse case. The remaining factors except "copyable" are use cases because they may support availability of tokens. "Copyable" refers to how easy it is to copy a token. For example, the content of a low cost RFID tag with no additional security could easily be read by anyone but a high cost RFID tag that comes with additional security may ensure that only authorized readers have access to its content. Hence, "Copyable" factor threatens "Mobility" which in turn affects the confidentiality goal.

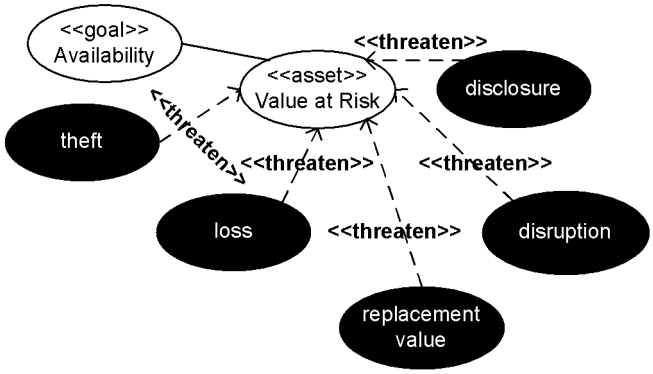

Figure 8.9: EMC Model for Value at Risk

Value at Risk: Figure 8.9 depicts the EMC model for the "Value at Risk" risk category. It relates to the availability NPG. When a token is lost, misused, disclosed, disrupted, stolen, it could prevent the availability of the token. When a token has a high replacement value, the security cost associated with its availability may be high for the intended user. Therefore, all the factors in this category are misuse cases because they prevent availability of the token. The analysis reveals that there are no countermeasures for the misuse cases. We need countermeasures in order to operationalize the availability goal. Therefore, the availability goal is not operationalize in this EMC model.

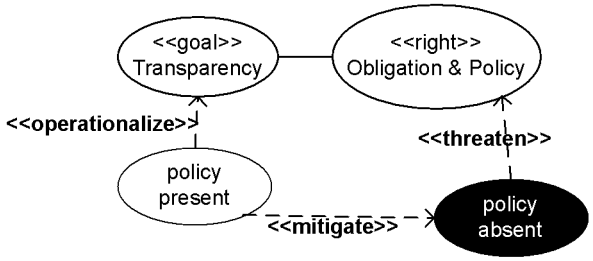

Figure 8.10: EMC Model for Obligation & Policy Risk Contributing Factors

Obligation & Policy: Figure 8.10 depicts the EMC model for the "Obligation & Policy" risk category. It relates to the transparency goal. Transparency includes the ability to plan, check and evaluate data collection and protection operations [131]. When a good policy is available, it may help plan, reproduce, check and evaluate data collection and processing operations. Absence of such policy may contribute to misuse of a token. Therefore, presence of a good policy is a use case while the absent of a policy is a misuse case.

8.5 Modeling of EMC with Colored Petri Nets

This section implements the EMC model with CPNs. CPNs provide an excellent technique for modeling workflow and concurrent systems [77]. They provide tools for verification,

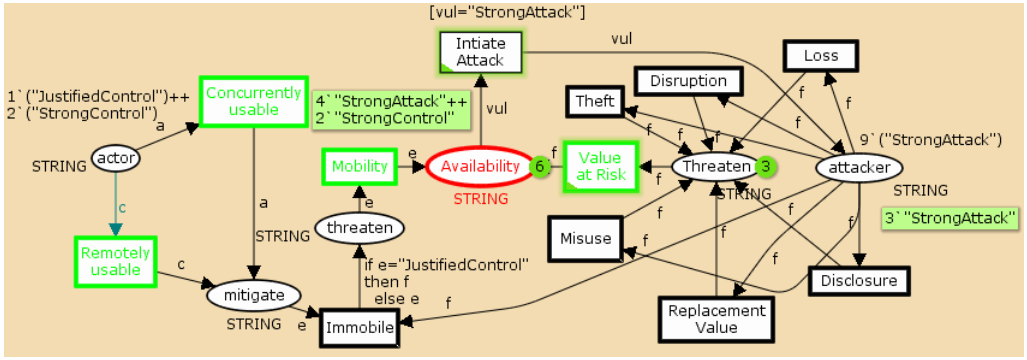

Figure 8.11: Extended Misuse Case for Value at Risk category

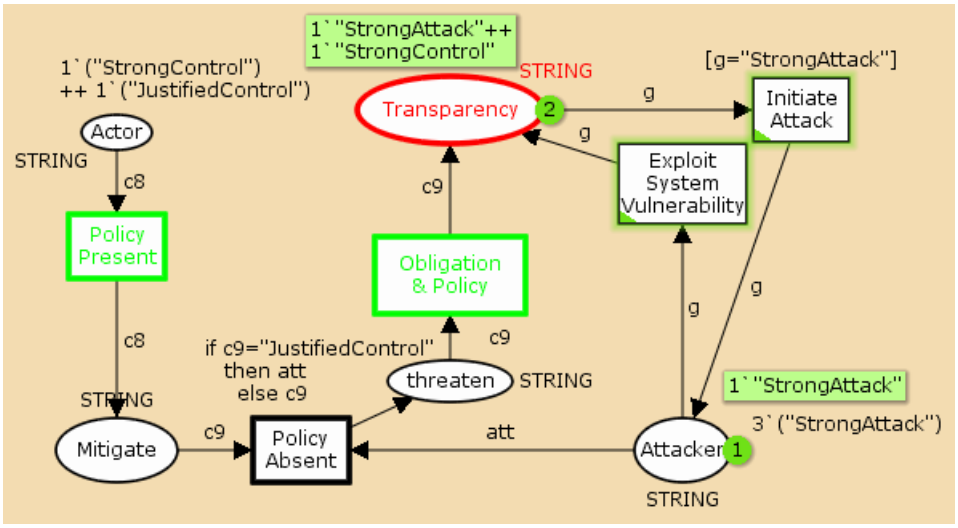

Figure 8.12: Extended Misuse Case for Obligation & Policy risk category

validation and automatic analysis of the model. A CPNs' model consists of **places, transitions (events), input and output** arcs. Places are represented by ellipses, transitions by rectangles, input/output arcs by directed arcs [77]. An arc always connects a place to a transition and vice-versa. A place represents a system state and may hold a collection of tokens.

A token in CPNs is a variable with data type and a value. The data type is known as color set and the data values are the colors. The transitions represent the events or actions that can cause a system to change its state [77]. Each place can be marked with one or more tokens, and each token has a data value or token color attached to it. The number of tokens and the token colors on the individual places represent the state of the system. The tokens on a specific place constitute the marking of that place [77]. The marking of a CPNs' model at the beginning of an analysis is called initial marking. The places where initial markings are located are known as initial places.

The CPNs model in Figure 8.11 represent the CPNs implementation of the "Value at Risk" and the "Mobility" EMC models in Figure 8.9 and Figure 8.8. Figure 8.12 represents the "Obligation & Policy" EMC model in Figure 8.10. The EMC "nets" in Figure 8.11 and Figure 8.12 use the conjunctive attack net model proposed by McDermott [95] and Zhou [176]. Use case and Misuse cases conjugate to determine how a system goal is satisfied.

We map the CPNs' specification to that of the EMC model. The initial markings "Justi-

fiedControl" and "StrongControl" correspond to the capabilities or nature of countermeasure available for the system actor. On the other hand, the initial marking "StrongAttack" corresponds to the capabilities or nature of the attack fired by an attacker. The initial token colors describe the nature of the countermeasure or the attack. For example, in Figure 8.11 and Figure 8.12 the actors have two types of countermeasures - strong and justified controls (countermeasures), representing the capability of the actors. Justified controls are chosen for economic reasons [62]. The attackers in Figure 8.11 and Figure 8.12 use strong attack to attack the system. Therefore, the token color is "StrongAttack". We assume that a strong countermeasure always mitigates strong attack. However, strong attack always threatens justified countermeasure. The attacker is the agent that initiates the misuse case while the actor initiates the use cases.

The token movement corresponds to the progress of an attack or a counter measure [95]. If an attacker's token reaches a goal state (place) then the countermeasures was ineffective or the actor has no mitigation or countermeasures for the misuse cases.

Use case and misuse cases correspond to transitions. We model the Use and Misuse cases as atomic events or actions. The set of pre-conditions and post-conditions correspond to either an arc expression or a guard. We use places to represent security and privacy goals. Similarly, we use places to represent relationships such as "include" and "mitigate".

Every system has vulnerabilities and attackers will always try to exploit them. We introduce additional transitions to show how an adversary would exploit vulnerabilities in a system when all the threats are mitigated. For example, we introduced the transitions "Exploit System Vulnerability" in Figure 8.12 to show how an adversary would exploit unknown vulnerability in the system. In Figure 8.11, the known vulnerabilities are the unmitigated misuse cases. They correspond to the transitions "Loss", "Misuse", "Disclosure", "Disruption", "Theft" and "Replacement value". The transition "Initiate Attack" is to enable or illustrate the continual attack by the adversary through the unmitigated vulnerabilities.

We color the use case green and misuse cases black in both Figure 8.11 and Figure 8.12. An actor uses a use case to mitigate the misuse cases. The attacker uses misuse case to threaten the use cases. The color set or data type for each place is the models in Figure 8.11 and Figure 8.12, is STRING. Also, all the variables are of the type STRING.

8.6 Validation of the EMC Formal Model

This section validates the behavioral correctness of the EMC CPNs model in Figure 8.11 and Figure 8.12. CPNs have two types of properties – behavioral and structural properties. While the behavior properties depend on the initial marking the structural properties are independent of the initial markings [29]. Behavior correctness criteria include soundness, boundedness, and fairness. The soundness property determines proper termination of CPNs' execution and that no anomaly such as deadlock or livelock occurs in the model [157]. We can validate a behavioral property of a CPNs model with state space analysis [76]. "The basic idea behind a state space is to construct a directed graph which has a node for each reachable marking and an arc for each occurring binding element" [76]. State space analysis automatically evaluates the basic behavioral properties – reachability, boundedness, home, liveness and fairness [79] and are explained as follows:

- Reachable property determines whether we have a path between two given markings M_1 and M_2 of a CPNs model [76].

- Boundedness property determines the maximal and minimal number of tokens which may be located on the individual places in the reachable markings [76]. Boundedness property determines the safety of a model. For example, a control system is safe if it is 1-bound. A buffer may be safe if it is k-bound i.e. no buffer overflows [175].

- Fairness property determines whether a process can starve another process in a system by denying it the resources needed to function [145]. It also determines how often the individual transitions take place [76].
- Home property determines the markings which can be reached from any reachable marking [76].
- Liveness property determines if a system is deadlock free or we have a set of binding elements who are always active [98], [175].

We perform two separate full state space analysis, one for each of the models in Figure 8.11 and Figure 8.12 using CPNTools [76]. The report for Figure 8.11 is as follows:

- Boundedness Properties - Best Integer Bounds
  - Availability'Availability 1 9 0
  - Availability'actor 1 3 0
  - Availability'attacker 1 9 0
  - ....
  - Availability'threatens 1 9 0
- Liveness Properties
  - Dead Markings - None
  - Dead Transition Instances - None
  - Live Transition Instances - Availability'Disclosure 1, Availability'Disruption 1, Availability'Intiate_Attack 1, Availability'Loss 1, Availability'Misuse 1, Availability'Replacement_Value 1, Availability'Theft 1, Availability'Value_at_Risk 1
- Fairness Properties
  - Impartial Transition Instances - Availability'Intiate_Attack 1, Availability'Value_at_Risk 1
  - Fair Transition Instances - None
  - Just Transition Instances - None
  - Transition Instances with No Fairness - Availability'Concurrently_usable 1, Availability'Copyable 1, Availability'Disclosure 1, Availability'Disruption 1, Availability'Immobile 1, Availability'Loss 1, Availability'Misuse 1, Availability'Mobility 1, Availability'Remotely_usable 1, Availability'Replacement_Value 1, Availability'Theft 1

The two models in Figure 8.11 and Figure 8.12 have similar behavior properties so we show the report for only Figure 8.11, the "Availability" model. The report focuses on the main behavioral properties relevant for the evaluation of the models. We analyze the reports as follows:

1. Boundedness Property: The best integer upper bound for Figure 8.11 is nine and that of Figure 8.12 is three. The full state space analysis performed on the two models means both models have no state space explosion problem. Therefore, they could be regarded as safe since their places represent message places [175].

2. Liveness property: Both models have no dead marking and transitions under the liveness property. The dead transition property determines the transitions that are never been enabled. If initially dead transitions exist, then the model may be badly designed [175]. The dead markings property determines which markings have no enabled binding element. If the number of dead markings reported by state space analysis tool is more than expected, then there must be errors in the design [175]. No dead marking and dead transition in both models indicate a good design and a possible deadlock free model. In addition, both models have live transitions indicating that an attacker would always exploit the vulnerabilities of a system. For example, in Figure 8.11 "Disclosure" and "Disruption" are live transitions because they have not been mitigated. An attacker will always exploit such unmitigated vulnerabilities.

3. Fairness: Both models have some transition instances with no fairness. This is expected in a misuse case model as both use and misuse cases try to compute for resources. They do so by starving the other transitions of resources. The No Fairness transitions show the potential conflicts that occur among the use and misuse cases as they try to accomplish their goal.

4. Reachability: We use this property to determine the soundness of the threat model. A threat model is sound with respect to a certain behavior only if the threat can take place (i.e. no deadlock or livelock)[171]. A threat will always materialize or reach a system goal in both models if the countermeasures are not strong. In addition, both models have live transitions which shows that an attacker can reach a goal through unmitigated or unknown vulnerabilities.

We conclude from the analysis above that the models in Figure 8.11 and Figure 8.12 may be behaviorally correct or corresponds to the static model.

8.7 Application of the EMC Model

The EMC model inherits the simple and easy to use properties of use cases [5], [38]. The CPNs implementation of the model can be used in risk analysis to effectively communicate risk to both technical and non-technical stakeholders. For example, the simulation results for the CPNs models in Figure 8.11 and Figure 8.12 show that the strong controls are able to realize the system goal but justified controls are unable to do so. Thus, both the attacker and the actor are able to operationalize their goals depending on the nature of the controls. The dynamic behavior of adversaries and the actors can be analyzed using the CPNs implementation of the EMC model.

As shown in Figure 8.11 and Figure 8.12, the places "Transparency" and "Availability" contain both "StrongControl" and "StrongAttack" tokens depicting the success and failure of the security and privacy controls in the two models.

8.8 Conclusion

This article introduces an extended misuse case (EMC) model for modeling privacy and security risk and formalizes the model with a tool capable of communicating risk to both technical and non-technical stakeholders. The EMC model integrates both privacy and security requirements in a single model by incorporating system goals, assets and rights. The EMC model is validated by means of colored petri nets (CPNs) and employed to automatically detect privacy and security violations in identity management systems (IDMSs). The CPNs instantiation of the EMC model can be extended with additional concepts to visually communicate risk to non-technical stakeholders [67].

Chapter 9

Article V

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

in Springer LNCS 2013 Proceedings of the 10th international conference on Trust, privacy and security in digital business

Ebenezer Paintsil and Lothar Fritsch

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 the risk analysis process. 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.

9.1 Introduction

Identity management systems (IDMSs) create and manage identities of end-users [69]. 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 [91] 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 [22]. 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 the privacy and security risks analysis. It relies on system behaviors or characteristics to analyze privacy and security risks in IDMSs rather than data on past events or intuitions of a risk assessor. Consequently, it can reduce subjectivity and uncertainty in the risk analysis of IDMSs. In addition, EM-BRAM employs Colored Petri Nets (CPNs) [77] to analyze a system's risk. CPNs are easy to learn and use because they hide large portion of complex mathematics and have a high degree of automation.

The rest of the article is organized as follows: Section 9.2 discusses the related work. Section 9.3 presents part of the privacy and security risks model for IDMSs. Section 9.4

is a case study on how the risk analysis method works. Finally, Section 9.5 concludes the article.

9.2 Related work

Gajek et al. [43] 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 CarsdSpace 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.

Gross's security analysis of the security assertion markup language (SAML) single sign-on protocol discovered flaws that allow attacks on the protocol. The attacks discovered include man-in-the-middle attack, replay attack, attack by information leakage, and impersonation [46]. He identified that the end-user's browser can connect to inter-site and transfer URL and other data without authentication. The end-user's browser cannot necessarily verify the IdP's certificate in order to identify it. This lack of verification is a cornerstone of man-in-the-middle attacks on the communication between browser and IdP's site. Gross analyzed the attacks in each of the six steps of the SAML protocol.

However, the above methods rely on manual inspections. Manual inspections are time consuming and therefore expansive. Rather than manual analysis, this article attempts to automate privacy and security risks in IDMSs.

Current risk analysis methods for IDMSs are mainly qualitative, rely on manual inspections and incomplete because the stakeholders' interests are ignored [22], [166]. However, the metric-based framework proposed by Cabarcos et al. [22] 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. In addition, the framework by Cabarcos et al. [22] is based on the assumption that subjectivity or uncertainty is removed in risk assessment by combining probability of past event with stochastic processes. However, this hybrid approach may only minimize the uncertainty in the subjective risk estimation [113].

Suriadi et. al. [154] formally evaluated two security and privacy goals of IDMSs. They showed that end-users could maintain anonymity throughout multiple single sign-on (SSO) 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.

9.3 Risk Analysis Model

This section presents an aspect of the risk analysis model that can be employed to analyze privacy and security risks in IDMSs. The full risk model can be found in [113]. The risk model is developed from a Delphi study on characteristics of tokens or information that flow in IDMSs [115]. We studied tokens because they are personal data sources and gateway to personal data [103]. 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 9.1 represents a partial risk model for IDMSs (full model can be found in [113]). It focuses on the characteristics of tokens that can threaten privacy and security of IDMSs. The external risk contributing factors consist of factors that may be outside the

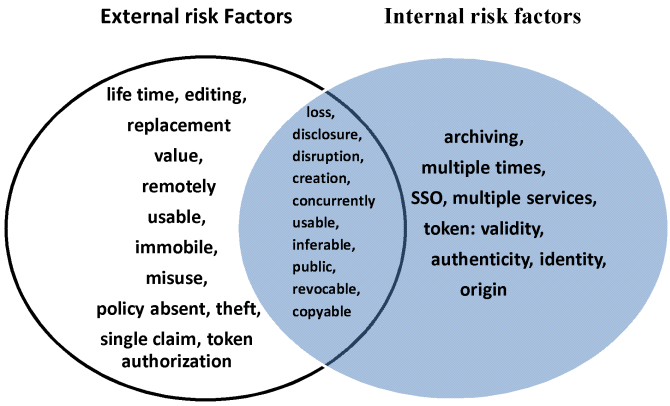

Figure 9.1: Internal and External Risk Factors [115]

control of IDMSs or factors that could not be easily verified. For example, risk contributing factors that affect privacy policy or access control may lie outside the control of IDMSs. Internal factors are those under the control of IDMSs and they may be relied upon to automate the privacy and security risks analysis. The intersection represents both internal and external factors. We discuss the internal factors as follows:

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 [169]. Although SSO reduces human error, it leads to sharing of valuable information across services or domains [89].

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 [89]. 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 [101], [128]. 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 [101]. For example, the character space for English keyboard is 94. The entropy of a biometric template can be found in [128].

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. Revocation can occur internally or externally.

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 low cost RFID tag with no additional security could easily be read by anyone with an appropriate reader but a high cost 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. In order to mitigate loss of tokens, the IDMS must be free of conflict. Token loss can also occur externally.

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.

To enhance security, an IDMS should have a mechanism for checking the authority who issues a token if the token is a credential. The credential should contain the necessary data to facilitate the authentication. If the token is a mere assertion then the IDMS should provide a different mechanism to ensure the authenticity of the assertion. In order to enhance token security, there should be a means of ensuring the validity, identity and authenticity of the token [87].

Token's origin: Refers to the origin of a token. The authority who issued the token should be clearly identified.

Token's authenticity: Determines if a token belongs to the entity presenting it to the IDMS. Authenticity of a token must be checked in order to mitigate privacy and security risks.

Token's identity: Determines if a token identifies the subject or the entity possessing the token. The IDMS should have a mechanism for identification.

Token's validity: Determines if a token has not expired, its lifespan is within the validity period or has passed the validity test.

We use state space analysis [77] to analyze whether these risk factors exist in a targeted IDMS. The external factors can guide privacy and security policy formulation for IDMSs. Note that security token is different from CPNs tokens. A CPNs token is a variable with data type and a value.

9.4 Case Study and Application of the Model

As a case study, we apply the model in Figure 9.1 to assess the privacy and security risks of SAML SSO service (SAML-SSOS) for Google Apps [45]. The SAML-SSO for Google Apps has been analyzed by Alessandro et al. [12] using LTL, and a weakness is exposed. In this case study, we show that more privacy and security risks are inherent in SAML-SSO for Google Apps analyzed by Alessandro et al. [12] using our EM-BRAM. The attack scenario for the SAML-SSOS IDMS is shown in Figure 9.2 and explained below:

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.

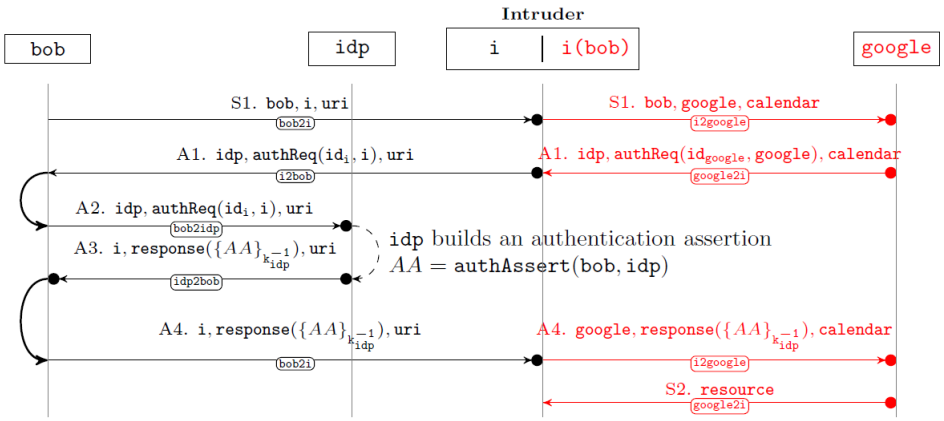

Figure 9.2: SAML SSO Service for Google Apps and an Intruder SP [12]

6. A4 and S2: The Intruder SP sends a response to the Google cloud service and receives bob's resources.

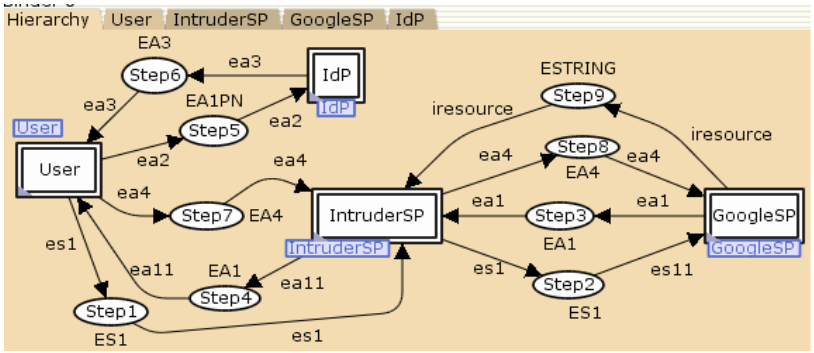

Figure 9.3: Hierarchical CPNs Model for SAML-SSO Service for Google Apps

CPNs' models consist of **places, transitions (events), input and output** arcs. We represent the places by ellipses, transitions by rectangles, input/output arcs by directed arcs [77]. 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 9.3 is the hierarchical CPNs model for the SAML-SSOS attack scenario described in Figure 9.2. We use the hierarchical CPNs to make a large model manageable and compositional. Figure 9.3 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 9.2. The numbered places represent the possible sequence of information flow in the top-level model. For ex-

ample, "Step1" in Figure 9.3 represents the first place to receive token in the hierarchical model.

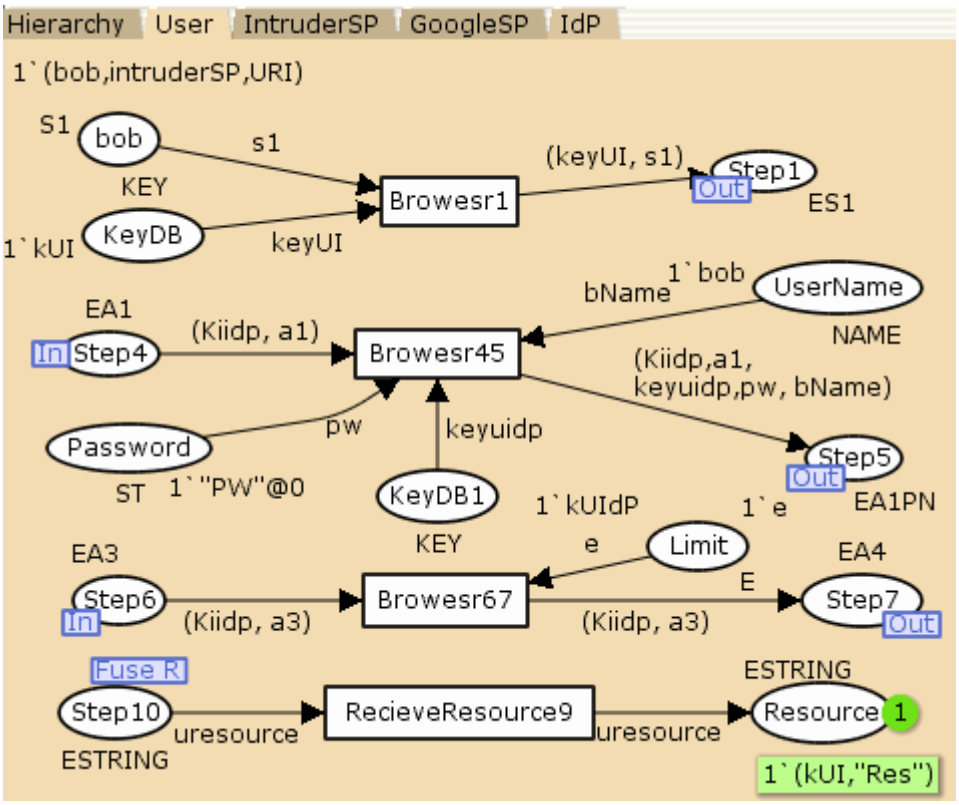

Figure 9.4: User

The system is initiated by the "User" sub-model. The "User" sub-model is in Figure 9.4. In Figure 9.4, 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 9.3, the top-level model. The token is sent through "Step1" in Figure 9.3 to the "IntruderSP" sub-model in Figure 9.5. 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 9.6, 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 9.5, 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 9.7. The "IdP" transition authenticates the token and issues an assertion via "Step6" output port. The "User" sub-model in Figure 9.4 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".

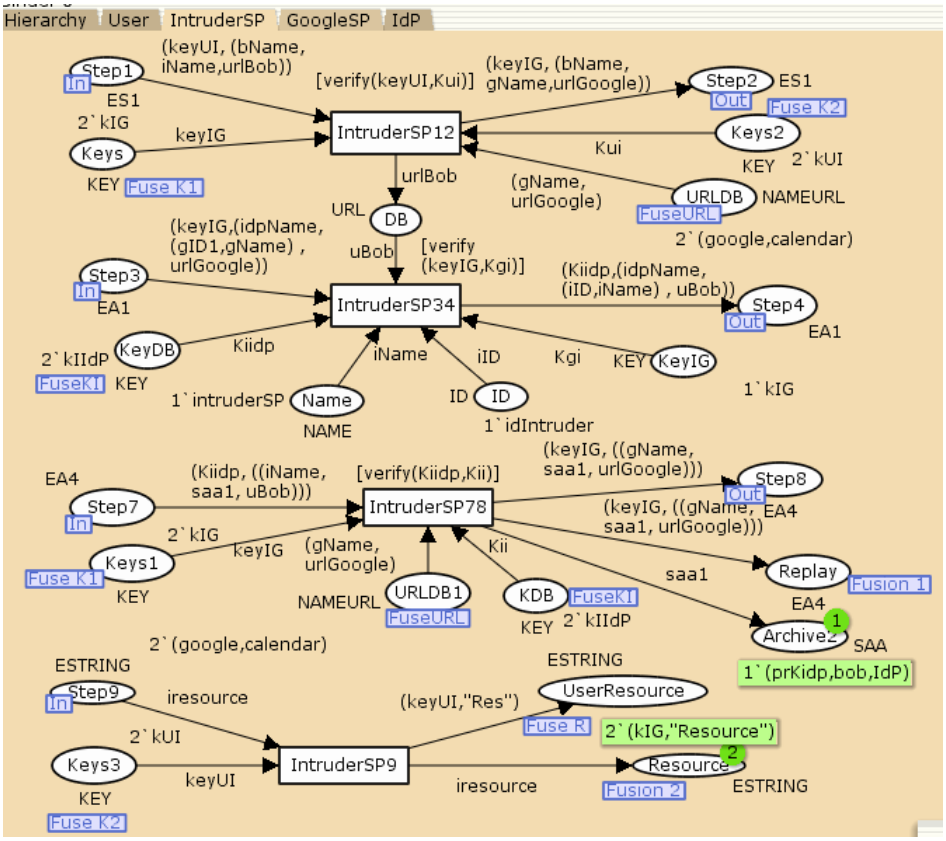

Figure 9.5: Intruder SP

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 compound. The atomic fields or data types include keys and IDs. Compound 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)=if k1=k2 then true else false;
val users =[bob,alice];val pwDB=["PW", "PW1"];
```

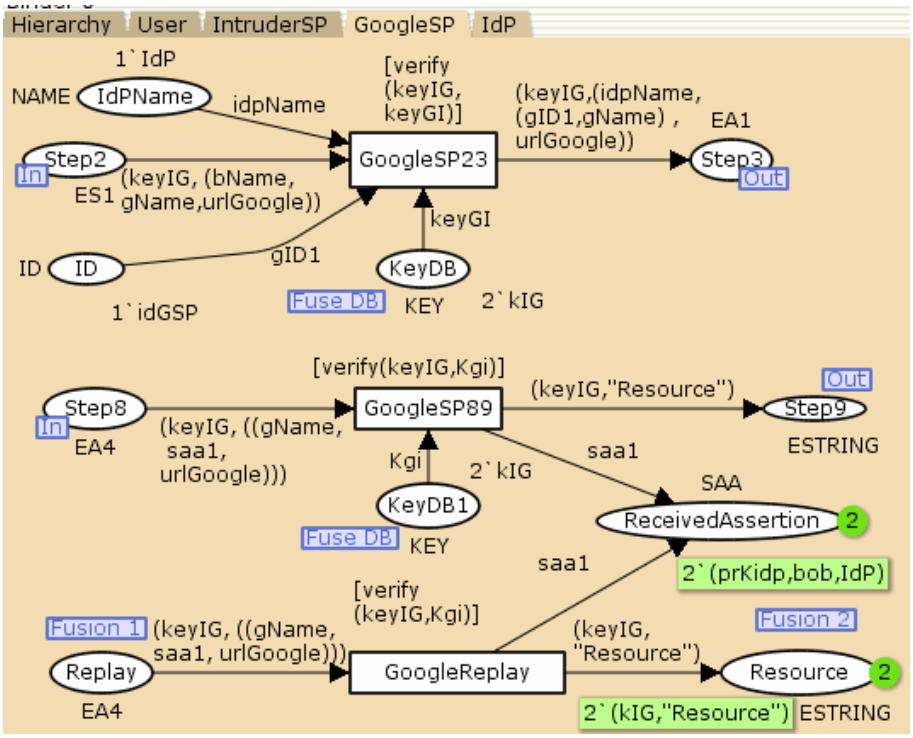

Figure 9.6: Google SP

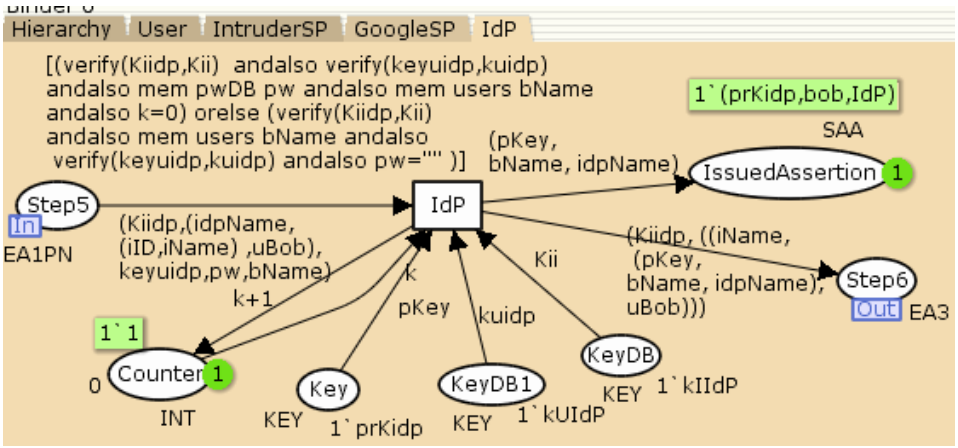

Figure 9.7: IdP

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.

9.4.1 Privacy and Security Risks Analysis

This section shows how privacy and security risks of the SAML-SSOS model in Figure 9.3 can be analyzed. The objective of risk analysis is to identify and assess all risks and then to suggest a set of controls that will reduce these risks to an acceptable level [44]. 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 [113].

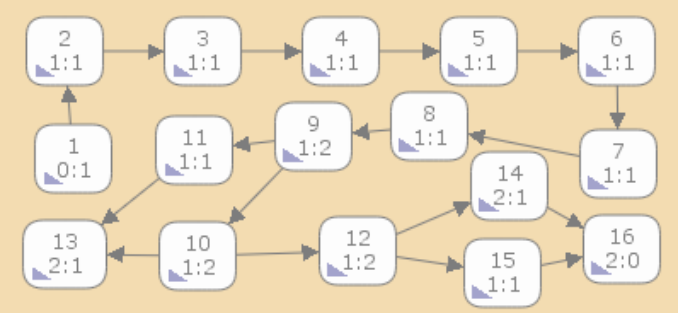

Figure 9.8: State Space Graph

We use CPNs simulation tools to check the correctness of the model. We then use the CPNTools [77] to generate the state space graph of the system model and search through the graph for privacy and security risks. Figure 9.8 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 9.6 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. [174]. The SAML-SSOS IDMS requires four attributes for assertions i.e. AuthAssert(ID,User, IdP, SP) [12], 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 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 mean the entire state spce graph. The second argument maps each node into a boolean and uses the nodes that are true for the analysis. The third argument 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 or function combine the results obtained from the fourth and the fifth arguments [80].

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" domain but occurs in state 5 and 6 in the IdP and User domain. 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 secret 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 concurrent use 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 [77]. 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 or deadlock 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 the "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.

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 No	Token has insufficient number of attributes 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 9.1: Risk Analysis Report for SAML-SSOS for Google Apps

Table 9.1 is the summary report of the risk identified in the vulnerable SAML-SSOS for Google Apps IDMS. We can now use the full risk model in [113] to determine the appropriate security control.

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 analysis revealed more detailed privacy and security risks than the ones discovered by Alessandro et al. [12]. 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 the subjectivity.

Chapter 10

Article VI

Towards Automation of Risk Analysis in Identity Management Systems,

in Trust, Security and Privacy in Computing and Communications (TrustCom2013) IEEE CPS 12th International Conference

Ebenezer Paintsil

Abstract

Currently, risk analysis methods for identity management systems (IDMSs) mainly rely on manual inspections. Manual inspection is time consuming and expensive. This article introduces the executable model-based risk analysis method (EM-BRAM) with the aim of automating privacy and security risks analysis in IDMSs. The EM-BRAM identifies risk factors inherent in 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. We apply the method to analyze privacy and security risk in the OpenID IDMS.

10.1 Introduction

Identity management systems (IDMSs) create and manage identities of end-users [69]. 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).

In order to analyze a system, we need a clear understanding of the system as well as what can go wrong in it. Risk assessors often rely on system stakeholders or people familiar with a system such as end-users and system owners to understand the target system [91], [58]. Similarly, system stakeholders rely on risk assessors to understand the risk analysis process. Risk analysis can be improved if the system stakeholders can understand the risk analysis process. Model-based risk analysis methods use graphical models to facilitate participation, risk communication and documentation.

However, current model-based risk analysis methods provide general support for risk analysis but do not target IDMSs. In addition, due to lack of data on past events, model-based risk analysis methods either rely on the subjective intuitions of risk assessors and system stakeholders (end-users and system owners) or complex mathematical validation techniques to determine a system's risk. Subjective risk analysis can lead to high uncertainties in the risk analysis result [25]. Moreover, complex mathematical risk modeling and validation techniques such as UMLSec [71] are difficult to learn and can impede risk communication among system stakeholders [52].

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 privacy and security risks analysis in IDMSs. It uses Colored Petri Nets (CPNs) [77] which hide large portion of complex mathematics thereby making the method relatively easy to use and less expensive.

The rest of the article is organized as follows: Section 10.2 is the related work. Section 10.3 discusses some of the advantages and disadvantages of CPNs. Section 10.4 presents the privacy and security risks model for IDMSs. Section 10.5 is a case study on how the risk analysis method works. Finally, Section 10.6 concludes the article.

10.2 Related work

Gajek et al. [43] manually analyze the Microsoft CardSpace IDMS. The analysis focused on how the vulnerabilities of a browser can threaten the security of the Microsoft CardSpace. They described 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 CarsdSpace 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. The SOP is a security policy universally supported in browsers. SOP policy enforces that the access of scripts is limited to objects originating from the same source.

Gross's security analysis of the security assertion markup language (SAML) single sign-on protocol discovered flaws that allow attacks on the protocol. The attacks discovered include man-in-the-middle attack, replay attack, attack by information leakage, and impersonation [46]. He identified that the end-user's browser can connect to inter-site and transfer URL and other data without authentication. The end-user's browser cannot necessarily verify the IdP's certificate in order to identify it. This lack of verification is a cornerstone of man-in-the-middle attacks on the communication between browser and IdP's site. Gross analyzed the attacks in each of the six steps of the SAML protocol.

The above methods rely on manual inspections. Rather than manual analysis, this article uses CPNs tools to support privacy and security risks in IDMSs.

Current risk analysis methods for IDMSs are mainly qualitative, rely on manual inspections and incomplete because the stakeholders' interests are ignored [22], [166]. However, the metric-based framework proposed by Cabarcos et al. [22] 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. In addition, the framework by Cabarcos et al. [22] is based on the assumption that subjectivity or uncertainty is removed in risk assessment by combining probability of past event with stochastic processes. However, this hybrid approach may only minimize the uncertainty in the subjective risk estimation [113].

Suriadi et al. [154] formally evaluated two security and privacy goals of IDMSs. They showed that end-users could maintain anonymity throughout multiple single sign-on (SSO) 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.

10.3 Colored Petri Nets

This section introduces and discusses some of the advantages and disadvantages of CPNs. CPNs' models consist of **places, transitions (events), input and output** arcs. We represent the places by ellipses, transitions by rectangles, input/output arcs by directed arcs [77]. 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. 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.

The EM-BRAM is supported by CPNs modeling and analysis. We begin the assessment by modeling a given IDMS with CPNs and analyze how the characteristics of information that flows in the IDMS contribute to privacy and security risks. CPNs stand out among the model-based formal methods (methods that rely on abstract state machine or state space analysis [6]) [172]. They hide large portion of complex mathematics and have a high degree of automation making it relatively easy to learn and use. They provide tools for verification, validation and automatic analysis of system models [77]. CPNs tools are able to model, debug and test a large scale, critical and complex concurrent system. They are suitable for a system that requires a large number of possible executions. CPNs are graphical language supported by a tool with the capabilities of a high-level programming language. It includes time concepts making it suitable for performance analysis. CPNs have a concise mathematical definition which contains very few but powerful primitives making it easy to learn, use and to develop strong analysis method by which properties of system models can be proved [164]. It is flexible in terms of token definition and manipulation. Various elements such as use cases, messages and task can be represented by different types of tokens.

However, CPN Tools are not interactive and difficult to implement new techniques such as new more efficient state space methods [165]. CPNs' models suffer from state space explosion. This can affect risk assessment of an IDMS with very large states. CPNs model and verify a system model and not the actual system itself therefore any obtained result is as good as the system model.

10.4 Risk Analysis Model

This section presents the factors used for the privacy and security risks analysis in IDMSs. The risk contributing factors were obtained from a Delphi study [115]. The Delphi study focused on the characteristics of token that can contribute to privacy or security risk in IDMSs. We studied tokens because they are personal data sources and gateway to personal data [103]. A token is a technical artifact providing assurance about an identity [115]. It is used for communicating authentication entitlement and attribute information in IDMSs. 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.

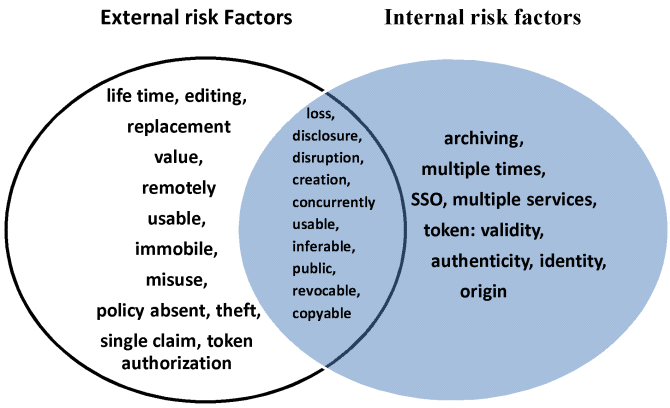

Figure 10.1: Internal and External Risk Factors [115]

The Figure 10.1 represents the potential risk contributing factors for IDMSs. The exter-

nal factors are not the focus of this article. The intersection represents both internal and external factors. We discuss the internal factors as follows:

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 token 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 sometimes without the consent of the end-user. IDMSs that support single sign-on (SSO) allow tokens to be used for multiple services sometimes in multiple domains upon a single authentication [169]. Although SSO reduces human error, it leads to sharing of valuable information across services or domains [89].

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 [89]. 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 [101], [128]. 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 [101]. For example, the character space for English keyboard is 94. The entropy of a biometric template can be found in [128].

When a token secret is revoked the user of the token could be identified or confidential information may be made available to unauthorized persons. Revocation can be internal or external.

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 low cost RFID tag with no additional security could easily be read by anyone with an appropriate reader but a high cost 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. In order to mitigate loss of tokens, the IDMS must be free of conflict. Token loss can also occur externally.

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.

To enhance security, an IDMS should have a mechanism for checking the authority who issues a token if the token is a credential. The credential should contain the necessary data to facilitate the authentication. If the token is a mere assertion then the IDMS should provide a different mechanism to ensure the authenticity of the assertion. In order to enhance token security, there should be a means of ensuring the validity, identity and authenticity of the token [87].

Token's origin: Refers to the origin of a token. The authority who issued the token should be clearly identified.

Token's authenticity: Determines if a token belongs to the entity presenting it to the IDMS. Authenticity of a token must be checked in order to mitigate privacy and security risks.

Token's identity: Determines if a token identifies the subject or the entity possessing the token. The IDMS should have a mechanism for identification.

Token's validity: Determines if a token has not expired, its lifespan is within the validity period or has passed the validity test.

10.5 Case Study and Application of the Model

This section applies the model in Figure 10.1 to evaluate the security and privacy risks the in OpenID [129] IDMS. OpenID is one of the popular IDMSs. The scenarios for OpenID are as follows:

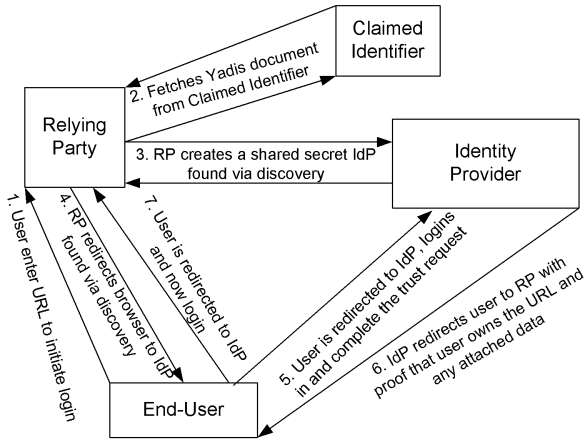

Figure 10.2: OpenID IDMS [129]

Scenario for OpenID:

1. End-user supplies his username, email address and OpenID password and obtains his OpenID URL from an IdP (e.g. http://username.myopenid.com/).
2. End-user can now visit a relying party's (RP) site.
3. The end-user supplies his OpenID URL (username.myopenid.com) to the OpenID login form on the RP's webpage.
4. The RP discovers the IdP from a "Claimed Identifier" provider and transforms the end-user's OpenID URL (e.g. http://username.myopenid.com/).
5. The end-user is redirected to the IdP to supply his credentials (e.g. username and password).
6. If the authentication is successful he is redirected back to the RP so that he can access the resources and services provided by the RP's site.

Figure 10.3 is the hierarchical CPNs model for the OpenID IDMS described in the scenario above. We use the hierarchical CPNs to make the large and complex model manageable and compositional. The transitions with double lines are the substitution transitions. They are used to divide large model into sub-models. The substitution transitions represent the system agents – "User", "IdP", "ClaimedID" and "RelyingParty". For example, all

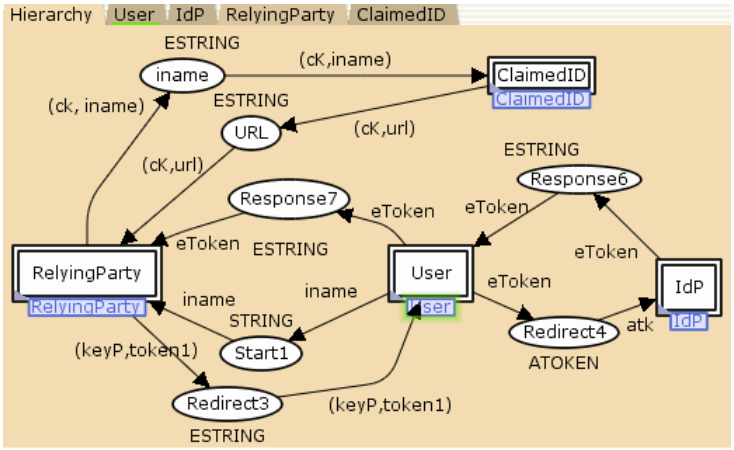

Figure 10.3: Hierarchical CPNs Model for OpenID

the events and states in the entity "User" agent is represented by the "User" substitution transition in Figure 10.3. In Figure 10.3, the "RelyingParty" substitution transition is the service provider and the "IdP" is the IdP. The numbered places enable information flow among the system agents. For example, the place "Start1" represents the first stage in the protocol.

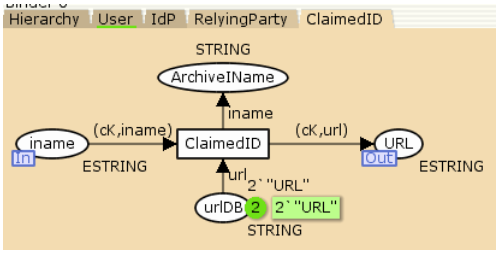

Figure 10.4: OpenID Claimed ID Model

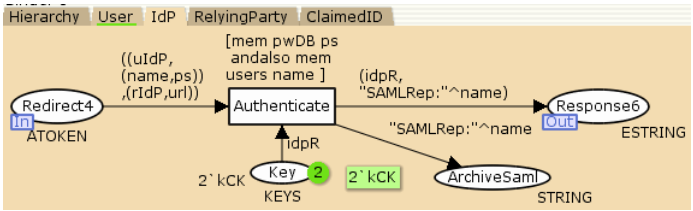

Figure 10.5: OpenID IdP Model

Figure 10.6 is the "User" sub-model for the OpenID IDMS. The "Browser1" is the first transition to execute in the sub-model. It takes the "url" from the "InitStart" place and sends it to the "RP". The "RP" is an output port so it sends the "url" to the top-level substitution transition "RelyingParty" in Figure 10.3.

In the "RelyingParty" sub-model in Figure 10.7, "start" place receives the "iname". The "claim" transition encrypt the "iname" and sends it to the "ClaimedID" sub-model in Figure 10.4. The input port "iname" receives the token for the sub-model. The "ClaimedID" transition generates the URL and forwards it to the "RelyingParty" in Figure 10.7. The

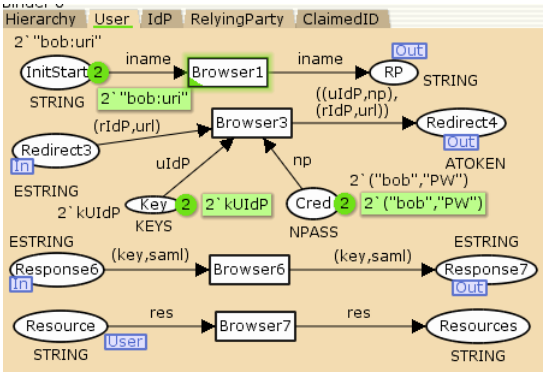

Figure 10.6: OpenID End-User Model

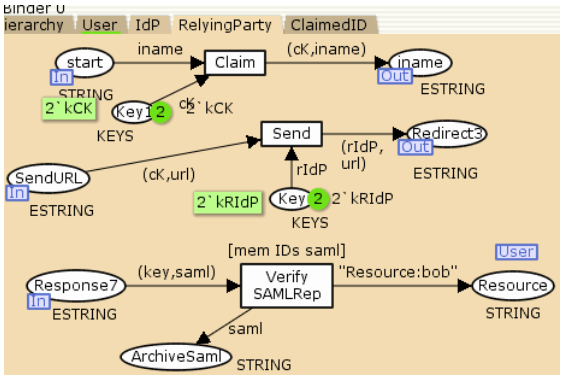

Figure 10.7: OpenID SP/Relying Party Model

"RelyingParty" receives the "URL" and forwards it to the "User" sub-model. The end-user includes the user name and password and forward the token to the "IdP" sub-model. The "IdP" sub-model authenticates the end-user and creates an assertion for her if the authentication is successful. The assertion is sent via the "User" sub-model to the "RelyingParty" sub-model. The "RelyingParty" then verifies the assertion and grant access to the resource if the assertion is correct. The assertion is archived by the "RelyingParty". The "Resource" place in the sub-model is a fusion set. A fusion set allows two identical places to hold the same tokens.

The tokens of the IDMS are composed of fields. Some of the fields are atomic others are compound. The atomic fields or data types include keys. Compound fields are constructed from the atomic ones. For example, a cipher is given by the order pair (K,Token) where K is an encryption key and Token is the message.

The data types or colors are declared as follows:

```
colset TOKEN=product STRING*STRING;colset ETOKEN=
product KEYS * TOKEN;colset ESTRING = product KEYS*STRING;
colset NPASS = product STRING*STRING;colset ENPASS=
product KEYS*NPASS;colset ESTRING = product KEYS*STRING;
colset ATOKEN = product ENPASS*ESTRING;colset KEYS = with
IdP12|puKsp|kGP|kRIdP|kUIdP|kUP;var auth, saml, saml2, url,
rep2, token1:STRING;var key,keyG,keyGP,keyIdP2,keyPI,
keyI, keyG:KEYS;val users =["bob","alice"];
val pwDB=["PW","PW1"];var atk:ATOKEN;var np:NPASS;
```

```
var keyP,key,keyG,keyGP,keyIdP2,kIdP2, idpR, keyPI, keyI,
keyG,cK,rIdP, uIdP:KEYS;var eToken:ESTRING;
colset AUTH=product ESTRING*ETOKEN;
```

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.

10.5.1 Validation of the CPNs Model

This section validates the behavioral correctness of the OpenID CPNs model above before the risk analysis. Validation loosely refers to "the process of determining that a design is correct" [36]. According to Holzmann [53], the three important general properties for a correct model are, absence of cycle, deadlock and improper termination. We used the CPNs simulation tools to validate the OpenID model and the state space tools to generate state space report for further analysis. The simulation result showed that the system model behaved as expected.

The statistics of the state space report generated by the state space tools [77] show that the state space of the OpenID model has 55 nodes and 90 arcs. Similar to the state space results, the strongly connected components graph (maximal subgraph in which it is possible to find a path from one node to another) has the same number of nodes and arcs. This means there is no infinite occurrence sequences or cycle in the model. Hence we can be sure that the model will terminate. The only worry is whether the model terminates prematurely or not. The state space analysis report shows that the model has only one dead marking which happens to be the home marking and the last node (node 55). This indicates that the model may terminate at the last node, hence deadlock and improper termination may not be possible. We conclude that the OpenID model in Figure 10.3 may be correct.

10.5.2 Privacy and Security Risks Analysis

This section shows how we analyze the privacy and security risks of the OpenID model in Figure 10.3. The objective of risk analysis is to identify and assess all risks and then to suggest a set of controls that will reduce these risks to an acceptable level [44]. Hence, we analyze if the characteristics of tokens in the OpenID IDMS threaten system privacy and security and suggest suitable controls based on the risk model in [113].

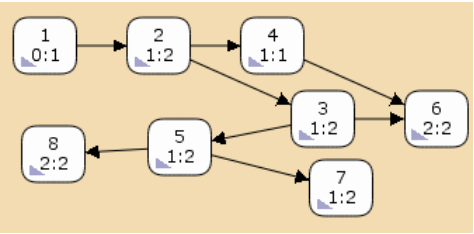

Figure 10.8: State Space Nodes of OpenID Model

We generate the state space graph in Figure 10.8 using the CPNTools [77] and search through the graph for the system risks. The state space graph has 55 nodes and 90 arcs. Figure 10.8 depicts only seven nodes. The nodes represent the system state and arcs represent the transitions from one state to another. The arcs consist of binding elements or data that flow from one node or state to another. We analyze the privacy or security risk as follows:

Multiple times: The place "RelyingParty'Resource" in Figure 10.7 stores all the tokens or assertions received by the "RelyingParty". To verify multiple use of token, we use the query "PredAllNodes(multipleUse())" to find the upper integer bound of all the nodes

where tokens on the "RelyingParty'Resource" place is greater than 1. The result shows that multiple use of tokens occurred at node 52. This means the end-user can be profiled by the relying party, hence we have profiling or linkability risk. The nodes are created in the space state graph generated by the CPNTools [77].

```
fun multipleUse()=fn n=>
size(Mark.RelyingParty'Resource 1 n)>1;
```

SSO/Multiple services: The alias or the identity "bob" is not supposed to be seen by the "ClaimedID" agent because "bob" requested a resource from the "RelyingParty". However, because the end-user's i-name may include the identity "bob", the identity "bob" may be used for discovery outside the requested domain. We verify this risk using the query "PredAllNodes(isMultipleServices())". The function "isMultipleService()" is defined below. The results show that the identity "bob" appears in ClaimedID's domain at node 7,8,...55.

```
fun isMultipleServices()= fn n => isSubstring "bob"
(st_Mark.ClaimedID'ArchiveIName 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. [174]. 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 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)
```

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(1,55)". The function returns true, which means 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.

The "SearchNodes" function 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 or function combine the results obtained from the fourth and the fifth arguments [80].

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 "RelyingParty"
(ArcDescriptor n) ) orelse (isSubstring "PW"
(ArcDescriptor n) andalso isSubstring "ClaimedID"
(ArcDescriptor n)), NoLimit, fn n => n, [],op ::)
```

The result shows that the password "PW" is not found outside the "User" and "IdP" domain. Hence, the password between the User and the IdP is not public or in an unauthorized domain.

copyable/Concurrently usable: We use the following query to determine if a token is copyable.

```
SearchNodes (EntireGraph,fn n => (isSubstring "empty"
(st_Mark.Hierarchy'Redirect3 1 n))=false andalso (isSubstring "k"
(st_Mark.Hierarchy'Redirect3 1 n))=false,NoLimit, fn n => st_Mark.
Hierarchy'Redirect3 1 n, [],op ::)
```

The query retrieves all tokens that pass through the intermediate places (Start1, Redirect1,..Response7) 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 through the system in plain text. The results of the query is empty list which indicates that no unencrypted token passes through "Redirect3". The query is repeated for all intermediate nodes.

Secondly, we verify if an encryption key can be found in an unauthorized database using the query below. The secret keys used in the sessions are kUIdP, kCK and kRIdP. kUIdP is the secret key for User and IdP while kCK is that of the ClaimedID and IdP. The session key kRIdP is for the "RelyingParty" and "IdP".

```
SearchNodes (EntireGraph, fn n => (isSubstring "kUIdP"
(ArcDescriptor n) andalso isSubstring "RelyingParty"
(ArcDescriptor n) ) orelse (isSubstring "kUIdP" (ArcDescriptor n)
andalso isSubstring "ClaimedID" (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 concurrent use 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 [77]. Hence, the model supports concurrent use of tokens.

loss, disclosure/disruption: The function "ListDeadMarkings()" returns the node 55 as the only dead node. In addition, the function "Terminal(55)" returns true which indicates that node 55 is the terminal node. Furthermore, the output from the "print(NodeDescriptor 55)" function shows that all the places in the system are empty except the archive places. This means all the tokens have been received at node 55, hence, there was no conflict or deadlock 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 using the query below.

```
SearchNodes (EntireGraph, fn n =>  (isSubstring
"empty" (st_Mark.Hierarchy'Redirect3 1 n))=false,
NoLimit, fn n => st_Mark.Hierarchy'Redirect3 1 n, [], op ::)
```

The query is repeated for all the intermediate places. The query retrieves all tokens that pass through the intermediate place "Redirect3". This can be examined for originators of the tokens. For example, a token from the "IdP" to the "User" must contain the originator of the token which is the "IdP". The tokens in both models have no originators because we considered only the high level specifications.

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

```
fun auth1()=(Mark.IdP'ArchiveSaml 1 55)=
(Mark.RelyingParty'ArchiveSaml 1 55);
```

The query returns true which indicates that the tokens were authentic.

We use the query below to verify whether identification of the end-user was successful. The query verifies if the authentication was successful and tokens used for authentication contain the identity of the end-user "bob". The query returns true indicating successful identification. If the token returns false then we need to examine the tokens on the places "RelyingParty'ArchiveSaml" and "IdP'ArchiveSaml" to see whether the inconsistency was caused by the identities or assertions.

```
fun isIdentity()=(Mark.IdP'ArchiveSaml 1 55)=
(Mark.RelyingParty'ArchiveSaml 1 55) andalso
isSubstring "bob" (st_Mark.RelyingParty'ArchiveSaml 1 55)
andalso isSubstring "bob" (st_Mark.IdP'ArchiveSaml 1 55);
```

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 "RelyingParty". To simplify the model, this scenario is not considered.

Factors	**Risk Value**	Meaning
multiple times	Yes	Tokens can be linked or profiled by a SP
single sign-on/ multiple services	Yes	Tokens may be linked or profiled by ClaimedID Token is not profiled outside the trusted domain
creation	- No	Not considered or modeled Token has no sensitive attributes
archiving	Yes	Token can be archived by Relying parties/SPs
public	No	Token secret is kept private between end-users and IdP
inferable	Yes	Tokens' secret can be guessed by Relying Party/SPs
copyable	No	Tokens cannot be copied
concurrently usable	Yes	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	-	Not considered or modeled
authentication	No	Tokens' authenticity test did not fail
identification	No	Tokens include the identity of the end-user
validation	-	Not considered or modeled

Table 10.1: Risk Analysis Report for OpenID

Table 10.1 is the summary report of the risk identified in the OpenID. We are unable to verify some of the risk factors because of the high level specifications used in the analysis. All the risk factors can be verified if we use low-level specifications. After the verification or analysis, the extended misuse case models in [113] can be used to determine some of the security controls for the IDMS. For example, if a token is used multiple times then one-time token is an appropriate security control.

10.6 Conclusion

This article presents a method for automation of risk analysis in identity management systems (IDMSs). The method models a targeted IDMS with colored Petri nets (CPNs) and generates a state space graph for the analysis. The security and privacy risks of the IDMS are verified using ML predicate functions and queries. The method has the potential of improving security, reducing cost and subjectivity in risk analysis of IDMSs. The strength of the method depends on the risk model and the specification used in the risk analysis. Future work will consider low level specifications with attack models and also investigate if more risk contributing factors can be identified.

Bibliography

[1] 17, 57

[2] ABU-NIMEH, S., AND MEAD, N. Privacy risk assessment in privacy requirements engineering. In *Requirements Engineering and Law (RELAW), 2009 Second International Workshop on* (sept. 2009), pp. 17–18. 31

[3] AICPA/CICA. Aicpa/cica privacy risk assessment tool. Tech. Rep. 2.0, American Institute of Certified Public Accountants and the Canadian Institute of Chartered Accountants, 2010. 13

[4] ALBERTS, C., AND DOROFEE, A. *Managing information security risks: the OCTAVE approach*. Addison-Wesley Professional, 2002. 43

[5] ALEXANDER, I. F. Initial industrial experience of misuse cases in trade-off analysis. In *Proceedings of the 10th Anniversary IEEE Joint International Conference on Requirements Engineering* (Washington, DC, USA, 2002), RE '02, IEEE Computer Society, pp. 61–70. Available from: http://dl.acm.org/citation.cfm?id=647648.731617. 18, 67, 68, 79

[6] ALMEIDA, J. B., FRADE, M. J., AND [ET AL.], J. S. P. . *Rigorous software development : an introduction to program verification*. Undergraduate topics in computer science. Springer, London, 2011. Available from: http://opac.inria.fr/record=b1132575. 95

[7] ALPCAN, T. Dynamic incentives for risk management. In *New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on* (may 2012), pp. 1 –5. 43

[8] ALVAREZ, G., AND PETROVIC, S. A new taxonomy of web attacks suitable for efficient encoding. *Computers and Security 22*, 5 (2003), 435 – 449. 44

[9] ARDAGNA, C., BUSSARD, L., DE CAPITANI DI VIMERCATI, S., NEVEN, G., PARABOSCHI, S., AND PEDRINI. PrimeLife Policy Language. In *W3C Workshop on Access Control Application Scenarios* (Luxembourg, 17-Nov-2009 2009). Available from: http://www.w3.org/2009/policy-ws/papers/Trabelisi.pdf. 47

[10] ARDAGNA, C., BUSSARD, L., ET AL. PrimeLife Policy Language. In *W3C Workshop on Access Control Application Scenarios* (Luxembourg, 17-Nov-2009 2009), PrimeLife. 16, 29, 30, 49, 56

[11] ARIAS-CABARCOS, A.-M., ET AL. A metric-based approach to assess risk for "on cloud" federated identity management. *Journal of Network and Systems Management 20* (2012), 513–533. 4, 5

[12] ARMANDO, A., CARBONE, R., COMPAGNA, L., CUELLAR, J., AND ABAD, L. T. Formal Analysis of SAML 2.0 Web Browser Single Sign-On: Breaking the SAML-based Single Sign-On for Google Apps. In *the 6th ACM Workshop on Formal Methods in Security Engineering (FMSE 2008)* (Hilton Alexandria Mark Center, Virginia, USA, 2008), ACM Press. xiii, 22, 23, 84, 85, 90, 92

[13] AVEN, T. A semi-quantitative approach to risk analysis, as an alternative to qras. *Reliability Engineering & System Safety 93*, 6 (2008), 790 – 797. 2, 12, 39, 44

[14] BANDARA, A., SHINPEI, H., JURJENS, J., KAIYA, H., KUBO, A., LANEY, R., MOURATIDIS, H., NHLABATSI, A., NUSEIBEH, B., TAHARA, Y., TUN, T., WASHIZAKI, H., YOSHIOKA, N., AND YU, Y. Security patterns: comparing modeling approaches. In *Software Engineering for Secure Systems: Industrial and Research Perspectives*, H. Mouratidis, Ed. IGI Global, Hershey, PA, October 2010, pp. 75–111. 7, 68

[15] BAR-OR, O., AND THOMAS, B. Openid explained. http://openidexplained.com/, 2010. [Online; accessed 18-Aug-2010]. 51

[16] BERINATO, S. Finally a real return on security spending, 2002. http://www.cio.com.au/article/52650/finally_real_return_security_spending/. 54

[17] BHARGAV-SPANTZEL, A., CAMENISCH, J., GROSS, T., AND SOMMER, D. User centricity: A taxonomy and open issues. In *DIM '06: Proceedings of the second ACM workshop on Digital identity management* (New York, NY, USA, 2006), ACM, pp. 1–10. 27

[18] BHASKAR, R., CHANDRASEKARAN, K., LOKAM, S. V., MONTGOMERY, P. L., VENKATESAN, R., AND YACOBI, Y. Vulnerabilities in anonymous credential systems. *Electron. Notes Theor. Comput. Sci. 197* (February 2008), 141–148. 29

[19] BRAMHALL, P., HANSEN, M., RANNENBERG, K., AND ROESSLER, T. User-centric identity management: New trends in standardization and regulation. *IEEE Security and Privacy 5* (2007), 84–87. 16, 49, 56

[20] BROWN JAMES, S. P., AND H., W. C. U-Prove CTP R2 Whitepaper. Tech. rep., Microsoft Corporation, Feb 2011. 29

[21] BYGRAVE, L. A. *Data Protection Law Approaching Its Rationale, Logic and Limits*. Kluwer Law International, 2002. 15, 28, 49, 56, 72, 74

[22] CABARCOS, P. Risk assessment for better identity management in pervasive environments. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on* (march 2011), pp. 389–390. 2, 31, 81, 82, 94

[23] CAMENISCH, J., AND HERREWEGHEN, E. V. Design and implementation of the idemix anonymous credential system, 2002. 29, 46, 47, 50, 74

[24] CAMENISCH, J., AND LYSYANSKAYA, A. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *EUROCRYPT '01: Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques* (London, UK, 2001), Springer-Verlag, pp. 93–118. 46

[25] CAMPBELL, H. Risk assessment: subjective or objective? *ENGINEERING SCIENCE AND EDUCATION JOURNAL* (APRIL 1998). 2, 13, 22, 93

[26] CANTOR, S., KEMP, J., PHILPOTT, R., AND MALER, E. Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0. Tech. rep., OASIS, March 2005. 27

[27] CAO, Y., AND YANG, L. A survey of identity management technology. In *Information Theory and Information Security (ICITIS), 2010 IEEE International Conference on* (dec. 2010), pp. 287–293. 27, 28

[28] CARRERAS, X., MÀRQUEZ, L., PUNYAKANOK, V., AND ROTH, D. Learning and inference for clause identification. In *Proceedings of the 13th European Conference on Machine Learning* (London, UK, UK, 2002), ECML '02, Springer-Verlag, pp. 35–47. 13, 41

[29] CHITRAKAR, R. Business process execution language - specification, modelling and analysis. Tech. rep., Department of Computer Science Univeristy of Illinois at Chicago, July 2006. http://www.cs.uic.edu/ rchitra1/analysis.html. 77

[30] CHORPPATH, A., AND ALPCAN, T. Risk management for it security: When theory meets practice. In *New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on* (may 2012), pp. 1 –5. 13, 41, 43

[31] CLARKE, R. A sufficiently rich model of (id)entity, authentication and authorisation. http://www.rogerclarke.com/ID/IdModel-1002.html, 2010. 45

[32] CRANG, M., AND COOK, I. *Doing Ethnographies*, 2 ed. SAGE Publication Ltd, 2007. 10

[33] DAHL, O. M. Using coloured petri nets in penetration testing, 2005. 43

[34] DELFT, B., AND OOSTDIJK, M. A security analysis of openid. In *Policies and Research in Identity Management*, vol. 343 of *IFIP Advances in Information and Communication Technology*. Springer Berlin Heidelberg, 2010, pp. 73–84. 2, 28, 31

[35] DIREKTORAT FOR FORVALTNING OG IKT (DIFI). Minid–security and privacy, 2011. http://www.difi.no/elektronisk-id/minid/security-and-privacy. 1

[36] EDWARDS, S., LAVAGNO, L., LEE, E., AND SANGIOVANNI-VINCENTELLI, A. Design of embedded systems: formal models, validation, and synthesis. *Proceedings of the IEEE 85*, 3 (1997), 366–390. 22, 100

[37] EEKELS, J., AND ROOZENBURG, N. A methodological comparison of the structures of scientific research and engineering design: their similarities and differences. *Design Studies 12*, 4 (1991), 197 – 203. 10

[38] ELAHI, G., AND YU, E. S. K. Modeling and analysis of security trade-offs - a goal oriented approach. *Data Knowl. Eng.* (2009), 579–598. 18, 24, 43, 67, 68, 79

[39] ENISA. Determining your organizations information risk assessment and management requirements and selecting appropriate methodologies. Tech. Rep. 2.0, ENISA, September 2008. 37, 38

[40] FELTEN, E. Drm, and the first rule of security analysis, 2003. https://freedom-to-tinker.com/blog/felten/drm-and first-rule-security-analysis/. 4

[41] FRITSCH, L. Profiling and Location-Based Services. In *Profiling the European Citizen - Cross-Disciplinary Perspectives*, M. Hildebrandt and S. Gutwirth, Eds. Springer Netherlands, Dordrecht, April 2008 2008, pp. 147–160. Available from: http://www.springer.com/new+&+forthcoming+titles+(default)/book/978-1-4020-6913-0. 16, 49, 56

[42] FRITSCH, L., AND ABIE, H. Towards a research road map for the management of privacy risks in information systems. In *Sicherheit* (2008), A. Alkassar and J. H. Siekmann, Eds., vol. 128 of *LNI*, GI, pp. 1–15. isbn: 978-3-88579-222-2. 54

[43] GAJEK, S., SCHWENK, J., STEINER, M., AND XUAN, C. Risks of the cardspace protocol. In *Proceedings of the 12th International Conference on Information Security* (Berlin, Heidelberg, 2009), ISC '09, Springer-Verlag, pp. 278–293. 2, 28, 30, 82, 94

[44] GERBER, M., AND VON SOLMS, R. From risk analysis to security requirements. *Computers and Security 20*, 7 (2001), 577 – 584. 12, 21, 22, 37, 38, 39, 40, 44, 89, 100

[45] GOOGLE. SAML Single Sign-On Service for Google Apps, 2012. Available from: https://developers.google.com/google-apps/sso/saml_reference_implementation. 23, 84

[46] GROSS, T. Security analysis of the saml single sign-on browser/artifact profile. In *Computer Security Applications Conference, 2003. Proceedings. 19th Annual* (dec. 2003), pp. 298 – 307. 2, 30, 82, 94

[47] GROSS, T., AND PFITZMANN, B. SAML artifact information flow revisited. In *In IEEE Workshop on Web Services Security (WSSS)* (Berkeley, May 2006), IEEE, pp. 84–100. 31

[48] HANSEN, M. Concepts of Privacy-Enhancing Identity Management for Privacy-Enhancing Security Technologies. In *PRISE conference proceesings: Towards privacy enhancing security technologies - the next steps* (Wien, Feb-2009 2009), J. Cas, Ed., pp. 91–103. 47

[49] HANSEN, M., SCHWARTZ, A., AND COOPER, A. Privacy and identity management. *IEEE Security and Privacy 6*, 2 (2008), 38–45. 16, 49, 56, 72

[50] HERNANDEZ-ARDIETA, J. L. *Enhancing the reliability of digital signatures as non-repudiation evidence under a holistic threat model*. Ph.D. thesis, UNIVERSITY CARLOS III OF MADRID, February 2011. 44

[51] HEVNER ALAN R., SALVATORE T. MARCH, J. P., AND RAM, S. Design science in information systems research. *Management Information Systems Quarterly 28* (2004). 4, 6, 9, 10, 23, 24, 43, 53

[52] HOGGANVIK, I., AND STØLEN, K. A graphical approach to risk identification, motivated by empirical investigations. In *Proceedings of the 9th international conference on Model Driven Engineering Languages and Systems* (Berlin, Heidelberg, 2006), MoDELS'06, Springer-Verlag, pp. 574–588. 2, 14, 93

[53] HOLZMANN, G. J. *Design and Validation of Computer Protocols*. Prentice-Hall, 1991. ISBN:0-13-539834-7. 22, 100

[54] HOUMB, S. H., AND HANSEN, K. K. Towards a uml profile for security assessment. In *In UML2003, Workshop on Critical Systems Development with UML* (2003). 67

[55] HOVORKA, D. Incommensurability and multi-paradigm grounding in design science research: Implications for creating knowledge. In *Human Benefit through the Diffusion of Information Systems Design Science Research*, J. Pries-Heje, J. Venable, D. Bunker, N. Russo, and J. DeGross, Eds., vol. 318 of *IFIP Advances in Information and Communication Technology*. Springer Boston, 2010, pp. 13–27. 10

[56] IBM COOPERATION. Overview of token types. Framework document, IBM, June 2010. http://publib.boulder.ibm.com/infocenter/wasinfo/v6r1/index.jsp?topic=/com.\ibm.websphere.express.doc/info/exp/ae/cwbs_tokentype.html. 46

[57] IBM RESEARCH. Specification of the identity mixer cryptographic library. Tech. Rep. 99740, IBM, Zurich, December 2010. 1, 16, 28, 29, 49, 56

[58] IDA, H. *A Graphical Approach to Security Risk Analysis*. Ph.D. thesis, University of Oslo, October 2007. ISSN:1501-7710. 2, 13, 26, 32, 39, 41, 43, 67, 93

[59] INDULSKA, M., AND RECKER, J. C. Design science in is research : a literature analysis. In *4th Biennial ANU Workshop on Information Systems Foundations* (Canberra, Australia, 2008), S. Gregor and S. Ho, Eds., ANU E Press. 10

[60] INFORMATION COMMISSIONER OFFICE. Privacy impact assessment handbook - version 2. Tech. rep., ICO, London, UK, June 2009. 13, 31, 39, 41, 47

[61] INITIATIVE, K. Liberty alliance, 2012. http://kantarainitiative.org/. 27

[62] INTERNATIONAL STANDARD ORGANIZATION (ISO). Iso 27005 information security risk management. Tech. rep., International Organization for Standardization, July 2008. 2, 4, 5, 12, 33, 39, 43, 77

[63] INTERNATIONAL STANDARD ORGANIZATION (ISO). Iso 24762 - information technology security technique - for identity management system. Tech. rep., International Organization for Standardization, 2011. 1

[64] ISACA. *The Risk IT Practitioner Guide*. ISACA, 3701 Algonquin Road, Suite 1010 Rolling Meadows, IL 60008 USA, 2009. isbn: 978-1-60420-116-1. 13, 15, 18, 32, 33, 41, 43, 48, 54, 55, 62, 67

[65] IWAIHARA, M., MURAKAMI, K., AHN, G.-J., AND YOSHIKAWA, M. Risk Evaluation for Personal Identity Management Based on Privacy Attribute Ontology. In *Conceptual Modelling - ER 2008*, Q. Li, S. Spaccapietra, E. Yu, and A. Oliv, Eds., vol. 5231 of *Lecture Notes in Computer Science (LNCS)*. Springer, Berlin, 134-Oct-2008 2008, pp. 183–198. Available from: http://www.springerlink.com/content/42460h2554003036/. 47

[66] JESELON, P., AND FINEBERG, A. An analysis of the traditional IS security approaches: implications for research and practice, November 2011. 39

[67] JØRGENSEN, J., AND BOSSEN, C. Requirements engineering for a pervasive health care system. In *Requirements Engineering Conference, 2003. Proceedings. 11th IEEE International* (September 2003), pp. 55 – 64. 68, 79

[68] JORGENSEN, J. B., AND LASSEN, K. B. Requirements engineering for the adviser portal bank system. In *Proceedings of the 13th Annual IEEE International Symposium and Workshop on Engineering of Computer Based Systems* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 259–268. Available from: http://portal.acm.org/citation.cfm?id=1126179.1126208. 26

[69] JØSANG, A., AND POPE, S. User centric identity management. *AusCERT Conference* (2005). 1, 26, 27, 30, 53, 70, 81, 93

[70] JØSANG, A., ZOMAI, M. A., AND SURIADI, S. Usability and privacy in identity management architectures. In *Proceedings of the fifth Australasian symposium on ACSW frontiers - Volume 68* (2007), ACSW '07, Australian Computer Society, Inc., pp. 143–152. 27

[71] JÜRJENS, J. *Secure System Development with UML*. Springer, 2005. ISBN:978-3-642-05635-2. 2, 13, 32, 39, 42, 43, 68, 93

[72] JÜRJENS, J., SCHRECK, J., AND BARTMANN, P. Model-based security analysis for mobile communications. In *Proceedings of the 30th international conference on Software engineering* (New York, NY, USA, 2008), ICSE '08, ACM, pp. 683–692. 69

[73] KARABACAK, B., AND SOGUKPINAR, I. Isram: information security risk analysis method. *Computers & Security* 24, 2 (2005), 147–159. 43

[74] KATSAROS, P., ODONTIDIS, V., AND GOUSIDOU-KOUTITA, M. Colored petri net based model checking and failure analysis for e-commerce protocols. 267–283. 32

[75] KORMANN, D., AND RUBIN, A. D. Risks of the passport single signon protocol. *Computer Networks 33* (2000), 51–58. 30

[76] KRISTENSEN, L. M., CHRISTENSEN, S., AND JENSEN, K. The practitioner's guide to coloured petri nets. *International Journal on Software Tools for Technology Transfer 2* (1998), 98–132. 77, 78

[77] KURT, J., AND LARS, K. M. *Colored Petri Nets: Modelling and Validation of Concurrent Systems: Modeling and Validation of Concurrent Systems*. Springer-Verlag Berlin Heidelberg, 2009. ISBN:978-3-642-00283-0. 7, 13, 18, 20, 21, 22, 32, 42, 43, 75, 76, 81, 84, 85, 89, 91, 93, 94, 95, 100, 101, 102

[78] KURT, J., AND LARS, K. M. *Colored Petri Nets: Modelling and Validation of Concurrent Systems: Modeling and Validation of Concurrent Systems*. Springer-Verlag Berlin Heidelberg, 2009. ISBN:978-3-642-00283-0. 39, 42

[79] KURT JENSEN, S. C., AND KRISTENSEN, L. M. *CPN Tools State Space Manual*. University of Aarhus, Department of Computer Science Aabogade 34 DK-8200 Aarhus N, Denmark, 2002. 77

[80] KURT JENSEN, S. C., AND KRISTENSEN, L. M. Cpn tools state space manual. Tech. rep., University of Aarhus, January 2006. 90, 101

[81] LANDOLL, D. J. *Security Risk Assessment Handbook*. Auerbach Publications, Taylor & Francis Group, New York, 2006. isbn:10: 0-8493-2998-1. 2

[82] LIU, L., YU, E., AND MYLOPOULOS, J. Security and privacy requirements analysis within a social setting. In *In Proc. of RE03* (2003), IEEE Press, pp. 151–161. 68

[83] LODDERSTEDT, T., BASIN, D. A., AND DOSER, J. Secureuml: A uml-based modeling language for model-driven security. In *Proceedings of the 5th International Conference on The Unified Modeling Language* (London, UK, 2002), UML '02, Springer-Verlag, pp. 426–441. Available from: http://dl.acm.org/citation.cfm?id=647246.719477. 68

[84] LORRIE CRANOR, M. L., ET AL. The platform for privacy preferences 1.0 (p3p1.0) specification, 2002. http://www.w3.org/TR/P3P/. 29, 30

[85] LUTZ, D., AND DEL CAMPO, R. Bridging the gap between privacy and security in multi-domain federations with identity tokens. In *Mobile and Ubiquitous Systems - Workshops, 2006. 3rd Annual International Conference on* (July), pp. 1–3. 47

[86] LUTZ, D. J. Secure aaa by means of identity tokens in next generation mobile environments. In *ICWMC '07: Proceedings of the Third International Conference on Wireless and Mobile Communications* (Washington, DC, USA, 2007), IEEE Computer Society, p. 57. 47

[87] MAC GREGOR, W., DUTCHER, W., AND KHAN, J. An Ontology of Identity Credentials - Part 1: Background and Formulation. Tech. rep., National Institute of Standard and Technology, Gaitersburg, MD, USA, 2006. Available from: http://www.jon.grimsgaard.no/rudrevyen/index.html. 15, 16, 17, 28, 49, 50, 56, 57, 84, 96

[88] MAHLER, T. *Legal Risk Management Developing and Evaluating Elements of a Method for Proactive Legal Analyses, With a Particular Focus on Contracts*. Monograph, The Faculty of Law, University of Oslo, Postboks 6706 St Olavs Plass, 0130 Oslo Norway, February 2010. 10

[89] MALER, E., AND REED, D. The venn of identity: Options and issues in federated identity management. *Security Privacy, IEEE 6*, 2 (march-april 2008), 16 –23. 27, 72, 83, 96

[90] MARCH, S. T., AND SMITH, G. F. Design and natural science research on information technology. *Decis. Support Syst. 15* (December 1995), 251–266. xv, 6, 9, 10, 14, 17, 18, 21, 23, 24, 53, 54, 63

[91] MASS SOLDAL LUND, BJØRNAR SOLHAUG, K. S. *Model-Driven Risk Analysis, The CORAS Approach*, 1 ed. Springer, 2011. 978-3-642-12322-1. 2, 4, 12, 32, 40, 43, 81, 93

[92] MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT). Kerberos, 2012. http://web.mit.edu/kerberos/. 27

[93] MATULEVICIUS, R., MAYER, N., AND HEYMANS, P. Alignment of misuse cases with security risk management. In *Proceedings of the 2008 Third International Conference on Availability, Reliability and Security* (Washington, DC, USA, 2008), IEEE Computer Society, pp. 1397–1404. 64, 68, 69

[94] MATULEVICIUS, R., MAYER, N., MOURATIDIS, H., DUBOIS, E., HEYMANS, P., AND GENON, N. Adapting secure tropos for security risk management in the early phases of information systems development. In *CAiSE* (2008), pp. 541–555. 37

[95] MCDERMOTT, J. P. Attack net penetration testing. In *Proceedings of the 2000 workshop on New security paradigms* (New York, NY, USA, 2000), NSPW '00, ACM, pp. 15–21. 21, 76, 77

[96] MCKENZIE ROBIN, C. M., AND WALLIS, C. Use cases for identity management in e-government. *IEEE Security and Privacy 6*, 2 (Mar. 2008), 51–57. 1, 4

[97] MICROSOFT CORPORATION. The identity metasystem: Towards a privacy-compliant solution to the challenges of digital identity. White paper, MICROSOFT CORPORATION, October 2006. 27, 28, 45, 51

[98] MURATA, T., AND KOMODA, N. Liveness analysis of sequence control specifications described in capacity designated petri net using reduction. In *Robotics and Automation. Proceedings. 1987 IEEE International Conference on* (mar 1987), vol. 4, pp. 1960 – 1965. 78

[99] MYERS M, A. D. *Qualitative Research in Information Systems*, 2 ed. SAGE Publication Ltd, The University of Auckland, New Zealand, 2002. 10

[100] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY(NIST). An introduction to computer security - the nist handbook. Tech. Rep. Special Publication 800-12, National Institute of Standards and Technology, October 1995. 71, 72

[101] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY(NIST). Electronic authentication guideline. Tech. Rep. 1.0.2, NIST Special Publication 800-63, April 2006. 83, 96

[102] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY(NIST). Guide for conducting risk assessment. Tech. Rep. Special Publication 800-30, National Institute of Standards and Technology, October 2011. 4, 33

[103] NAUMANN, I., AND HOGBEN, G. Privacy features of european eid card specifications. Tech. Rep. 1.0.1, ENISA, January 2009. 3, 14, 16, 32, 49, 56, 82, 95

[104] NORSIS. Identitetstyveri og identetssvindel, 2012. http://www.norsis.no/. 1

[105] OFFICE OF AUSTRALIAN PRIVACY COMMISSIONER. National privacy survey: Id theft, id scanning and online privacy concerns are on the rise, August 2007. 1

[106] OKUBO, T., TAGUCHI, K., AND YOSHIOKA, N. Misuse cases + assets + security goals. In *Computational Science and Engineering, 2009. CSE '09. International Conference on* (August 2009), vol. 3, pp. 424 –429. 18, 19, 67, 68, 69

[107] ORACLE CORPORATION. Using saml v2 or opensso enterprise cdsso, 2010. http://download.oracle.com/docs/cd/E19316-01/820-3740/gilfp/index.html. 28

[108] OSTERWALDER, A. *THE BUSINESS MODEL ONTOLOGY A PROPOSITION IN A DESIGN SCIENCE APPROACH*. Monograph, De l'Universite de Lausanne, 2004. 10

[109] OSTROWSKI, L., HELFERT, M., AND HOSSAIN, F. A conceptual framework for design science research. In *Perspectives in Business Informatics Research*, J. Grabis, M. Kirikova, W. Aalst, J. Mylopoulos, M. Rosemann, M. J. Shaw, and C. Szyperski, Eds., vol. 90 of *Lecture Notes in Business Information Processing*. Springer Berlin Heidelberg, 2011, pp. 345–354. 10.1007/978-3-642-24511-4_27. 10

[110] OXFORD ENGLISH DICTIONARY. Methodology, 2012. 10

[111] PAINTSIL, E. Towards legal privacy risk assessment and specification. In *Proceedings of the 8th international conference on Trust, privacy and security in digital business* (Berlin, Heidelberg, 2011), Springer-Verlag, pp. 174–185. 6, 19, 69

[112] PAINTSIL, E. Executable model-based risk assessment method for identity management systems. In *Norwegian Information Security Conference (NISK 2012)* (Akademika forlag, Postboks 2461 Sluppen, 7005 Trondheim, November 2012), Tapir Akademisk Forlag. ISBN 978-82-321-0012-5. 6

[113] PAINTSIL, E. A model for privacy and security risks analysis. In *New Technologies, Mobility and Security (NTMS), 2012 5th International Conference* (may 2012), pp. 1–8. xiii, 5, 6, 11, 18, 19, 21, 22, 24, 25, 31, 35, 82, 89, 92, 94, 100, 103

[114] PAINTSIL, E. Taxonomy of security risk assessment approaches for researchers. In *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)* (nov. 2012), IEEE, pp. 257 –262. xiii, 5, 6, 7, 11, 12, 25, 35

[115] PAINTSIL, E. Evaluation of privacy and security risks analysis construct for identity management systems. *Systems Journal, IEEE 7*, 2 (2013), 189–198. xiii, 5, 6, 11, 14, 17, 21, 23, 24, 25, 35, 70, 82, 83, 95

[116] PAINTSIL, E. Towards automation of risk analysis in identity management systems. Accepted for publication in Trustcom 2013 proceedings, July 2013. 5, 6, 11, 23, 24, 25, 35

[117] PAINTSIL, E., AND FRITSCH, L. A taxonomy of privacy and security risks contributing factors. In *Privacy and Identity Management for Life*, vol. 352 of *IFIP Advances in Information and Communication Technology*. Springer Boston, 2011, pp. 52–63. xv, 3, 5, 6, 11, 14, 15, 23, 24, 25, 35, 53, 54, 58, 60, 62, 64, 70

[118] PAINTSIL, E., AND FRITSCH, L. Executable model-based risk analysis method for identity management systems:using hierarchical colored petri nets. ISBN 978-3-642-40342-2. 5, 6, 11, 23, 24, 25, 35

[119] PAINTSIL, E., AND FRITSCH, L. Executable model-based risk assessment method for identity management systems. 6

[120] PANUSUWAN, V., AND BATLAGUNDU, P. Privacy risk assessment case studies in support of square. Tech. Rep. CMU/SEI-2009-SR-017, Software Engineering Institute, July 2009. 43

[121] PARTY, A. . D. P. W. Opinion 4/2007 on the concept of personal data. Tech. rep., ARTICLE 29 DATA PROTECTION WORKING PARTY, June 2007. 74

[122] PEFFERS, K., TUUNANEN, T., ET AL. A design science research methodology for information systems research. *J. Manage. Inf. Syst. 24*, 3 (Dec. 2007), 45–77. 9, 10, 11, 17

[123] PETERSON, G. Introduction to Identity Management Risk Metrics. *IEEE Security & Privacy 4*, 4 (Jul-2006 2006), 88–91. 15, 33, 47, 48, 54, 55

[124] PFITZMANN, A., AND HANSEN, M. Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management A Consolidated Proposal for Terminology - v0.29. http://dud.inf.tu-dresden.de/Anon_Terminology.shtml, July 2007. 47

[125] PHILIP L. CAMPBELL, J. E. S. A classification scheme for risk assessment methods. Tech. Rep. SAND2004-4233, SANDIA National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, August 2004. 13, 24, 38, 41, 42

[126] PHILIP S. ANTON, ROBERT H. ANDERSON, R. M., AND SCHEIERN, M. The vnerability assessment & mitigation methodology. Tech. rep., RAND, 2003. ISBN 0-8330-3434-0. 24

[127] PINUELA, A. Competitiveness and innovation framework programme ict policy support programme (ict psp). Tech. Rep. ICT-PSP/2007/1, STORK-eID Consortium, November 2008. 3, 14

[128] RATHA, N. K., CONNELL, J. H., AND BOLLE, R. M. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal 40*, 3 (2001), 614–634. 83, 96

[129] RECORDON, D., AND REED, D. Openid 2.0: a platform for user-centric identity management. In *Proceedings of the second ACM workshop on Digital identity management* (2006), ACM, pp. 11–16. xiii, 1, 23, 28, 30, 97

[130] REDWINE, S. Introduction to modeling tools for software security, 2005-2012. Available from: `https://buildsecurityin.us-cert.gov/bsi/articles/tools/modeling/698-BSI.html`. 43

[131] ROST, M., AND BOCK, K. Privacy by design and the new protection goals*. *DuD* (January 2011). 18, 24, 63, 64, 68, 69, 70, 72, 75

[132] SALIM, F., REID, J., DULLECK, U., AND DAWSON, E. Towards a game theoretic authorisation model. In *Proceedings of the First international conference on Decision and game theory for security* (Berlin, Heidelberg, 2010), GameSec'10, Springer-Verlag, pp. 208–219. 43

[133] SAMUEL-OJO, O., SHIMABUKURO, D., ET AL. Meta-analysis of design science research within the is community: Trends, patterns, and outcomes. In *Global Perspectives on Design Science Research*, R. Winter, J. Zhao, and S. Aier, Eds., vol. 6105 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2010, pp. 124–138. 9

[134] SERGIO, S. *An introduction in Science and Technology Studies*, 2 ed. Blackwell Publishing, 2010. 10

[135] SHABALIN, P. Model checking umlsec. Tech. rep., University of Munchen, August 2004. 71

[136] SIEMENS ENTERPRISE. CRAMM, 2011. Available from: http://www.cramm.com/. 43

[137] SIMON, H. A. *The Sciences of the Artificial*, first ed. MIT Press, Cambridge, Massachusetts, 1969. 3, 9

[138] SINDRE, G., AND OPDAHL, A. L. Eliciting security requirements with misuse cases. *Requirements Engineering 10*, 1 (2004), 34–44. 18, 19, 67, 68, 69

[139] SIPONEN, M. T. An analysis of the traditional is security approaches: implications for research and practice. *Eur. J. Inf. Syst. 14*, 3 (November September), 303–315. 38

[140] SIPONEN, M. T., AND OINAS-KUKKONEN, H. A review of information security issues and respective research contributions. *SIGMIS Database 38*, 1 (2007), 60–80. 50

[141] SKJEGSTAD, M., JOHNSEN, F., AND NORDMOEN, J. An emulated test framework for service discovery and manet research based on ns-3. In *New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on* (may 2012), pp. 1 –5. 14, 42

[142] SMOJVER, S. Selection of information security risk management method using analytic hierarchy process (ahp). *Central European Conference on Information and Intelligent Systems, CECIIS – 2011* (2011). 37

[143] SOLOVE, D. A taxonomy of privacy - GWU Law School Public Law Research Paper No.129. *University of Pennsylvania Law Review 154*, 3 (Jan. 2006 2006), 477. 31, 47, 54

[144] SOVIS, P., KOHLAR, F., AND SCHWENK, J. Security analysis of openid. In *Sicherheit* (2010), vol. 170 of *LNI*, GI, pp. 329–340. ISBN:978-3-88579-264-2. 30, 31

[145] SREENIVAS, R., AND KROGH, B. On fairness and conflicts in petri nets. In *Circuits and Systems, 1989., Proceedings of the 32nd Midwest Symposium on* (aug 1989), pp. 406–409 vol.1. 78

[146] STEFAN TAUBENBERGER, JAN JÜRJENS, Y. Y., AND NUSEIBEH, B. Problem analysis of traditional it-security risk assessment methods - an experience report from the insurance and auditing domain. In *SEC* (2011), pp. 259–270. 12, 32, 37, 38, 39, 40, 41, 44

[147] STEIN, T., CHEN, E., AND MANGLA, K. Facebook immune system. In *Proceedings of the 4th Workshop on Social Network Systems* (New York, NY, USA, 2011), SNS '11, ACM, pp. 8:1–8:8. 1

[148] STEPHENSON, P. R. A formal model for information risk analysis using colored petri nets. In *Proceedings of the Fifth Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, Aarhus, Denmark, October 8-11, 2004, DAIMI PB - 570 / Kurt Jensen (Ed.)* (Oct, 2004), pp. 167–184. 13, 42, 43

[149] STEWART, A. On risk: perception and direction. *Computers & Security* (2004), 362–370. 12, 39, 41, 44

[150] STRINGER, E. *Action Research in Education (2nd Edition)*, 2 ed. Prentice Hall, July 2007. 10

[151] SUH, B., AND HAN, I. The is risk analysis based on a business model. *Information & Management 41*, 2 (2003), 149–158. 43

[152] SUNYAEV, A., TREMMEL, F., ET AL. A reclassification of is security analysis approaches. In *AMCIS* (2009), p. 570. 13, 38, 39, 41

[153] SURIADI, S., FOO, E., AND DU, R. Layered identity infrastructure model for identity meta systems. In *Proceedings of the sixth Australasian conference on Information security - Volume 81* (Darlinghurst, Australia, Australia, 2008), Australian Computer Society, Inc., pp. 83–92. 28

[154] SURIADI, S., FOO, E., AND JØSANG, A. A user-centric federated single sign-on system. *J. Netw. Comput. Appl. 32* (March 2009), 388–401. 31, 43, 82, 94

[155] THE ASSOCIATED PRESS. Study: Id theft spurs web privacy concerns, 2008. 1

[156] THE OPEN WEB APPLICATION SECURITY PROJECT (OWASP). Threat risk modeling, 2012. https://www.owasp.org/index.php/Threat_Risk_Modeling. 4, 13, 41

[157] TIPLEA, F., AND BOCANEALA, C. Decidability results for soundness criteria of resource-constrained workflow nets. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 42*, 1 (jan. 2012), 238 –249. 77

[158] TRACY MITRANO, D. R. K., AND MALTZ, L. What does privacy have to do with it?: privacy risk assessment. *EDUCAUSE* (04 2005). 19, 68, 69

[159] TUROFF, M., AND LINSTONE, H. A. *The Delphi Method - Techniques and Applications*. http://is.njit.edu/pubs/delphibook/, New Jersey Institute of Technology and Portland State University, 2002. 17, 23, 54, 57

[160] VAN LAMSWEERDE, A. Elaborating security requirements by construction of intentional anti-models. *Software Engineering, International Conference on 0* (2004), 148–157. 68

[161] VORSTER, A., AND LABUSCHAGNE, L. A framework for comparing different information security risk analysis methodologies. In *Proceedings of the 2005 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries* (Republic of South Africa, 2005), SAICSIT '05, South African Institute for Computer Scientists and Information Technologists, pp. 95–103. 2, 13, 37, 38, 41

[162] WANG, A. J. A. Information security models and metrics. In *Proceedings of the 43rd annual Southeast regional conference - Volume 2* (New York, NY, USA, 2005), ACM, pp. 178–184. 41, 43

[163] WANG, L., ISLAM, T., AND OTHER. An attack graph-based probabilistic security metric. In *Proceeedings of the 22nd annual IFIP WG 11.3 working conference on Data and Applications Security* (2008), Springer-Verlag, pp. 283–296. 43

[164] WANG, R., AND DAGLI, C. H. Executable system architecting using systems modeling language in conjunction with colored petri nets in a model-driven systems development process. *Syst. Eng. 14*, 4 (Dec. 2011), 383–409. 32, 95

[165] WESTERGAARD, M., AND KRISTENSEN, L. The access/cpn framework: A tool for interacting with the cpn tools simulator. In *Applications and Theory of Petri Nets*, G. Franceschinis and K. Wolf, Eds., vol. 5606 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2009, pp. 313–322. 95

[166] WINKELVOS, T., RUDOLPH, C., AND REPP, J. A property based security risk analysis through weighted simulation. In *Information Security South Africa (ISSA), 2011* (august 2011), pp. 1–8. 2, 4, 12, 31, 32, 41, 43, 82, 94

[167] WOODCOCK, J., AND DAVIES, J. *Using Z: specification, refinement, and proof*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996. 4, 13, 18, 39, 41

[168] WP2. D 2.1: Inventory of topics and clusters. Deliverable 2.0, Future of Identity in the Information Society, September 2005. 47

[169] WP3. D3.1: Structured overview on prototypes and concepts of identity management systems. Deliverable 1.1, Future of Identity in the Information Society, September 2005. 1, 27, 28, 29, 46, 47, 72, 83, 96

[170] WP3. D3.17: Identity management systems - recent development. Report D3.17, Future of Identity in the Information Society, August 2009. 3, 14, 28, 29, 54

[171] XU, D., AND NYGARD, K. A threat-driven approach to modeling and verifying secure software. In *Proceedings of the 20th IEEE/ACM international Conference on Automated software engineering* (New York, NY, USA, 2005), ASE '05, ACM, pp. 342–346. 79

[172] XU, J., AND KUUSELA, J. Analyzing the execution architecture of mobile phone software with colored petri nets. *International Journal on Software Tools for Technology Transfer (STTT) 2* (1998), 133–143. 10.1007/s100090050022. 95

[173] YACOUB, S. M., AND AMMAR, H. H. A methodology for architecture-level reliability risk analysis. *IEEE Trans. Software Eng. 28*, 6 (2002), 529–547. 42, 43

[174] YAMADA ET AL. Information security incident survey report. Tech. rep., NPO Japan Network Security Association (JNSA), 2006. 54, 90, 101

[175] YANG, Y., TAN, Q., AND XIAO, Y. Verifying web services composition based on hierarchical colored petri nets. In *Proceedings of the first international workshop on Interoperability of heterogeneous information systems* (New York, NY, USA, 2005), ACM, pp. 47–54. 77, 78, 79

[176] ZHOU, S., QIN, Z., ZHANG, F., ZHANG, X., CHEN, W., AND LIU, J. Colored petri net based attack modeling. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, vol. 2639 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2003, pp. 583–583. 76

[177] ZWINGELBERG, H., AND HANSEN, M. Privacy protection goals and their implications for eid systems. In *Privacy and Identity Management for Life*, J. Camenisch, B. Crispo, S. Fischer-Hbner, R. Leenes, and G. Russello, Eds., vol. 375 of *IFIP Advances in Information and Communication Technology*. Springer Berlin Heidelberg, 2012, pp. 245–260. 2, 18, 31