NTNU – Trondheim
Norwegian University of
Science and Technology

Comparison of i*-based and Use Case-based Security Modelling Initiatives for Software Requirements Engineering

An empirical comparison of Secure Tropos and Misuse Cases

Yushan Pan

Master in Information Systems
Submission date: June 2012
Supervisor: Guttorm Sindre, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

NTNU
Det skapende universitet

Comparison of i*-based and Use Case-based Security Modeling Initiatives for Software Requirements Engineering

An empirical comparison of Secure Tropos and Misuse Cases

Yushan Pan

Master of Science in Information Systems Engineering
Submission date: June 10, 2012
Supervisor: Guttorm Sindre
Co-supervisors: Peter Karpati, Olawande Dararmola

Norwegian University of Science and Technology
Department of Computer and Information Science

Problem description

In the specialization project (TDT4501) prior to this thesis, according to a previously established framework, all in all, eight modeling techniques were investigated through a synthesis of available literature, using published papers about and surveys of the relevant modeling techniques. The results illustrated that both of these two categories modeling initiatives have their advantages and disadvantages.

Purely analytical evaluations of modeling approaches will not always reflect their practical usefulness. Hence it is important to complement such evaluations with empirical investigations, to see if advantages claimed analytically also come true in practice. Thus, the objective of this master thesis project is to make an experimental comparison. Since such comparisons are time-consuming it is not realistic to try out all the eight techniques that were investigated analytically. Hence, two techniques will be chosen Secure Tropos and Misuse Cases. These shall be compared in a controlled experiment to investigate:

1. The participants' performance.
2. The participants' preference for the two techniques.

Based on the results we will analyze whether the advantages and disadvantages claimed from the analytical evaluation also were indicated in the experiment. It shall also be discussed whether the analytical and experimental evaluation together presents some ideas for improvement of the two modeling techniques.

Assignment given: 15th. January 2012

Supervisor: Guttorm Sindre

Abstract

In the course TDT4501 - Specialization Project - "ReqSec project", the preparatory course to this thesis, through purely analytical evaluation of the eight modeling approaches, the advantages and disadvantages were illustrated based on the categories - i*-based modeling approach and Use Case-based modeling approach.

However, only a purely analytical evaluation of the modeling approaches does not always reflect their practical usefulness. Hence, the **[motivation]** of the thesis was selecting two modeling approaches, those are Secure Tropos and Misuse Cases, using an empirical investigation for such evaluations to guide the researchers and practitioners a better overview and understanding of the benefits of the two modeling approaches in a real life usage. The objective was to see if the advantages claimed analytically in the previous project also come true in practice. **[Questions]** Through a controlled experiment, two core problems shall be investigated: a) How about the participants' performance when they applied the two modeling approaches to finish tasks in the experiment and b) Their preference for the two modeling approaches after the experiment. The **[principle]** was using two modeling approaches to perform the experiment, through the participants' performance on the identified number of threats and mitigations for the experiment cases, and their perception of the two modeling approaches by means of asking them to estimate the usage of modeling diagrams, textual description of cases, and memory in the experiment. And combining with the evaluation of post-questionnaire analysis, the conclusions were summarized based on the empirical study of statistical results and the previous analytical study results, to investigate whether the empirical evaluation could match well with analytical evaluation or not.

[Contribution] The experiment project was the first time to compare the Secure Tropos and Misuse Cases comprehensibly. The results illustrated that both modeling techniques had no significant difference of identifying threats but they had significant difference of identifying mitigations in this controlled experiment with 50 students who apply to both modeling approaches with relevant cases. And through analyzing the same case with the same modeling approach or different modeling approach of the experiment, it was found that Net Shopping case was identified more mitigations and threats by the participants when considering the aspect of technique criteria of threats and mitigations. The participants were complementary regarding goal-based modeling approach in some security issues and performed non-techniques threats and mitigations in this controlled experiment. Hence, Secure Tropos was investigated perceiving more favorable. In the last, comparing with the six dimensions from previous analytical comparison, the investigation shows that most of the two modeling approaches' advantages were confirmed, and the results also coincided to the previous analytical evaluation.

Keywords: Secure Tropos, Misuse Case, Empirical Study, Security Modeling

Preface

This master thesis was written as part of my MSc degree at Norwegian University of Science and Technology (NTNU), Department of Computer and Information Science, spring 2012. It extends the work done in the preliminary project - modeling techniques for security requirements as a part of "ReqSec" Project carried out by the same author in fall 2011.

Firstly, I must offer my profoundest gratitude to my supervisor Professor Guttorm Sindre for his encouragement and advising during the thesis. His guidance and assistance in this thesis has been invaluable. I also want to offer my special thanks to two Post-Doctor Researchers - Peter Karpati and Olawande Daramola for their very helpful guidance during my entire master study at NTNU.

Secondly, I wish to thank Professor Tor Stålhane, for his helpful guidance of the knowledge about Software Engineering Experiment and Requirements Engineering. And I would like to thank those students who participated in the experiment. I would also like to thank Hallvard Andreas Eriksen, Thomas Hoberg Burnett, Amanpreet Kaur and Zheng Wang for their helpful proof-reading and correction of grammar errors after I finished this thesis.

Finally, I would like to thank my friends and families, who have always supported me.

Yushan Pan

June 1, 2012

Acronyms

ST Secure Tropos Modeling approach

MUC Misuse Cases Modeling approach

MUCM Misuse Case Maps

MUSD Misuse Sequence Diagrams

SQUARE Security Quality Requirements Engineering Methodology

XP Extreme Programming

UC Use Case Diagrams

SD Strategic Dependency

SR Strategic Rationale

SRE Security Requirement Engineering

EAS Elicitation and Analysis

SRs Security Requirements

S Specification of SRs

DS Documentation of SRs

IMA Identification and Modeling of Assets

VT Vulnerabilities and Threats

VVS Verification and Validation Support

UML Unified Modeling Language

SUC Security Use Case

IT Information Technology

TAM Technology Acceptance Model

HID Health Insurance Department

DHI Department of Health Insurance

NT Net Shopping Case

HIS Health Insurance System

KNOW_UC Knowledge of Use Cases

KONW_MUC Knowledge of Misuse Cases

KNOW_IST Knowledge of i* Modeling

KNOW_ST Knowledge of Secure Tropos

KNOW_MOD Knowledge of System Modeling

THR Threat

MIT Mitigation

THMI Threat and Mitigation

PU Perceived Usefulness

ITU Intention to Use

PEOU Perceived Ease of Use

ICT Information and Communications Technology

USD UML Sequence Diagrams

TDD Test-driven Development

RCIS'12 Sixth International Conference on Research Challenges in Information Science, 2012

ReqSec Requirements for Secure Information Systems

ESMUS Enhanced Sustained Use Monitoring System

IDI Institutt for datateknikk og informasjonsvitenskap (Department of Computer and Information Science)

NTNU Norges teknisk-naturvitenskapelige universitet (Norwegian University of Science and Technology)

UiB Universitetet i Bergen (University of Bergen)

SINTEF Stiftelsen for industriell og teknisk forskning

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter gives an introduction to the thesis. Firstly, the problem definition and the motivation behind exploring the problem were illustrated. And then come up with the problem's context. At the end of the chapter there was an outline of the rest of the thesis.

1.1 Problem description

Security is becoming more and more important in software systems in recent years. Hence many techniques and methods were developed to support the secure systems engineering. However, there is no clear guideline on what is the difference between these modeling techniques: which one has what advantages and disadvantages, and why it is better than other techniques in a specific scenario. Therefore, in order to realize the research purpose, researchers and practitioners might struggle with the characterization of these modeling approaches and tries to use each modeling approach in their research in order to figure out the best and suitable one for their goals due to the fact that existing modeling approaches have some similarities.

Hence, it is very important and useful to address clearly each modeling approach and to help the researcher and practitioners solve this situation when they apply those relevant modeling approaches in their objective cases. In the previous analytical comparison[54], eight modeling approaches with disadvantages and advantages were presented according to the classification framework for these modeling approaches[40]. However, whether these findings claimed the same ideas in practice were still uncertain issues for research. For instance, the previous work expressed six dimensions of characterization that were based on purely analytical comparison[53]; it was lacking investigation of their practical usefulness. Therefore, the two important modeling approaches (Secure Tropos & Misuse Cases) were chosen to perform an experiment to see if the previous claim analytically also come true in practice. The results of this empirical evaluation can provide practitioners and

researcher a reference and a guideline of the two modeling approaches in case of avoiding wasting time to choose relative approach in a really complex system analysis and design.

1.2 Research motivation

1.2.1 Purpose

The purpose of this master thesis project was to figure out whether the pervious analytical comparison claim was also indicated in this empirical comparison. According to the experiment results and its findings, a discussion based on the analytical evaluation can help to conclude whether the advantages of the two modeling approaches were realized in practice and whether the analytical and experimental evaluation together presents some ideas for improvement of the two modeling techniques.

1.2.2 Objective

The objective of this master thesis project was to make an experimental comparison. Through the empirical comparison can be investigated whether the claimed findings and conclusions coincided to the previous work by the way of investigating the participants' performance and preference for the two modeling approaches. Thus, the thesis gave a discussion of the ideas about the improvement of the two modeling approaches based on the view of two modeling approaches' advantages and disadvantages that claimed in the analytical study and the empirical study.

1.3 Context

The objective method contains two well-known modeling approaches - Secure Tropos and Misuse Case, with the focus on two special cases. Secure Tropos is normally good for goal-based analysis and easily to find security concerns between each goal and soft-goal, but it is also believed that the Misuse Case has a good ability to identify malicious act against a system. Hence, there was an assumption that the two modeling approaches had equal preference during the experiment as a condition at the beginning of the experiment design.

This thesis has been carried out with the supervision of Professor Guttorm Sindre, at the Department of Computer and Information Science at the Norwegian University of Science and Technology (NTNU). The project was part of "ReqSec" project [12] that was lead by Professor Guttorm Sindre of NTNU and Professor Andre Opdahl of UiB. The Research Council of Norway finances the project from 2008-2012.

1.4 Outline

The remainder of this thesis was organized in the following parts:

Preliminary Studies and Representation This contains background information of previous published papers and, state of art and the evaluation of approach before conducting the experiment.

Experiment This contains the experiment, its objective, execution and aftermath.

Discussion This contains the findings after execution of the experiment. And combining of the previous analytical comparison and the empirical comparison, the ideas of improvement for these two modeling approaches were presented regarding the results of empirical study.

Conclusion This contains concluding remarks and points for future work.

Reference and Appendices This contains the reference used in the thesis, the documents used in the experiment and the results. Also, a snapshot of published paper on an international conference was attached at the end of the thesis.

Chapter 2

Preliminary Studies

Even though most of the preliminary studies were conducted in the TDT4501 Specialization Project[54], but the most essential for understanding the content of the thesis was repeated and expanded in chapter 2 and chapter 3. This chapter was the result of the literature study performed at the start of the project. It explained the general principles of requirements engineering process, characteristics of the requirements, security requirements engineering and some relevant empirical studies of modeling approaches for security requirements engineering.

2.1 Requirements Engineering

2.1.1 Requirements Engineering Process

Software requirements engineering is the process of determining the features of the system and other requirements from the customers. A requirements engineering process used for requirements engineering vary widely depending on the application domain, the people involved and the organization developing the requirements [7]. This process is regarded as one of the most important issue of a software development because through this method the software development team can decide precisely what should be build and what will be built [7].

In general, this process is run as the stakeholders of the software to determine functional and non-functional requirements, a close interaction between developers and end-users, and finally expected software will be developed. In the entire development cycle, however, most people designed software is not considering security issues before the phase of determining requirements or they always secure the security issues of the software after it has been built. Thus, in some cases, software can be treated as a non-successful product when it delivered to the end-users, such as mobile bank system, online transaction system and so forth.

2.1.2 Characteristics of Requirements

After the general concepts of the requirements process was clear, there might be a question about what are the requirements must be. In general software development, the requirements must be [2]:

- correct, complete, consistent, non-ambiguous, verifiable and traceable.

1. Correct[2]: A requirement shall be correct if it describes something that a system must match or a constraint on the way it must be done .
2. Complete[2]: A requirement is complete to the extent that all parts are present and each part is fully developed. The requirement set is complete if it contains all of the complete requirements.
3. Consistent[2]: A requirement is consistent if it does not conflict with another requirement.
4. Non-ambiguous[2]: A requirement is non-ambiguous if there is only one interpretation of its meaning.
5. Verifiable[2]: A requirement is verifiable if and only if there is a finite cost effective process that a person or machine can use to check that the as built system meets the requirement.
6. Traceable[2]: An analysis requirement is traceable if it can be traced backward to a feasibility study, a white paper, meeting notes, or an interview. A design requirement is traceable if it can be traced backward to one or more essential requirements.

2.2 Security Requirements Engineering

Software security requirements are defined as [23]:

"A security requirement is complementary to the functional requirement of a system. It is a manifestation of a high-level organizational policy into the detailed requirements of a specific system. Security requirements should be based on an analysis of the assets and services to be protected and the security threats from which these assets and services should be protected."

Functional security requirement[32] is a capability or condition need in a system to control or limit the fulfillment of requirements. And non-functional security requirement[32] is a property of a system required to ensure fulfillment of requirements in the face of misuse or abuse.

This approach covers a large scope of security requirements in[21]. For instance, the quality properties of requirements are: *design independent, unambiguous, precise, understandable, traceable, verifiable, prioritized, complete, consistent, organized, and modifiable.*

Firesmith defined that these quality properties of requirements can be categorized into *identification, integrity and privacy requirement*[32]. For example, bank system users must authenticate their password is a security requirement, and the bank system shall not allow unauthorized individuals access to any communications is a privacy requirement. There are several approaches to security requirements engineering handle requirement phase tasks[62]. For instance, SQUARE, Charles Haley and his colleagues[23], Gustav Boström and his colleagues[24], CLASP, Microsoft[48][57][44], Axelle Apvrille and Makan Pourzandi[15], Eduardo Fernandez[30], Kenneth van Wyk and Gray McGraw[63], and Gunnar Peterson[56]. All these approaches are focusing on specific issues. For example, SQUARE is based on interaction between requirement engineers and the project stakeholders. Boström and his colleagues suggest security requirements in agile development with a focus on XP practices. Eduardo Fernandez and Gunner Peterson suggest use cases as a starting point for identifying security requirements[30][56]. Kenneth van Wyk and Gray McGraw prefer to use abuse cases to identify security requirements[63].

The MUC is an security requirements artifacts that can be used for eliciting security requirements[59]. The approach is an extension of regular UML use case diagrams with MUC that specify activity not wanted in the system. Secure Tropos is founded on the i* modeling for agent-oriented software development. It can be used for modeling the security aspects of systems[66]. Other security requirements artifacts were also discussed in[62], such as Abuser stories, attack trees and threat trees, soft-goal interdependency graphs.

2.3 Modeling Techniques for Security Requirements and Relevant Empirical Studies

Modeling techniques for the system development is an important vital ingredient (Bray 2002)[16]. System models are a very important bridge between the system analysis and system design process. Combining different modeling techniques can also considerably enrich system models and the system development. The system modeling techniques for the requirements engineering can be divided into two half. One is based on internal model where the process oriented, data oriented and process/data combination modeling techniques are used to express the systems interaction functions, interaction objects and dataflow between the functions[16]. And another one is based on external model where the models are used for the system appearance and system behavior.

From the meaning of aspect of the model's representation, if the representational modeling provide snapshots of the system appearance, then the behavioral modeling really defines the relationship between the input and output from the system. For instance, the Use Case modeling technique is used for presenting what the system does and what the relationship between the input and output of a system in the stage of functional requirements desire(Bray 2002)[16].

Some relative works were done to compare the MUSD, MUCM, Mal-activity diagrams, textual Use Cases for the empirical studies. Karpati and his colleagues performed an

experiment to an visualizing cyber attacks with MUCM[41]. The experiment involved the consideration of architectural context, using architecture with MUCM to investigate the MUCM appearance; they found that MUCM can help the less experienced stakeholders to gain an understanding of intrusion cases easily.

Vikash Katta et al [64] compared Misuse Sequence Diagrams (MUSD) and Misuse Case Maps (MUCM) for intrusion visualization. They investigated MUSD and MUCM were perform equally well in a controlled experiment, and the experiment participants were prefer to use MUSD when they were complementary regarding architectural issues and sequences of actions.

In 2010, an experimental evaluation of MUCM for eliciting SR were presented by Karpati et al[55]. Through a controlled experiment they investigated that the participants significantly better understanding of the intrusion and ability to identify mitigations when they applied the MUCM. MUCM was better performed by the participants than those who applied to MUC with a system architecture diagrams to understand the intrusion. However, both two modeling techniques were not perceived by their participants for the ease of use.

The empirical study of Misuse Case and Mal-activity Diagrams for modeling social engineering attacks in 2011[39] solved one question that whether Mal-activity diagrams was really more effective than Misuse Case for understanding social engineering attacks and finding prevention measures. The results shows that both modeling techniques were perform equally well, only one small difference was the participants perceived mal-activity was easy to use.

Stålhane et al[61] presented a controlled experiment to compare safety hazard identification by means of Misuse Cases based on Use Case Diagrams and Textual Use Cases. Although they explained that the Textual Use Cases were able to identify more failure modes or threats in most cases due to the reason that the use case encourages analysts to focus on threats related to the use case function of the system, they were still find Use Case Diagrams and Textual Use Cases were equally easy to use from their experiment.

In the comparison of attack trees and misuse cases for security threat identification[14], Opdahl et al cited that the attack trees have more effective for identify threats, especially when there was no use case diagram was pre-drawn for the system to analyze. But the investigation shows the participants had similar perception for the two techniques. The researchers also suggested a way of threats categorization, but they did not mention how they classified the threats into the sub-categories. So categorizing of the threats or mitigations is still a tough task in the experiment until now, since the responses from the participants might ambiguous in most cases.

As mentioned above, all of these empirical studies of security modeling were just cover the Use Case-based modeling approaches but did not touch any i*- based modeling approaches. Thus, it was worthy to compare i*-based modeling approaches with the Use Case-based modeling approaches. Hence, the rest of the project was focus on designing an experiment to compare ST with MUC in order to cover i*-based modeling techniques and the Use Case-based modeling techniques in empirical study of the security requirement modeling techniques field.

Chapter 3

Overview the Knowledge of i*-based and Use Case-based modeling Techniques

This section presents the characterization of i*-based and Use Case-based initiatives based on a framework[40]. The framework was a product of a depth study of the literature review about those were relevant for security modeling characterization in recent years[53]. The rest section expressed the concepts of two modeling categories as it was presented already in the previous project work[54] and the conference paper[53]. However, the purpose of presenting here again was to review the fundamental concepts of relevant modeling techniques before the empirical study of the two modeling approaches.

3.1 Goal-oriented Modeling Technique

In requirements engineering, a goal-oriented modeling approach has been recognized to be useful on the design of information system, especially the internet applications and web-based systems. Generally, goals present the objectives that the system should achieve through the cooperation of actors in the system environment and in the system-to-be[25]. The goal-oriented modeling can capture "why" data and functions of a system, and whether they are suggest for achieving the high-level objectives that could be rise in a requirements engineering process naturally[25]. The goal-oriented modeling approach provides a criterion for requirements completeness, i.e., the requirements can be judged as complete if they are sufficient to express the objectives that they are refining.

3.2 i*-based Modeling Techniques

The i*-based modeling approach was developed for modeling and reasoning about organizational environments and their information systems[67]. The i* modeling covers both actor-oriented and goal-oriented modeling techniques. It answers the question about "why and who" in a system, but not including "what" in a system. It is suitable for an early phase of system modeling in order to understand the problem domain[68]. Hence i* modeling approach inherits the goal-oriented modeling approach features as well. There are two main components of the i* modeling: 1) the strategic dependency (SD) model is used to describe the dependency relationships among various actors in an organizational context[67]. 2) The strategic rationale (SR) model is used to describe stakeholder interests and concerns, and how they might be addressed by various configurations of systems and environments[67]. The i*-based modeling techniques involve several modeling approaches such as Secure Tropos (ST), Enhanced Secure Tropos (EST) and Secure i* . Since just ST was selected for the experiment, thus its advantages and disadvantages based on the previous work[54] were presented for later use in Chapter 10.

3.2.1 Secure Tropos modeling approach

Secure Tropos

Secure Tropos is based on the i*-based modeling framework for agent-oriented software development. Secure Tropos is an extension of the Tropos methodology[34] that incorporates concepts for modeling the security aspects of systems. The key additional concepts of Secure Tropos are:

1. **Security Constraint:** This is a restriction that is related to security issues, such as privacy, integrity and availability, which can influence the analysis and design of a multi-agent system being developed by Mouratidis [51].

2. **Secure Dependency:** This presents one or more security constraints that must be fulfilled for a dependency to be satisfied[34].

3. **Secure Entity:** This represents any secure goal/task/resource of the system.

Therefore, the Secure Tropos process [46] also extends the Tropos process with phases to analyze and model the new concepts. These activities produce different kinds of diagrams, which are used as input to the later activities. The activities are: secure reference modeling, security constraint modeling, secure capabilities modeling[26].

Figure 3.2 illustrates the notations[50] of the ST and an example[8] of ST diagram in figure 3.1.

Actor entities that have strategic goals and intentionality.

Goal an actor's strategic interests.

Soft-goals goals without clear criteria whether they are satisfied or not.

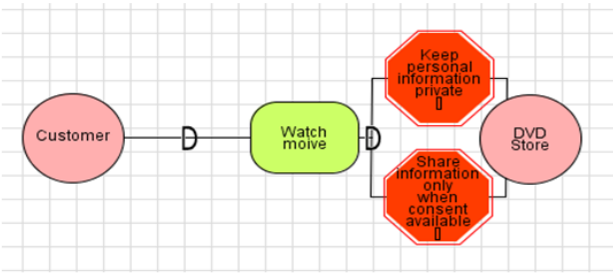

Figure 3.1: Full dependency links with a security constraint

Task represents at an abstract level, a way of doing something.

Resource represents a physical or information entity.

Dependencies indicate that one actor depends on another in order to attain some goals, execute some tasks, or deliver a resource.

Figure 3.1 shows the full dependency link with a security constraint. Customer requires DVD from the DVD store. Both customer and DVD store are actors in this case. Watch movie is a secure goal. Customer is a depender with constraints to the dependee DVD store. DVD store has the responsibility to keep the customer's personal information private and just could share the information only when consent is available.

3.2.2 Advantages of Secure Tropos

Representation perspective and level of abstraction Secure Tropos is typically goal-based approach. The objective of Secure Tropos is to satisfy the security goals of a system despite security constraints. Certainly, the security goals can be high level in many cases[27]. Hence, this modeling technique can involve multiple stakeholders in a system development, such as the project managers and end-users or those non-technique professionals. This method would be easier for stakeholders to seek the security concerns in an IT project.

Kind of security requirement engineering (SRE) tasks/activities There are seven sub-classifications for SRE tasks/activities[40], *security objectives, identification and modeling of assets, identification of vulnerabilities and threats, elicitation and analysis of SRs, Specification of SRs, documentation of SRs and verification and validation support.* Secure Tropos can fulfill with security objectives, elicitation and analysis of SRs, specification of SRs. It can do the high-level goals analysis and sub-goals analysis that might be most important for customers, also have the ability to derive security requirements from the other sources such as legislative request about security, to analyze the security requirements to achieve complete and unambiguous requirements[53].

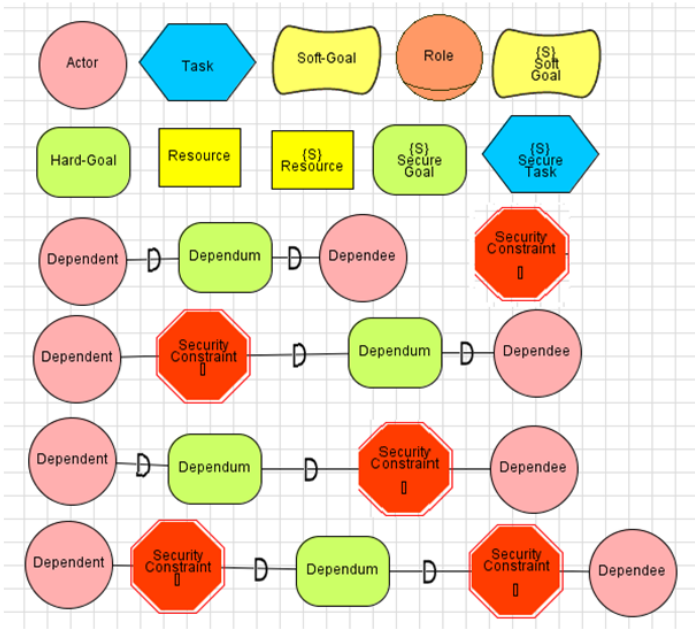

Figure 3.2: Tropos, Secure Tropos notation and Tropos concepts graphically

Technical criteria and specification criteria Secure Tropos have good "internal verification support" except the "validateable" and "complete" specification criteria where the ST just partially support these two criteria[53]. And for the other 5 criteria, e.g., *requirements reuse, support for other development stages, help support and easy to use,* ST modeling technique has good ability on these specification criteria as well.

Modeling language, process and method ST modeling technique has good ability to "formulate basic security requirements", "usage scenarios", "represent security mechanisms" and low level security requirements[53]. There is a tool support to draw the ST models[8]. It also has similarity with software specification languages and can be used for the stages of software development. Most reusable artifacts can be used during the early and late requirements[53].

Relevant SRE notation and software evolution and other perspectives ST covers the basic notions of security requirements modeling, such as goal, soft-goal, security feature/constraint and security entity. ST is catering the notion of risk and mitigation, also can specifically in security goal during the stage of software development. The stakeholder is actor who can find threats, vulnerabilities and risk/mitigation in the requirements engineering process of a software development stage[53].

3.2.3 Disadvantages of Secure Tropos

Representation perspective and level of abstraction ST is typically goal-based approach, where objective is to satisfy the security goals of a system despite security concerns. But secure goals are in high level in many cases[27], hence it is not most appropriate for analyzing low level security concerns.

Kind of security requirement engineering (SRE) tasks/activities ST cannot support the criteria of "identification and modeling of assets" - the representation of entities that are considered valuable, a possible target of the threats that should be protected[52][41] [42]. It also cannot support "identification of vulnerabilities and threats" - the identification of software defects that can be exploited by an attacker to cause harm, and what could be an attacker's target[52] [41][42]. "Verification and validation support (VVS)" is also cannot fulfilled as well - peer review, walk through, formal methods or other verification and validation approaches that could be used to guarantee that the security requirements are correct, complete, consistent and unambiguous[49] [41][42].

Technical criteria and specification criteria ST has very good "internal verification support", but bad "external verification support" for the specification criteria of "correct". Specification criteria - "validatable" which is a sub-dimension of "External verification support" of technical criterion is also cannot fulfilled. Another defect in this dimension is the standards integration support for ST modeling technique, there are *understandable, consistent, verifiable and complete are just partially supported in the technical criterion*[53].

Modeling language, process and method ST modeling technique cannot fulfill with the requirements of reusing the provided artifacts in later phase in this dimension of "modeling language criteria"[53]. Only most ideas and concepts for ST are just used during the early and late requirements process, and it is also no evidence that ST can be used in the industry.

Relevant SRE notation and software evolution and other perspectives Unlike other i* modeling approaches, ST embraces the notion of risk and mitigation. But ST has no support for "component architecture" - the level of support for a component-based structure that allows software modules to be added or removed with ease. ST also has very low support for the "modularity", "change propagation", and "change impact analysis" criteria in the software evolution support dimension.

3.2.4 Other i*-based Modeling Techniques

It is also valuable to present other i*-based modeling techniques. Those have high similarities with ST. For instance:

1. **Enhanced Secure Tropos** is an extension of the Secure Tropos methodology. To achieve this, the underlining i* ontology [46] on which the i*-based modeling framework is based on was extended in order to fully model security. The approach [46]

introduced an enhanced ontology with three main notions, namely ownership-used to model the transfer of resources that an actor controls, delegation-used to model the transfer of entitlements and responsibilities between actors, and trust-used to model the belief of actors about the behavior of other actors, which together form the very foundation of all security concerns.

Typically, delegation is defined as a relation between two actors called the delegator and delegate, and a goal, task or resource called the delegatum. The two types of delegation are delegation of execution and delegation of permission [21]. Also, the notion of trust is used to differentiate delegation between trusted and un-trusted actors. Trust is defined as a relation between two actors called the trustor and the trustee, and a goal, task, or resource called the trustum. Again, the two types of trust that are defined are trust of execution and trust of permission [21].

2. **Secure i*** is an extension of Yu's i* modeling framework that is used for modeling and analyzing security trade-offs [22]. It tries to align security concerns with other equally important requirements such as usability, performance, and functional requirements. The approach is based on a meta-model of security concepts containing some important notions and their relationships [51].

   The important notions are: actors, assets, threats, and vulnerabilities. All of these cannot be represented by the i* modeling notations. Actors are entities that have security goals. Security goals are those that prevent or detect threats, or recover the system from threats. Actors are interested in Assets, can own, or delegate permission of usage of Assets to other actors. Threats targets Assets and occur through vulnerabilities. Threats can also be unintentional, or result from accident, human error, or natural disasters. Secure i* uses malicious representations equivalent of the original i* motions to model this additional aspects. Also, Secure i* includes a trade-off analysis method that makes use of the Secure i* notation in order to enable software engineers find trade-offs to mitigate threats or attacks [22].

3.3 Problem-based and model-based modeling techniques

Problem-based modeling techniques that bring informal and formal aspects of software development together in a single theoretical framework for software engineering design. The approaches conceive development as the representation and step-wise transformation of software problem [54]. Problem-based modeling allows modeling and analysis of different threat using formal and informal means to identify vulnerabilities and possible mitigations to threats [54].

Model-based modeling techniques are based on the use of models that would help requirements analysts to understand complex software problem, and identify potential solutions through abstraction[53]. The modeling notations are based on the UML or variants, or extensions of it.

The problem-based modeling approaches cover the Misuse Cases (MUC), Abuse Cases,

Misuse Case Map (MUCM) and Misuse Sequence Diagram (MUSD). The model-based modeling approaches cover the techniques: Misuse Cases, Abuse Cases, Security Use Case, Misuse Case Map and Misuse Sequence Diagram. Besides Security Use Case, other modeling approaches can be treated as both problem-based and model-based.

3.4 Use Case-based Modeling Techniques

The Use Case-based modeling approaches were typically defining interactions between a role (known in UML as an "actor") and a system, to achieve a goal[10]. The Use Case-based approach covers only functional goals of a system, with one or more actor directly involved in the software's operation. The actor can be a human or an external system. The notation of the Use Case is originally based on Ivar Jacobson's Objectory notation[37]. The Use Case model is also not a kind of standalone model, it can be used in conjunction with other models[58], such as i* models. The Use Case modeling approaches cover the Misuse Cases, Abuse Cases, Security Use Case, Misuse Case Map and Misuse Sequence Diagram. In order to parallel with i* -based modeling techniques in this empirical study, and considering the most frequency mentioned technique in recent academic research, Misuse Case was chosen as the candidate in the master thesis. Well, other modeling techniques were also presented since there are some similarities between them.

3.4.1 Misuse Cases Modeling Approach

Misuse Cases (MUC) [59] are used to capture the various ways that a user with malicious intentions can be a threat to a system. MUC extends regular Use Case diagram with new concepts such as Misusers[35], Misuse Cases[35] and Mitigation Use Cases[35][38] in order to elicit the security requirements of a system. It also incorporates two new relations, threat and mitigate, in addition to usual relations such as extend, generalize, and association of the Use Case diagram. Misuse Case diagrams use an opposite method of the Use Case notation to represent malicious intentions of a user towards a system. The MUC can be defined as Misuse Case and Misuser.

Misuse Case [59] A sequence of actions, including variants, that a system or other entity can perform, interacting with Misusers of the entity and causing harm to some stakeholder if the sequence is allowed to complete.

Misuser [59] An actor that initiates Misuse Case, either intentionally or inadvertently.

Compared to regular Use Cases, the inverted notation (figure 3.3) indicates both: similarity (because the same symbol shapes are used) and negation (because of the inverted graphics). Use Case and Misuse Case can, therefore, be shown in the same diagram without confusion.

Ordinary Use Case relationships such as include, extend, and generalize can be used between Misuse Case too, and ordinary association relationships can be used between Mis-

Figure 3.3: The notation of MUC

user and their Misuse Case. Misuse Case also can threaten a Use Case when the Use Case is threatened by the Misuse Case.

Here is an example of inverted graphics (figure 3.4) shows that Misuse Case together with regular Use Case in a high-level specification of part of Health Insurance System (see chapter 5). The figure 3.4 illustrates that attacker can attack the DHI department to modify older persons' personal health insurance data in the database. He also can illegally get the password and user name from the mobile device that used for collecting health data of the older person. After the attacker attacks the bank system, he also can make fake bills and illegally get account number and password from insurance company.

3.4.2 Advantages of Misuse Case

Representation perspective and level of abstraction Misuse Cases is a kind of model-based or problem-based modeling approach [43]. It is able to represent the scenarios that depict both low and high-level security concerns by using a combination of formal and informal techniques. Hence, MUC is more relevant for evaluating low-level security concerns when it compared with ST. And since Use Case-based modeling techniques are model-based modeling approaches in that they leverage significantly the UML notations framework making Misuse Case is easier to learn, use and integrate with conventional system design process.

Kind of security requirement engineering (SRE) tasks/activities The MUC approach adequately cater to 4 tasks/activities such as "elicitation and analysis (EAS)", "specification of SRs (S)", "identification and modeling of assets (IMA)" and "identification of vulnerabilities and threats (VT)" [43]. Hence, MUC will be relatively more useful for systems where mitigation of threats and defense against attacks are primary objectives.

Technical criteria and specification criteria Comparing with ST, MUC just has two significant benefits in this classification. For instance, MUC has good "internal verification support" in the technical criteria and specification criteria dimension, and when it compared with ST, MUC can partially support external verification - "correct"[27].

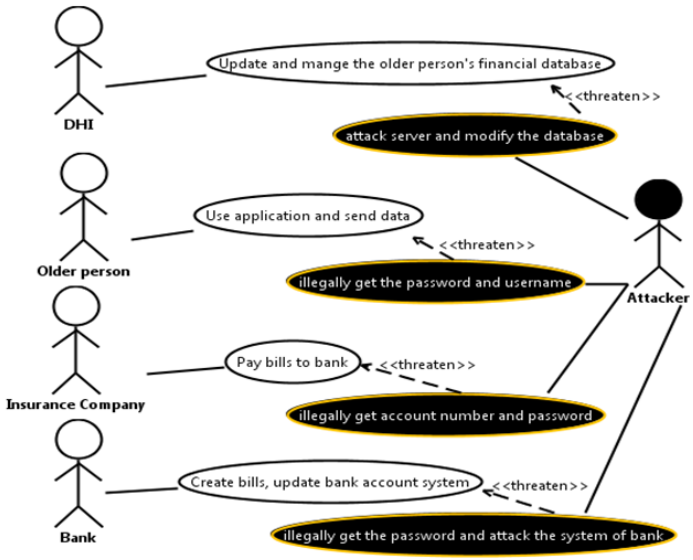

Figure 3.4: The example of MUC

Modeling language, process and method The same with ST, MUC offer very good support for aspects such as ability to formulate basic security requirements, and representation of usage scenario. There is also substantial tool support for MUC and also have evidence in the literature that suggests increasing interest and adoption of the MUC in industry[13]. MUC offer support for "reuse of provided artifacts" in later stages, especially in the "testing" stage of a system development[59]. It means MUC will be more relevant for test-driven development (TDD) of secure systems.

Relevant SRE notation and software evolution and other perspectives MUC embrace the notion of risk and mitigation[27] that guarantees MUC can be more suitable to use when issues of immunity against attacks or attack mitigations are paramount interest. MUC also have generally better support for "impact analysis" and "modularity"[43].

3.4.3 Disadvantages of Misuse Case

Representation perspective and level of abstraction Misuse Case is not goal-based modeling approach[43]. Hence, Misuse Case cannot support analyzer to model and analyze the system from the aspect of systems' security goals which can be insist of security constrains. Misuse Case is also not process-based. Thus, analyzing security requirements is also cannot fulfill when using MUC via process modeling to analyze secure concerns.

Kind of security requirement engineering (SRE) tasks/activities Compare to Secure Tro-

pos, Misuse Case can partially fulfill with classification of "verification and validation support" - only one dimension of "internal verification support" regarding to the classification of the kind of security requirement engineering (SRE) tasks/activities. And it is also only can identify threats but no mitigations can suggest according to the analysis the sub-dimension of "identification and modeling of assets" [43].

Technical criteria and specification criteria As well as Secure Tropos, Misuse Case is also cannot fully support for the dimension of Technical criteria and Specification criteria, such as Misuse Case is just partially support for "external verification", and no support for "standards integration" [27].

Modeling language, process and method Misuse Case is just partially support for requirements reuse except one of its character - "appropriate", and cannot fully support for the completeness of other development stages. In contrast with Secure Tropos that is able to "represent security mechanisms" and "low-level security requirements" in a system, Misuse Case has no ability to do either. Misuse Case also cannot provides "reuse of provided artifacts" in later phase since the reusable artifacts for the MUC is just only used on stage of "testing" [43][59].

Relevant SRE notation and software evolution and other perspectives Meanwhile, as cited in "software evolution support" for Secure Tropos, Misuse Case is not the best benefit of this dimension as well. Even though some of characteristics are higher support to do evolution, such as medium support for "modularity" and "change impact analysis", Misuse Cases are still cannot support for "change propagation". Hence, there is no chance to track the change of the system, and cannot promise that change is "correctly propagated", for example there still left several inconsistent dependencies in the MUC diagrams.

3.4.4 Other Use Case-based modeling Techniques

1. **Abuse Case** is used to show the complete interaction between a system and one or more actors, where the results of the interaction are harmful to either the system, one of the actors, or one of the stakeholders in the system[47]. Abuse case also presents abuse of patronage used to complete the use case. Abuse case uses the same notations as use case diagrams, it enables the creation of abuse case specifications in standard notation such as UML[33]. The difference between Abuse Case and MUC are that Misuse Cases use an inverted form of the UC notation but Abuse Case use the standard notion to show the interaction between a system and actors.

2. **Security Use Cases** [31] are used to specify how a system is able to successfully protect itself from relevant security threats. It allows the analysis and specification security requirements. Unlike **Misuse Cases** [59] focus on detecting successful attack against a system and analysis security threats, Security Use Case shows a system's capability to handle security threats. Security Use Case is used by the requirement team, other than security experts. The composition of a Security Use Cases depends on the type of security requirements a system is supposed to fulfill [31].

3. **Misuse Case Map** combines notions from Misuse Case and Use Case Maps to capture the security requirements of a system from an architectural viewpoint. Misuse Case Maps can be used to envision the effect of an intrusion on the architecture of the system. It extends the notation of the Use Case Maps notation with vulnerabilities, exploit paths and mitigation. MUCM has potential to significantly improve the understanding of intrusions and identification of mitigations when compared to MUC combined with a system architecture outline[41].

4. **Misuse Sequence Diagram** is a sequence diagram in a UML is a kind of interaction diagram[42]. The sequence diagram is used to show the process operation and operation orders. Misuse Sequence Diagram[42] is a threat modeling approach designed to show the sequence of attacks from an intruder to a system. MUSD is based on the UML and MUC diagrams and utilizes security concepts like vulnerability exploitation and mitigation. The notation of MUSD is very like the MUC notation. The benefit of MUSD is that the MUSD can show the sequence of steps of the intrusion.

Chapter 4

Experiment Management

This chapter defines the overall approach for experiment's management. The management highlights roles, schedule, risks and how to handle the experiment results.

4.1 Experiment Organizers

The experiment organizers were the co-supervisors, Peter Karpati and Olawande Daramola, they assisted the participants, organized them and helped to answer any question during the experiment. They also record time spent, problem raised by participants.

4.2 Participants

The experiment subjective consists of undergraduate students group - XCom'13 at IDI, who was always collecting money for excursion. The reason why they were chosen as the subjective of the experiment because we can pay some money to support their excursion. Hence, the payment may motivates the participants to seriously take part in the experiment. If not,in future nobody dares to ask them to participant in any experiment or research anymore due to their bad attitudes of the experiment. There are several research experiments in a year, thus their attitudes were also an emphasis issue for other researchers to reference. The participants were currently at their second year studies at NTNU, and they had the background of ICT as preferred. Through the pre-questionnaire of the experiment, the participants were asked to report their general background such as the man-month work and the lengths of their finished semester studies at the IDI department. Most participants reported three semesters of ICT studies that means they had been taught Use Cases modeling techniques or at least UML activate sequence modeling in their software engineering course as one of the compulsory courses for undergraduate students at IDI. They also reported 0-2 man-month work off the campus studies. For instance, the internship or

summer job in an enterprise. Thus, they have ability to use one of the mentioned modeling approaches in the pre-experiment sheets (see appendix A). In the line with their relevant student projects and exercises on campus studies, also with their ability to understand the basic of this experiment content according to their experience on relevant enterprise internship, the participants can be believed have ability to finish the experiment. However, the experience on any enterprise work was not a mandatory criterion for seeking participants since the participants were students at the university, they were not the expert. Therefore, the background of the man-month work was a threat to validity like other threats in this experiment. It was worthy to discuss the threats to validity of the experiment to ensure the possibility of replicating the investigation of the experiment for reuse later. It was also had value to discuss the possibility of replicating both the design and the results. Hence, the inevitable threats to the validity of why the experiment had been decided to use these students as the experiment sample and why not other participants were discussed in chapter 12.

4.3 Experiment Time Schedule

The experiment was performed at 14:15 PM at March 13th, 2012 in auditorium F1 at the Norwegian University of Science and Technology. Since there were 50 participants, hence the room F1 was big enough for the experiment. The experiment was expected 110 minutes to run, but the participants were too smart to finish the experiment before the planned time. Time threats that may impact on the results of the experiment were also discussed in chapter 12.

4.4 Experiment Risks

There were definitely some risks that threaten the validity of the experiment which cannot be avoided. All the identified risks were analyzed and evaluated such as student background, experience on the modeling approaches and other relevant issues and so forth. Although the experiment was expected to valid for the research goal, there was still a possible that the experiment might be not in very appropriately method since this experiment was the first time to perform an empirical study of the two modeling approaches in the requirements engineering research field. Hence, it was very valuable to discuss the experiment reliability and these kinds of problems were also the threats to the validity of the experiment (see Chapter 12). All problems that cannot be avoided were reduced as less as possible according to the knowledge and experience of the author with the purpose to improve the experiment practicality and for the future replication.

4.5 Experiment Measurement

The measurements can be divided into four parts. Not only participants' background was compared but also the quantity of threats and mitigation they suggested were also analyzed as well. The experiment was designed to get the participants' feedback based on the participants' performance and preference on the modeling techniques. The performance can be analyzed through the number of identified threats and mitigations where the participants have read and applied from the relevant cases with modeling techniques. And their self-assessed the usage of diagrams, textual description and memory were counted in percentage with the purpose to evaluate the preference of the participants on the two different modeling techniques. It also can further analyze their preference through their feedback from the post-questionnaire (TAM format[20]). The data reported by the participants by their self-assessment can help to analyze their perception of the two modeling approaches. Hence, the measurements that used in this experiment could be more precisely in quantity standard other than just asking them whether they will use which modeling technique in the future like in the post questionnaire (see Appendix A). The quality data can also help to analyze the experiment from the statistics view and can enhance the results more dependable. The further experiment sheets design details was discussed in chapter 6 with the experiment variables and hypotheses. Other experiment results such as time used, problems that occurred during the experiment were also recorded by experiment organizer and analyzed further in chapter 12.

4.6 Experiment Measurement Priorities

The experiment measurement priorities in the experiment can be expressed as three main higher priorities. The primary objectives of procedure were: 1) the numbers of threats and mitigations that the participants have identified and suggested, and 2) their estimated the usage percentage of textual description and diagram, 3) their perception of the two modeling techniques. The reason why these three measurements were higher priorities since through a multi-dimensional view of the experiment, the excepted results can be more credible for the research. And because the main concerns of the experiment was two core problems - the participants' performance and their preference, thus the three measurements can fully fulfilled with evaluation criteria. Therefore, these three measurements have been decided as the high priorities in the experiment results analysis. The general categories of threats and mitigations according to the participants' responses were also cannot be ignored in the experiment. However, the qualities of threats and mitigations were considered as low priorities because each evaluator has his/her own rule and criteria for the qualities. Therefore, the qualities of the threats and mitigations and its sub-categories of threats and mitigation were difficult to evaluate, so the main core points of the research were not focus on the qualities in this time. Last but not the least, time spend and problems occurred during the experiment were considered as low priorities, how they impacted on the experiment results were also discussed in the chapter 12.

Chapter 5

Description of System Design

To test the procedure in an experiment, two systems were designed and paired with two modeling approaches. After each case description, a possible system threats description was presented in order to help the readers to understand the system quickly.

5.1 The Background of Cases

The background of the first case - Health Insurance System came from the author's customer driven project course in the autumn 2010. The first case extends the SINTEF student "ESMUS" project[69], combining Norwegian health insruance welfare with a home health monitoring system, that developed with the purpose of remotely monitoring the US Military veterans. The second case was using the background of eBay, Amazon and Tmall online shopping. Nowadays, the buyer browses the goods online, when he/she decides to send the order; he/she has to provide credit card information in the parallel with sending the order to the transaction system. In this case both buyers and salespersons were ranked by credit rating and the case add a new method to the current transaction e-commerce - the "trust", buyers can refer to the credit rating of the salespersons to decide whether to buy the goods or not.

5.2 System in Details

5.2.1 Health Insurance System

An application that runs on the Microsoft and Android Mobile platform could be used to collect an older person's health data everyday which includes heart beat rates, carbon dioxide in blood, body temperature, and blood pressure.

All of the data from an older person could be sent at once or one at a time to the Health Insurance Department (HID) server in the hospital. The most important and delicate matters for the health insurance department is the privacy of all older person's medical information and the sharing of it. When HID receives the data, a doctor analyses the data and gives a feedback, the feedback is sent to the older person by a system, and also a copy is sent to HID.

When HID receives the feedback form from the doctor, they will calculate fees and update the health insurance database in the hospital, then check the database of the status of older person's health insurance, to figure out whether the medical health fee is now above 1800NOK or not. According to the welfare and security system in Norway, if the medical health fee is above 1800NOK the older person will not need to pay the bills in a normal year rather, HID will request bank to send the bills to the insurance company, otherwise, HID will request bank to send bills to the older person that should be paid in a normal year.

When bank receives the message from the HID, they will create bills according to the messages received from HID, and send bills to the older person or insurance company. We are only concerned with the way the older person and insurance company pay the bills via Internet. Bank should guarantee that there is a secure interaction between the users and bank system.

5.2.2 Threats for Health Insurance System

The Health Insurance System extends the ESMUS project with a combination of Norwegian healthcare welfare case. The design of the system was considered the dataflow in the system. For instance, the application used on the smart mobile phone with anti-virus software, the attacker attacks that phone and eavesdrop on the dataflow between the phone and the system in the hospital. The attacker can also eavesdrop the dataflow within the system and the dataflow between the hospital and insurance companies. Hence, the threats exist vary in the system communication. Figure 5.1 presents an overview of the HIS system and the relationship between each agent of the HIS system.

5.2.3 Secure Tropos diagram for Health Insurance System

Figure 5.2 illustrates the Secure Tropos diagrams for Health Insurance System. The notation used in the diagram was presented in Chapter 3. Four security constraints were considered in this case: 1) When older person send his/her health data via the application on mobile, he/she has to verify the user name and password of the application on the platform. 2) When older person pays the payment to bank system, there are 3 security constraints to ensure security concerns between the interaction between the older person and bank system - Personal account number, personal number and the password of the account. 3) When insurance companies pay the payments to bank system, they also need to verify their account number, company's ID and the password. 4) Bank should double check the request messages from HID before the bills are created.

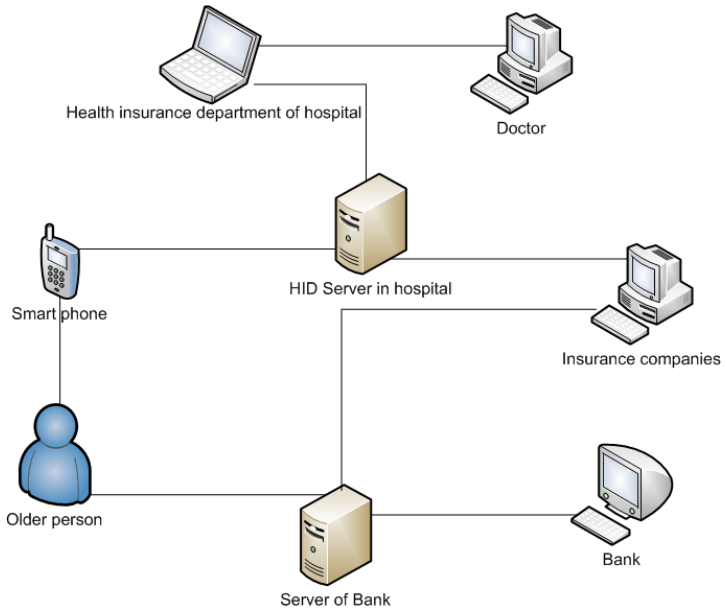

Figure 5.1: Relationship between different agents of the Health Insurance System

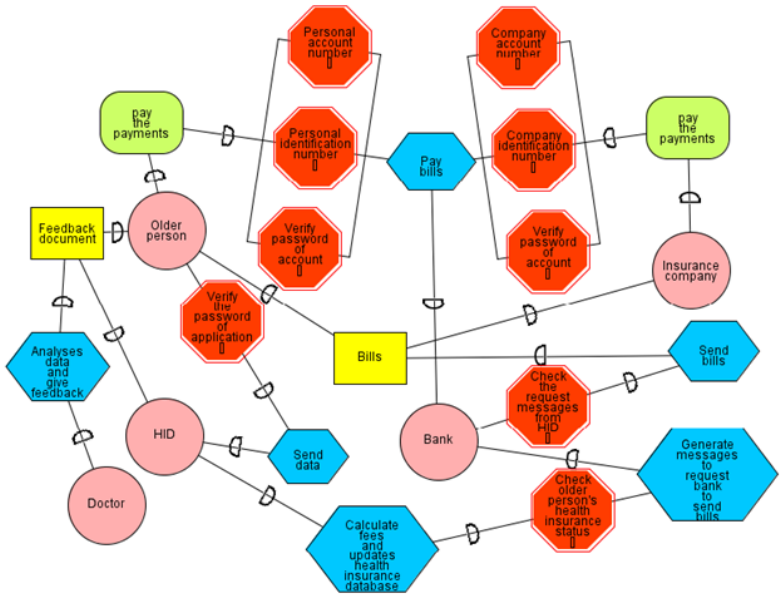

Figure 5.2: Overview of Health Insurance System

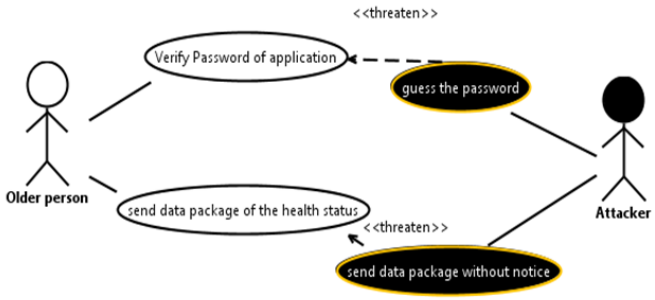

Figure 5.3: Attacker illegally get the password and send data to DHI

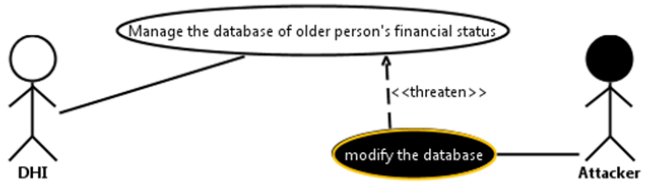

Figure 5.4: Attacker modifies the database after he/she attacked the system

5.2.4 Misuse Case diagrams for Health Insurance System

Figure 5.3 and 5.4 are the examples of Health Insurance System's threats. Attacker can get the password and send the data package to DHI server without notice by older persons in figure 5.3. Figure 5.5 presents attacker gets the password of the insurance company, then he can transfer money to his own account.

Figure 5.6 shows the attacker gets the password of the bank system and login illegally. He can create fake bills and collect the payments through his own account.

5.2.5 Net Shopping

The most known traditional way is that we browse goods, compare the goods provided by different salespersons, and decide to buy it, pay for it and wait for its delivery. Another way to buy stuff via the Internet is "trust rating". In trust rating, we pay the fees to pay-pal first, when we receive the goods, we check and use it within 3 days to ensure we are comfortable with the goods then we can confirm the deal with pay-pal to finish this transaction, pay-pal pays money to the seller after we confirm the deal.

"Trust rating" is a transaction for the salesperson. Each successful transaction allows buyers to mark a credit evaluation for it. Evaluation is divided into positive, neutral, and

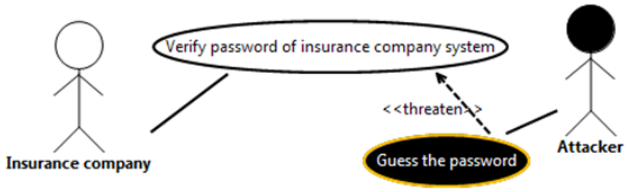

Figure 5.5: Attacker gets the password of the insurance company system

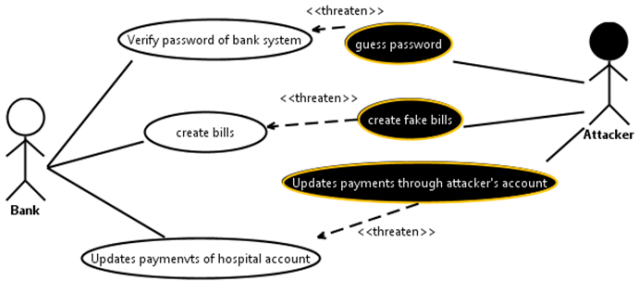

Figure 5.6: Attacker gets the password of bank system and create fake bills, collect payments by his own account.

negative feedback categories. Each credit is equal to a credit score. After each transaction, if there is a good evaluation for the salesperson, they will get one point. No point for medium evaluation, one point is deducted from a salesperson for a negative evaluation. If a salesperson has no medium and negative evaluations within 300 continuous transactions, he will get a diamond to illustrate the high trust rating. The more diamonds he has, the higher the good quality and trust during the transaction. In this case there are three important vulnerabilities:

1. How to avoid salespersons selling goods to themselves in order to earn more diamonds.
2. The salespersons could sell the virtual products (e.g. illegal scanned eBook) in a very low price to their partners (themselves or other salespersons), after they get the diamonds, they may sell other real products (e.g. print copy of a real book).
3. Some third party platforms may help salespersons to get diamonds in a dubious way, such that all the sold goods are later refunded to the seller but they will get a fee after the transaction (when the salesperson has already earned the diamond).

5.2.6 Threats for Net Shopping System

The Net shopping case refers to nowadays online transaction. An assumption was made for the system that there is a possible way that buyers buy goods after checking salespersons' credit ranking. In the case, "Trust" is used as a bridge, the buyers can try to use the goods with few days before he/she decides to whether buy it or not. The rule of credit ranking is based on the successful transactions. If there is no medium evaluation from buyers, the salespersons can earn 1 score. And if salespersons get 300 continues positive evaluation, they will get diamonds to show they are in the best trust rating. Hence, if we want to update nowadays transaction systems, there are several threats shall be avoided. For example, salespersons sale goods to themselves and evaluate transaction by themselves shall be avoided. We also have to make sure the dataflow is safe if a hacker is eavesdropping on the system. The third party companies are also forbidden to help salespersons to earn the diamond via fake transactions. For instance, salesperson sales goods to these companies, after these salespersons pay money to these companies, these companies will give a positive evaluation to salespersons and return the goods to salespersons too. Finally if buyers just want to try to use the goods but never want to pay for it on purpose, the avoiding method for this situation should also be realized in real life. Figure 5.7 illustrates the relationship between the agencies of the Net Shopping system.

5.2.7 Secure Tropos diagram for Net Shopping

Figure 5.8 illustrates Secure Tropos diagrams for Net Shopping. In the diagram, 1) there are two security constraints when the buyers confirm a transaction, the buyers' account number and its password. 2) When buyers decide to buy goods, the order will send to

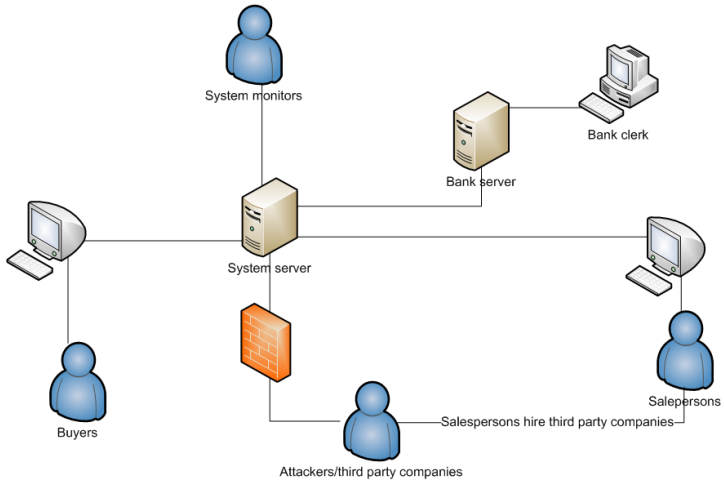

Figure 5.7: Relationship between different agencies in the Net shopping system

system trader monitors first in order to double check that whether the buyers and salesperson is the same person. 3) The information of goods on the website shall be published by system trader monitors and they have to check whether the goods price and whether the goods is a virtual product.

5.2.8 Misuse Case diagrams for Net shopping case

Figure 5.9 shows salesperson pretends as a buyer and sales goods to himself or herself. Figure 5.10 shows some third parties pretend as a buyer to buy goods to help salesperson to get more diamonds. Figure 5.11 shows how attacker can pretend as a system trade monitor to allow the virtual products transaction.

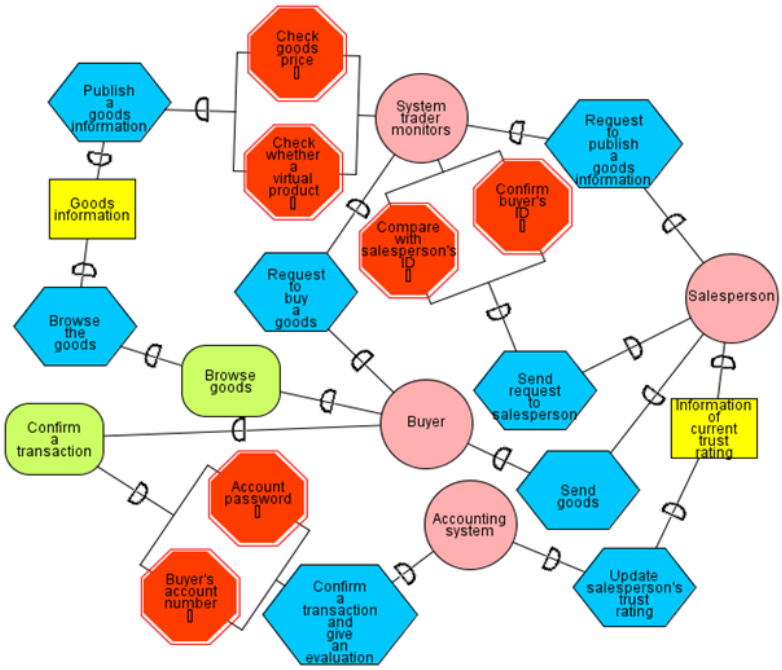

Figure 5.8: Secure Tropos diagrams for Net Shopping

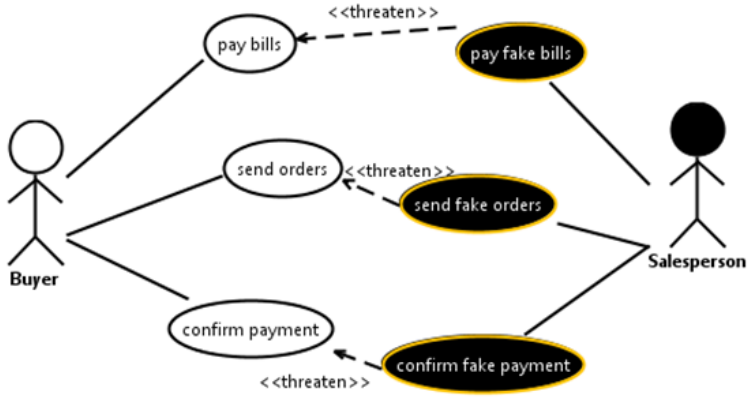

Figure 5.9: Salesperson pretends as a buyer

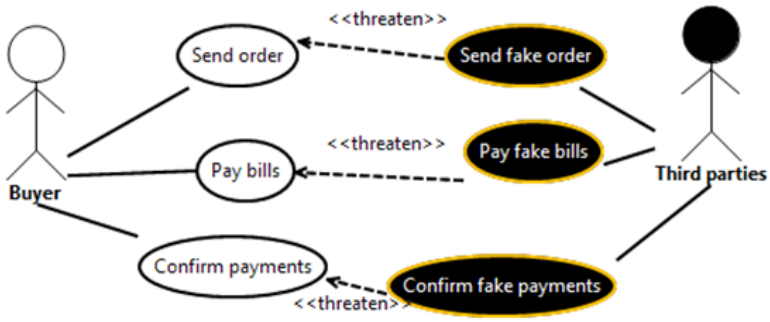

Figure 5.10: Third parties pretends as a buyer

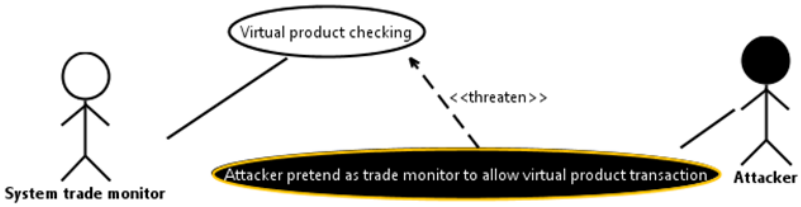

Figure 5.11: Attacker allows the virtual products trade

Chapter 6

Experiment - The Variables and Hypotheses of the Experiment

In order to get a basis for evaluation the performance and preference of the ST and MUC approaches, this experiment was performed. The experiment involved computer science undergraduate students in their second years study, and they already had relevant experience on the system modeling studies or some of them had internship experience in enterprise.

6.1 Experiment Objective Definition

As mentioned above in the problem description, the main objective of the experiment was to identify whether there were significant differences between the two modeling approaches when the participants were trying to apply approaches to two cases. The experiment was motivated that it was the first time to compare the two modeling techniques - ST and MUC in an empirical study, thus it was need to understand the difference in individual performance and preference of participants when they were using the two approaches regarding to their background. The method used in the experiment to determining the objective of the research was a Latin-Square experimental design to control the techniques and cases order. All the participants' responses were recorded on the sheets of experiment.

- **Object of Study** The object of the experiment study were the sophomore students at the Department of Computer and Information Science, Norwegian University of Science and Technology(NTNU). Before the experiment, It was already known that they have learned the UML use case modeling more or less in their software engineering course and they may also used this technique for their exercises during their recent two years study.

- **Purpose** The purpose of the experiment was to evaluate the participants' performance and preference based on the background of their two years study and probably their summer internships.

- **Perspective** The perspective was from the point of view of the author, the author's supervisor and co-supervisors from the "ReqSec" project team. The interesting thing concerns that if there was any systematic significant difference in the performance and preference between the two modeling approaches. And there were also people who might be interesting in the experiment results.

- **Quality Focus** The main effect in the experiment was the individual learning of UML Use case and other relevant modeling techniques study. Since the design of experiment was not plan to introduce ST and MUC modeling approaches to the participants in advance. Hence, both modeling techniques had the same complexity for the participants but may not involve the cases; the case validity was also discussed in the results and chapter 12.

- **Context** The experiment was run as an offline format - paper format. The context of the controlled experiment was constructed based on the author's project during the master study and current e-commerce transaction.

6.2 Planning

6.2.1 Context Selection

As described in chapter 5, the contexts were selected from the author's project and current online transaction in the world. Since the experiment was planned to perform at the university, so it was run off-line. The cases addressed the real problems in recent research of IT technology.

The use of the two modeling approaches with two cases as the experiment context provide not only the author and his supervisors but also other researchers a brand new view of the two modeling techniques according to the experiment results. The experiment sheets were designed with the answer sheets together, thus no extra forms were need for the experiment data collection and so forth. Furthermore, the experiment context also provides a good opportunity for replicating such as find threats and mitigations, estimate the usage of diagrams and textual description section in the two parts of experiment. Also, the participants have studied relevant modeling technique - UML. Hence, there was no need to spend too much time and effort to teach them the knowledge about that modeling method, so creating the background environment for the experiment to run was not an essential thing since all the participants were come from the same department with very similar background.

6.2.2 Variables and Its Connection with the Measurement

There were 12 variables were chosen by strictly following the criteria of independent variables and dependent variables (Claes Wohin 2000)[17]. In table 6.1, there were the variables and their explanation used for the experiment. The variables can be classified as four main areas - the background, performance, estimating the usage of two modeling approaches and the perception of the two modeling approaches.

1. Several general variables were used to evaluate the participants' background, there were KNOW_UC, KONW_MUC, KNOW_IST, KNOW_ST, KNOW_Mal-activity, KNOW_UML-activity, KNOW_MOD, STUDY and JOB.
2. THR, MIT and THMI were used for evaluating the participants' performance during the experiment.
3. The variables of THR_TXT,THR_DIAG, THR_MEM were used for asking the participants' how they separated their work in each part of the experiment by using threats to finish the tasks. On the contrary, MIT_TXT, MIT_DIAG, MIT_THR and MIT_MEM were used for asking the participants to estimate the percentage of usage in each part of the experiment in condition of they just use mitigations to finish the tasks. The number they estimated here for the two part - THR and MIT were counting into 100% for each part.
4. There were 12 questions in the post-questionnaire, which were used for asking participants report their perception of the modeling techniques after they use the modeling technique to finish the tasks. The 12 statements were marking from PER_1 to PER_12 and these 12 questions can be classified into 3 categories - PER_PU, PER_PEOU, and PER_ITU. The 3 categories were used for evaluating the participants' perception of the two modeling approaches. For instance, whether it was easy to use the modeling technique, or whether they will use the modeling technique in future.

Points 2 to 4 were the main focus of the empirical study as mentioned above, hence the hypotheses were also based on these concerns as planned. (see 6.2.3)

Table 6.1: Variables of the Experiment

Name	**Explanation**
TECH = MUC, TECH = ST	The technique used in that part of the experiment, either MUC(Misuse Case) or ST (Secure Tropos).
CASE = HIS, CASE=NS	The case used in that part of the experiment, either HIS (Health Insurance System) or NS (Net Shopping).

Continued on next page

Table 6.1 – continued from previous page

Name	Explanation
KNOW_UC, KNOW_MUC	The participants' self-assessed knowledge about use cases (KNOW_UC), misuse cases (KNOW_MUC) on a 5-point scale, where 1 is "Never heard of it" and 5 is "Expert".
KNOW_IST, KNOW_ST	The participants' self-assessed knowledge about i* (KNOW_IST), and Secure Tropos (KNOW_ST) on a 5-point scale, where 1 is "Never heard of it" and 5 is "Expert".
KNOW_Mal-activity, KNOW_UML_Activity	The participants' self-assessed knowledge about Mal-activity (KNOW_Mal-activity), UML activity diagram (KNOW_UML_Activity) on a 5-point scale, where 1 is "Never heard of it" and 5 is "Expert".
KNOW_MOD	The participants' self-assessed knowledge about system modeling (KNOW_MOD) on a 5-point scale, where 1 is "Never heard of it" and 5 is "Expert".
STUDY	The participants' self-reported semesters of ICT-studies.
JOB	The participants' self-reported man-month of ICT-relevant work experience.
THR, MIT	The number of unique threats (THR) and mitigation (MIT) identified by the participants.
THMI	The sum of unique threats and mitigation (THMI) identified by the participants.
THR_TXT, THR_DIAG, THR_MEM	The estimated extent of the use of the textual description (THR_TXT), diagram(s) (THR_DIAG) and the exclusive use of memory (THR_MEM) to identify threats in percentage (should sum up to 100).
MIT_TXT, MIT_DIAG, MIT_THR, MIT_MEM	The estimated extent of the use of the textual description (MIT_TXT), diagram(s) (MIT_DIAG), the exclusive use of the previously identified threats (MIT_THR) and the exclusive use of memory (MIT_MEM) to identified mitigation in percentage (should sum up to 100).
PER_1, PER_2,...PER_12	Scores on the 5-point Likert scales for the four statements about perceived usefulness (PER_PU), perceived ease of use (PER_PEOU) and intention to use (PER_ITU) of the techniques.

Continued on next page

Table 6.1 – continued from previous page

Name	Explanation
PER_PU, PER_PEOU, PER_ITU	Average scores on the 5-point Likert scales for the four statements about perceived usefulness (PER_PU), perceived ease of use (PER_PEOU) and intention to use (PER_ITU) of the techniques.
PER_AVE	Average scores on the 5-point Likert scales for all the twelve statement about the techniques.

6.2.3 Hypotheses Formulation

The hypotheses for the experiment were listed in table 6.2. There were 9 hypotheses of the experiment; all of them were covered with the variables that were mentioned in the table6.1.

Table 6.2: Hypotheses of the Experiment

ID	Description	Representation
$H1_0$	There will be the same number of identified threats between the each two groups.	THR[ST] = THR[MUC]
$H1_1$	There will be a significant difference in the number of identified threats between each two groups.	THR[MUC] $\neq$ THR[ST]
$H2_0$	There will be the same number of identified mitigations between the two groups.	MIT[ST] = MIT[MUC]
$H2_1$	There will be a significant difference in the number of identified mitigations between the two groups.	MIT[MUC] $\neq$ MIT[ST]
$H3_0$	There will be the same number of identified threats and mitigations between the two groups.	THMI[ST] = THMI[MUC]
$H3_1$	There will be a significant difference in the number of identified threats and mitigations between the two groups.	THMI[MUC] $\neq$ THMI[ST]
$H4_0$	There will be the same in the estimated use of the diagrams of the two techniques in identifying threats.	THR_DIAG[ST] = THR_DIAG[MUC]
$H4_1$	There will be a significant difference in the estimated use of the diagrams of the two techniques in identifying threats.	THR_DIAG[MUC] $\neq$ THR_DIAG[ST]
$H5_0$	There will be the same in the estimated use of the diagrams of the two techniques in identifying mitigations.	MIT_DIAG[ST] = MIT_DIAG[MUC]
$H5_1$	There will be a significant difference in the estimated use of the diagrams of the two techniques in identifying mitigations.	MIT_DIAG[MUC] $\neq$ MIT_DIAG[ST]

Continued on next page

Table 6.2 – continued from previous page

ID	Description	Representation
$H6_0$	The two techniques will be perceived very similarly regarding usefulness.	PER_PU[ST] = PER_PU[MU]
$H6_1$	The two techniques will be perceived significantly differently regarding usefulness.	PER_PU[MUC] $\neq$ PER_PU[ST]
$H7_0$	The two techniques will be perceived very similarly regarding ease of use.	PER_PEOU[ST] = PER_PEOU[MUC]
$H7_1$	The two techniques will be perceived significantly differently regarding ease of use.	PER_PEOU[MUC] $\neq$ PER_PEOU[ST]
$H8_0$	The two techniques will be intended to be used significantly similarly.	PER_ITU[ST]= PER_ITU[MUC]
$H8_1$	The two techniques will be intended to be used significantly differently.	PER_ITU[MUC] $\neq$ PER_ITU[ST]
$H9_0$	The two techniques will be perceived significantly similarly	PER_AVE[ST] = PER_AVE[MUC]
$H9_1$	The two techniques will be perceived significantly differently.	PER_AVE[MUC] $\neq$ PER_AVE[ST]

6.2.4 Experiment Design

All the problems concerned have been stated above already, and the independent and dependent variables also have been chosen. Hence, the following steps were determining the measurement scales for the variables. In order to reduce the learning effect, the experiment sheets were designed by following the rules of Latin Square experimental design[4]. The experiment involved student participants recruited from the sophomore students in March 2012. The students belonged to an organization that received some payment for each participant in the experiment. The research questions were stated in previous chapters, the payment in this experiment was the inducement in order to ensure that the participants took in the experiment seriously.

The participants have used two techniques individually on two different cases: Health Insurance System (HIS) and Net Shopping (NS). To control the techniques and cases order in the experiment, a Latin-Squares experimental design[4] was used as shown in the table 6.3. The within-experiment data regarding performance, estimated use of diagrams and perception thereby become paired, comparing two dependent samples.

Table 6.3: Latin-Squares experimental design

Case order: Technique order:	HIS before NS	NS before HIS
MUC before ST	Group 4	Group 3
ST before MUC	Group 2	Group 1

6.2.5 Data Collection

After controlling for the participants' background, for each combination of technique and case in the experiment sheets, three types of tasks were designed for collecting data in both parts of the experiment. There were a performance task, an estimation task (estimating the usage of the textual description, diagrams and their memory for the performance task in percentage) and a perception task (Post-questionnaire). All the variables were summarized in table 6.1. The primary distinction in the experiment was based on the technique used (variable TECH). All the other measurement variables were categorized and compared with each other depending on whether they related to one or the other technique. For example, the number of mitigations identified (MIT) when ST was used, was expressed as THR [ST]. The measurement variables were explained in section 6.2.1 and summarized here in the following paragraphs.

- **Background** was measured by a pre-task questionnaire addressing the participants' self-assessed knowledge of several modeling approaches (see table 6.1) on a 5-points scale [5], where 1 is "Never heard of it" and 5 is "Expert". They were also asked to report their numbers of completed semesters of ICT studies and man-month of ICT-relevant work experience.
- **Performance** was measured by asking the participants to identify and list as many as threats or mitigations as they can for two specific systems. The answers were coded qualitatively in order to determine exactly how many threats and mitigations each participant have identified according their responses. It was also planned to categorize the type of the threats and mitigations but due to the criticality of evaluation by different person make it was a difficult task to control. Thus, this work will left for future experiment where boilerplate[1][60] - a standard templates of the security threats and mitigations identification will be considered to help to solve this problem in future experiment. However, a general classification for the threat types were introduced in the experiment as a pioneer work for the future experiment.
- **The estimating of use of text, diagrams and memory** were used for solving the performance task when the participants spent their attention on estimating the percentage of use textual description, diagrams and memory in different tasks and modeling techniques. The percentage was counted into 100%.
- **Perception** was measured by a post-task questionnaire adapted from the Technology Acceptance Model (TAM) (Davis 1989)[20]. There were 4 questions addressed perceived usefulness (PU), 4 questions addressed perceived ease of use (PEOU) and 4 investigated the participants' intention to use (ITU) the technique in the future, each rated on a 5-point Likert scale. There were 3 negative questions, with a lower score reflection a positive opinion that were inverted before data analysis(see section 8.2.4) .

Chapter 7

Experiment - Execution

7.1 Preparation

The participants were informed that the experiment was planned to study the outcome for the "ReqSec" project, and their background match the research expectation. The participants were interesting in the experiment since they need money for their excursion. The experiment budget allows us to offer them a fee to support their excursion, they granted that they took part in the experiment seriously at least they tried their best to do the experiment, since a good attitude was very important for the experiment.

7.2 Execution

The experiment was executed within the expected minutes. Due to the cases and techniques order, the participants used different the experiment sheets and filled in their responses in the experiment sheets. Each group had to try to use the two modelling approaches with the cases and filled in the questionnaire forms after they read the modelling technique with case in each part of experiment. The detailed experiment procedure comprised the following steps:

1. Filling in the pre-experiment questionnaire.
2. Reading a short introduction to the experiment.
3. Using the first assigned technique on the first assigned case:
   (a) Reading the introduction to the first technique.
   (b) Reading the textual description of the first case.
   (c) Identifying as many threats and mitigation as possible.

(d) Estimating the usage percentages.

(e) Filling in the post-experiment questionnaire (TAM).

4. Short break.
5. Repeat steps 3a-3e for the second technique and case.

The duration of the steps was decided dynamically, there was always enough time to finish the steps.

7.3 Data validation

The data were collected from 50 participants. After checked their responses sheets, there was no evidence to remove any data from the participants' responses that means all the responses were valid. It was pleasant to see that they can took in the experiment seriously and the data cannot be better since all the groups finished the experiment and filled in all the tasks in the experiment sheets. Even though there was one student from the group 2 who did not response the task of estimating usage of diagram, textual and memory, it still no need to remove data and could kept his/her data in the sample, since this situation might be occurred when there was an occasionally issue during the experiment execution. Also, it was not essential to contract this student to fill in it again since s/he already knew the experiment text, there must exist a big learning effect to risk the experiment.

Chapter 8

Experiment - Analysis and interpretation

This chapter presents the data collection, how the data was analyzed and interpreted. How the data reflects the research goals were also discussed in this chapter.

8.1 Statistics description

8.1.1 Data representation

As the first step of analyzing the data, several graphical statistics were presented to visualize the data with the purpose of illustrating data distribution.

- **Background** According to the pre-experiment questionnaire responses, it was discovered that the most participants had experience or have tried to use UML use case and system modeling in their previous studies. But almost all the participants were not familiar with other modeling approaches. It was not a surprise since they were sophomore students and none of them were expert of any modeling technique. Hence, they match well enough for the experiment sample expectation. The following figures 8.1 to 8.6 show the detail information of the participants' background.

  In figure 8.1 and 8.2, the vertical axis value represents how many participants self-report their knowledge of the modeling techniques in the experiment sheets. The horizontal axis presents the knowledge level of relevant modeling techniques that cover from "never heard about it" to "expert". For example, it can be found that in the figure 8.1 and 8.2 there was no one expert on all these modeling techniques.

  Figure 8.3 represents the participants' completed study at the university level. Apparently, most of the participants had 3 semester studies in information technology.

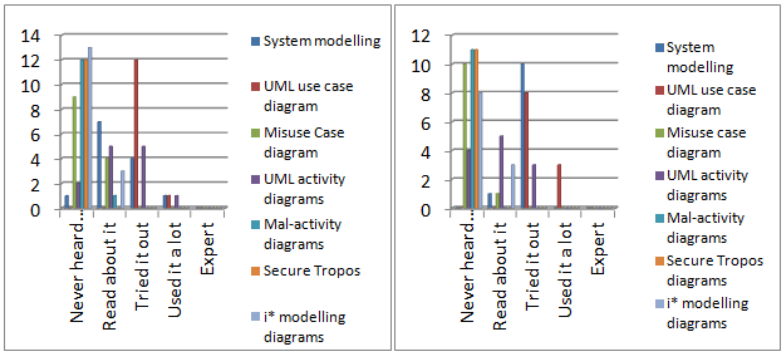

Figure 8.1: Pre-experiment questionnaire of Group 1 and Group 2

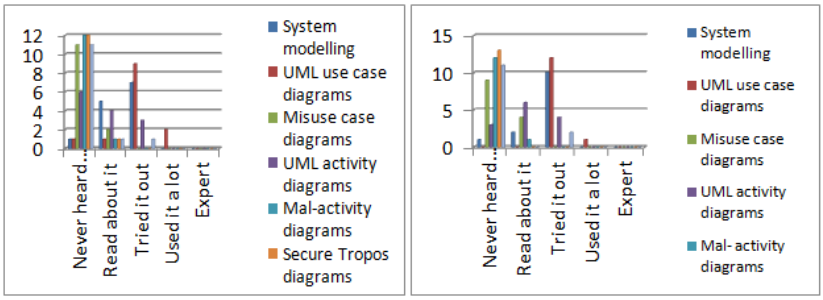

Figure 8.2: Pre-experiment questionnaire of Group 3 and Group 4

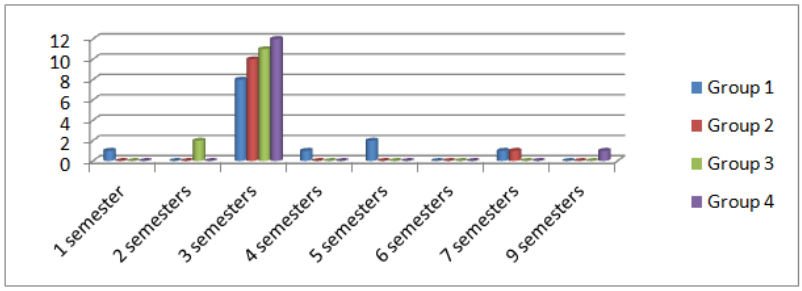

Figure 8.3: Completed study semesters

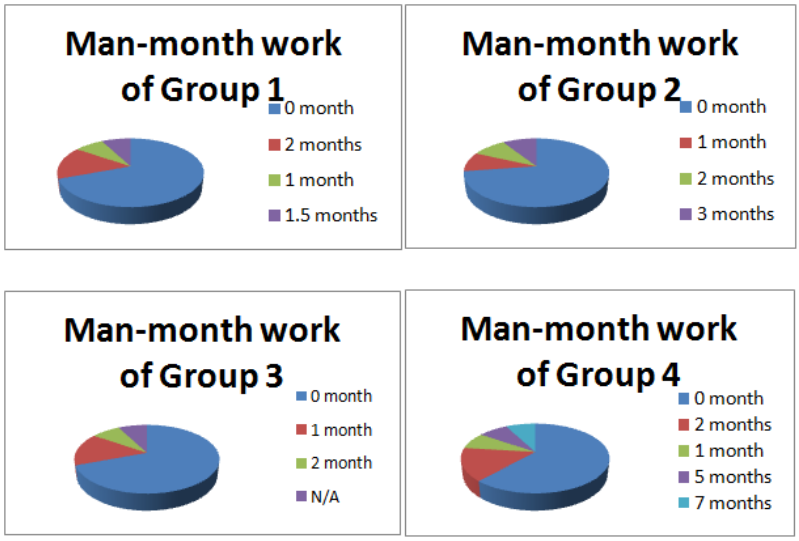

Figure 8.4: Man-month jobs

However, there were still 5 persons from the group 1 report they had more than or less than 3 semester studies, such as 1, 4, 5 and 7 semester studies. 2 persons from the group 2 reported they had 2 semester studies and 1 person from the group 4 reported 9 semester studies at the university level. But for those participants who have reported they had 5, 7 and 9 semesters study that can be divided into 2 parts according to their self-report that they had 3 or 4 semester studies in IT, and few semesters study in other study fields. Hence, they were still treated had around 3 semester studies at ICT courses.

Most of the participants reported that they had no work experience. One student from group 3 did not report the man-month job, so it was marked N/A in the figure. However, few participants had relevant internship or summer job work experience on ICT technology (see figure 8.4).

- **Finding Threats and Mitigation by Groups** The following figure shows that the total number of threats was identified by each group. Since each group have used two different cases with two modeling techniques, the data was extracted from the paired technique with case regarding to the group number (see figure 8.5 to 8.8). The comparison of performance was discussed in section 8.2.

  Figure 8.5 presents that group 1 and group 4 had the same experiment task in their first part of experiment. The modeling technique was ST with Net shopping case. The participants in group 1 found 36 threats and 40 mitigations when they have applied ST modeling technique with NS case. Group 4 found 36 threats as well, but they just suggested 31 mitigations in total when they have applied the same modeling

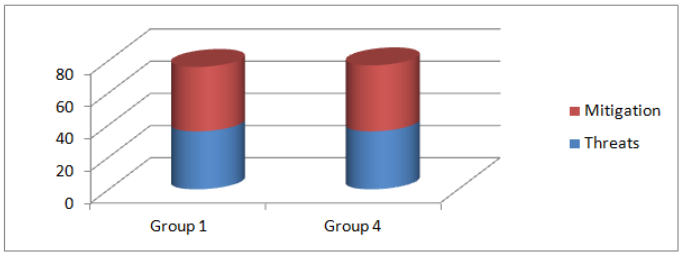

Figure 8.5: Secure Tropos with Net Shopping (Group 1 and 4)

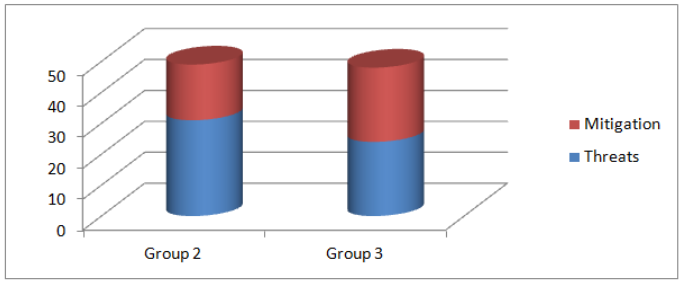

Figure 8.6: Secure Tropos with Health Insurance System (Group 2 and 3)

technique with group 1. Figure 8.6 expresses that group 2 and group 3 had the same experiment task in their second part of experiment due to the Latin-Squares experimental design (see table 6.3). The modeling technique was the same with group 1 and group 4 but with a different case - Health Insurance System. 24 threats and 24 mitigations were identified by the participants in the group 2. Meanwhile, 31 threats and 18 mitigations were identified by the participants in group 3.

For the second part of the experiment, group 1 and group 4 have used MUC modeling technique with Health Insurance System. 45 threats and 42 mitigations were identified by the participants from group 1, and 49 threats and 51 mitigations were identified by the participant from group 4. Meanwhile, it was the first part experiment for group 2 and group 3 who have used MUC modeling technique with Net shopping case. 22 threats and 18 mitigations were identified by the participant from group 2, and 29 threats and 36 mitigations were identified by the participant from group 3. The histograms show the number of threats and mitigations that the participants have identified in total. Each column is comprised of identified threats and mitigation together.

- **Estimating the usage of Textual Description, Diagrams and Memory** The participants' self-reported the estimated usage of textual description, diagrams and memory for each modeling approach they used and these numbers were also counted

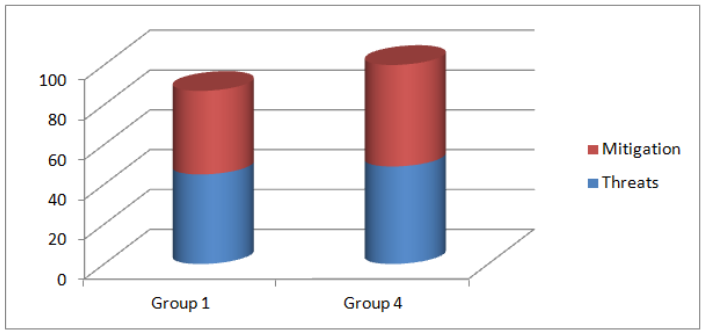

Figure 8.7: Misuse Case with Health Insurance System (Group 1 and 4)

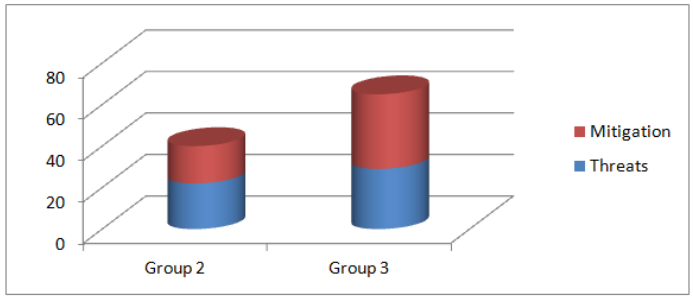

Figure 8.8: Misuse Case with Net Shopping (Group 2 and 3)

Identify threats through textual description = THR_TXT								
Identify threats through memory = THR_MEM								
Find mitigation for threats through diagrams = MIT_DIAG								
Find mitigation for threats through textual description = MIT_TXT								
Find mitigation for threats through memory = MIT_MEM								
Find mitigation for threats through only threats = MIT_THR								
Sequence ID	Student ID	THR_DIAG	THR_TXT	THR_MEM	MIT_DIAG	MIT_TXT	MITI_VUL	MITI_MEM
GROUP 1 <1>	tru	20	60	20	10	60	10	20
2	1441k	60	20	20	10	10	15	65
3	6	0	50	50	0	50	0	50
4	JB24	40	50	10	20	30	40	10
5	202	5	80	15	0	70	20	10
6	H33	30	50	20	50	20	20	10
7	G2K	10	0	90	0	0	0	100
8	VON	90	10	0	0	100	0	0
9	MJ14	10	70	20	70	15	10	5
10	88	40	40	20	30	10	50	10
11	winnowill	20	40	40	10	30	30	30
12	orange	0	50	50	0	0	0	100
13	~!!S	30	40	30	20	20	40	20
GROUP 4 <1	CYI	30	10	60	30	0	0	70
2	8526	30	40	30	30	50	10	10
3	1106	20	60	20	50	25	25	0
4	T	0	100	0	0	0	100	0
5	louie	0	80	20	0	80	0	20
6	HMO	5	85	10	0	0	0	100
7	3647	50	10	40	0	0	100	0
8	67	20	50	30	30	20	40	10
9	kena	0	100	0	100	0	0	0
10	A.N	0	40	60	0	10	10	80
11	47900	0	2	98	0	0	2	98
12	1105	10	70	20	20	50	10	20
13	333	25	25	50	20	25	15	40

Figure 8.9: Estimating the usage of Textual Description, Diagrams and Memory for Secure Tropos with Net Shopping Case

in this experiment in order to analyze the participants' perception as an evaluation methods. All data according to their responses were coded as the following tables in excel (see figure 8.9 - 8.12) for further analysis.

Figure 8.9 - 8.12 present that the participants' self-assessed how they have separated their work on each modeling technique they have applied in the experiment. The figures 8.9-8.12 illustrate the participants' responses in this experiment task. Both modeling approaches were estimated the usage of threats or mitigation by the participants in order to finish relevant tasks as mentioned in the experiment sheets (see Appendix A).

- **Post-experiment Questionnaire** There were two post-experiment questionnaires for the two parts of experiment. Both parts of experiment asked the participants to report their perception of the modeling technique. Histograms are used to illustrate the data. As it was mentioned in 6.2.3, the designed questions can be categoried into 3 classes. In appendix C - post-questionnaire, PU means perceived usefulness of the modeling technique, ITU means intention to use the modeling technique and PEOU means perceived ease of use the modeling technique. Since there were 3 negative questions in the questionnaire, therefore, in order to analyze them in statistical method, the values were inverted into reverse values by following the critical rule of "***Reverse value = 6-response value***".

Secure Tropos Health Insurance System. Group 2 & Group 3.								
Identify threats through diagrams = THR_DIAG								
Identify threats through textual description = THR_TXT								
Identify threats through memory = THR_MEM								
Find mitigation for threats through diagrams = MIT_DIAG								
Find mitigation for threats through textual description = MIT_TXT								
Find mitigation for threats through memory = MIT_MEM								
Find mitigation for threats through only threats = MIT_THR								
Sequence ID	Student ID	THR_DIAG	THR_TXT	THR_MEM	MIT_DIAG	MIT_TXT	MITI_VUL	MITI_MEM
Group 2 <1>	248357	50	30	20	50	10	10	30
2	!!!!	30	50	20	10	10	30	50
3	i1648	20	30	50	10	20	20	50
4		10	5	85	10	15	50	25
5	SH	40	50	10	30	40	10	20
6	123	25	75	0	20	40	40	0
7	MHH	40	20	40	30	15	15	40
8	he	25	60	15	10	15	25	50
9	APE123	100	0	0	100	0	0	0
10	Trallala	0	50	50	0	0	0	100
11	test	50	30	20	10	10	60	20
Group 3 <1>	C9A 1	40	30	30	30	15	5	50
2	#10	10	30	60	10	20	50	20
3	1247	5	80	15	0	30	30	40
4	1024	10	40	50	0	25	35	40
5	725	50	0	50	50	0	0	50
6		60	20	20	0	20	20	60
7	pfhoy	30	60	10	30	50	20	0
8	4747	80	5	15	10	10	0	30
9	029A	20	50	30	10	40	20	30
10	bo	80	10	10	60	10	20	10
11	hdr	0	80	20	0	80	0	20
12	AK	80	0	20	20	0	60	20
13	45	50	20	30	35	25	15	25

Figure 8.10: Estimating the usage of Textual Description, Diagrams and Memory for Secure Tropos with Health Insurance Case

Misuse Case_Health Insurance System. Group 1 & Group 4.								
Identify threats through diagrams = THR_DIAG								
Identify threats through textual description = THR_TXT								
Identify threats through memory = THR_MEM								
Find mitigation for threats through diagrams = MIT_DIAG								
Find mitigation for threats through textual description = MIT_TXT								
Find mitigation for threats through memory = MIT_MEM								
Find mitigation for threats through only threats = MIT_THR								
Sequence ID	Student ID	THR_DIAG	THR_TXT	THR_MEM	MIT_DIAG	MIT_TXT	MITI_VUL	MITI_MEM
GROUP 1<1>	tru	40	40	20	40	30	15	15
2	1441k	50	10	40	30	20	20	30
3	6	50	25	25	20	10	10	60
4	JB24	60	30	10	30	10	20	40
5	202	15	5	80	10	40	40	10
6	H33	70	10	20	80	0	0	20
7	G2K	70	10	20	50	0	20	30
8	VON	50	50	0	50	50	0	0
9	MJ14	80	5	15	0	0	100	0
10	88	80	0	20	70	10	10	10
11	winnowill	45	10	45	30	30	10	30
12	orange	60	10	30	10	10	40	40
13	~!!$	50	30	20	0	0	0	100
GROUP 4 <1>	CYI	0	0	100	0	0	0	100
2	8526	20	40	40	0	20	30	50
3	1106	60	30	10	30	25	35	10
4	T	80	0	20	10	0	80	10
5	louie	60	30	10	60	40	0	0
6	HMO	90	8	2	0	0	0	100
7	3647	50	30	20	10	5	50	35
8	67	30	0	70	40	0	0	60
9	kena	90	10	0	10	10	0	80
10	A.N	60	20	20	0	0	0	100
11	47900	2	3	95	0	0	0	100
12	1105	30	30	40	30	20	30	20
13	333	40	20	40	20	10	20	50

Figure 8.11: Estimating the usage of Textual Description, Diagrams and Memory for Misuse Case with Health Insurance Case

Misuse Case Ner Shopping. Group 2 & Group 3.								
Identify threats through diagrams = THR_DIAG								
Identify threats through textual description = THR_TXT								
Identify threats through memory = THR_MEM								
Find mitigation for threats through diagrams = MIT_DIAG								
Find mitigation for threats through textual description = MIT_TXT								
Find mitigation for threats through memory = MIT_MEM								
Find mitigation for threats through only threats = MIT_THR								
Sequence II	Student ID	THR_DIAG	THR_TXT	THR_MEM	MIT_DIAG	MIT_TXT	MITI_VUL	MITI_MEM
Group 2 <1>	248357	60	10	30				
2	!!!!	80	15	5	10	10	20	60
3	i1648	20	10	70	20	10	10	50
4		5	20	75	2.5	2.5	50	45
5	SH	70	20	10	10	50	10	30
6	123	50	40	10	35	35	20	10
7	MHH	15	40	45	10	40	25	25
8	he	40	40	20	20	10	20	50
9	APE123	0	50	50	0	0	0	100
10	Trallala	80	20	0	0	0	100	0
11	test	20	20	60	5	5	80	10
Group 3 <1>	C9A1	60	10	30	30	10	10	50
2	#10	20	40	40	20	10	50	20
3	1247	20	60	20	5	40	20	35
4	1024	0	50	50	0	10	10	80
5	725	70	0	30	10	0	0	90
6		0	20	80	0	10	20	70
7	pfhoy	20	60	20	20	40	20	20
8	4747	80	20	0	20	0	0	80
9	029A	30	50	30	15	45	20	20
10	bo	75	20	5	20	20	40	20
11	hdr	40	40	20	30	30	20	20
12	AK	0	80	20	0	0	80	20
13	45	40	40	20	30	20	30	20

Figure 8.12: Estimating the usage of Textual Description, Diagrams and Memory for Misuse Case with Net Shopping Case

8.1.2 Data Pre-analysis

There is always no standard answers that can provide the rules of removing data pointers because of some loose information or uncertain issues. This situation may risk the experiment results and impact on the conclusion. There are also no good results for any experiment by comparing the result based on a removed data with the ones that was not removed data. However, two separate ways of reducing data can be identified:

- Single data points can be removed, for example, outliers or
- All data can be analyzed and derive to the conclusion that due to high inter-correlation (Cohen 2002)[19] between some variables, in this case we should combine measures into some abstract measure.

But for this experiment, it was not obviously evident to support for removing any outliers, because the experiment sample size was not computed in advance. It was difficult to recruit people. Hence setting the total number of the participants was not a necessary task for the experiment. Thus, all data was treated as valid for the experiment. Another issue was the experiment just focus on both modeling diagrams, thus the main analysis was not the textual description and memory used in the experiment, even so, all these variables were analyzed in order to derive the conclusion from a multiple view of angle for the experiment. Hence the useable data was not only estimated usage of diagrams but also the variables of MIT_THR, MIT_TXT, and MIT_MEM used in the result analysis. Another reason why all variables were analyzed since there might be risks for the experiment if these variables were removed because considering the participants' background, they had

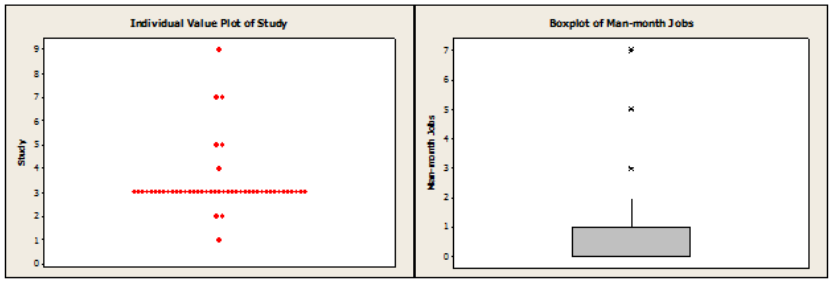

Figure 8.13: The outlier points of their study background and man-month jobs

limited experience on relevant modeling techniques study. For example, only the UML use case have been taught in their software engineering class and some non-avoiding risks (see chapter 12) in this experiment. Combined with all these factors in the experiment, all the variables were maintained during the data analysis. The comparison of results came from the non-removed variable set and how the results were analyzed are explained in next section (see 8.2).

The box plots (figure 8.13) were used to express the new data sets. Most of participants had no ICT relevant work experience according to the right boxplot figure but there were three outliers extremely higher than average value. The left figure shows that most of the participants had 3 semester studies at ICT.

There were 50 participants in total; all the participants were grouped into four groups where they have tried to use both modeling techniques with two cases. The responses were coded qualitatively in order to determine exactly how many threats and mitigations does each participant can identify. The counting rule for the threats and mitigation was that if the participant was using a conjunction word such as “and”, “or” conjunction word to conjunct several threats or mitigations, this kind of threat or mitigation was counted as two threats or mitigations in this experiment. Also, if the mitigation was suggested to several threats at the same time, this mitigation was counted several times according to which threats it relates to. When compared the participants’ background and the number of threats and mitigations they found, there was a simple figure 8.15 as follows which were illustrating that the participants’ performance were based on their background of relevant ICT work experience for the ST modeling technique. The data distribution was based on the background of the participants, the right figure shows the mean value of the threats they have identified, the left figure illustrates the relevant mitigations that refer to the threats which the participants have identified relate to their man-month work experience. All the participants also tried Misuse Cases diagrams modeling approach as their tasks in the experiment. According to the experiment design in order to reduce the learning effect during the experiment, each group had to apply different modeling technique-with-case regarding to their group number, for example group 1 and group 4 tried to use Secure Tropos diagrams first but they had different case order (NS or HIS) in the experiment. Figure 8.15 presents that the performance for identified threats and mitigations by two paired groups.

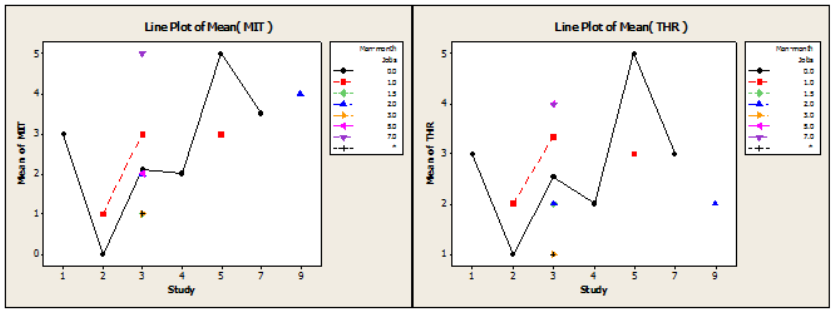

Figure 8.14: The mean value for Identified Threats and Mitigation compared with participants' background.

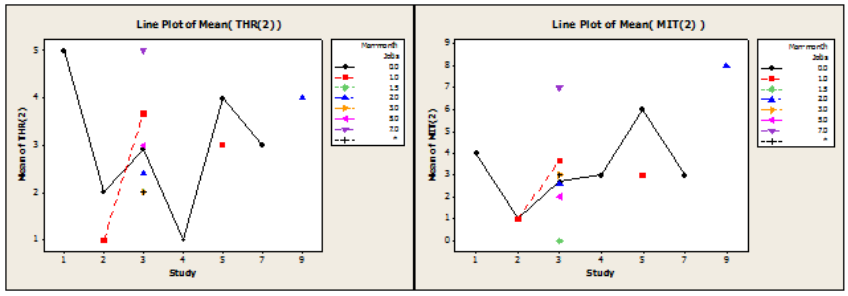

Figure 8.15: The mean value for Identified Threats and Mitigation compared with participants' background.

As well as the figure illustrating in Group 1 and Group 4, the left figure means that the threats were identified by the participants relate to their job experience. And the right figure means that the mitigations number refer to the number of threats were identified relate to their background - the man-month work experience. The estimated the usage of the textual description, diagrams and memory were designed to collect the response of how the participants use diagrams, textual description or memory to identify threats and suggest the mitigations. Hence, it was very important to compare how many percentage they have used of the two modeling methods to identify threats as expected. The following figure 8.16 gives visualization representative of these data that collected from each group. As mentioned above, whether the participants through modeling diagrams to identify threats or suggest mitigations were the most important issue for the results analysis, thus the issues of showing probability of textual description, memory they have used and other concerns in a plot figure was omitted here. The figure 8.16 illustrates that detected probabilities from each group for two modeling techniques in the two parts of the experiment according to the participants' self-assessments. The data were distributed from 5% to 95%, for the each modeling approach the data was distributed vary widely but most of them were range

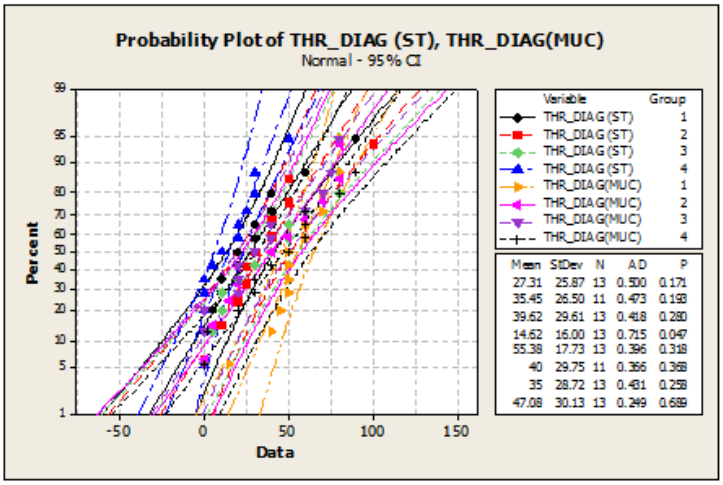

Figure 8.16: Probability plot of THR_DIAG(ST), THR_DIAG(MUC)

from 20% to 70% according the participants' response. This means that the modeling diagrams were useful for the participants to finish relevant tasks in the experiment due to their background with few experience on these two modeling approaches. But for the each individual, there was different percentage value estimated for using modeling diagrams. Hence, a further analysis was needed (see 8.2). The post-experiment questionnaire was used to collected data according to the participants' self-assessments, it represented their response as illustrated above (see section 8.1.1). Since the TAM (Davis 1989)[20] was used to design style to record the participants' self-assessed perception of the modeling approaches in the experiment, so no need to reduce any data from these data sets as well.

8.2 Data Analysis

8.2.1 background

Kruskal-Wallis H tests[3] was used for the four independent groups to test the participants' background. All background variables were tested and there was no significant different between the groups with respect to knowledge, study semesters and job months (See Appendix B). The 2-tailed Wilcoxon sign-rank tests[11] was also used to compare the participants self-assessed knowledge background in the areas of Use Case v.s. i* modeling diagrams, Misuse Cases v.s. Secure Tropos diagrams and Activity diagrams v.s. Mal-activity diagrams. According to the results (see appendix B), no obviously significant difference of background between the participants was found that they had more knowledge about UML use case diagrams and UML activity diagrams other than i* modeling diagrams and Mal-activity diagrams. According to the analyzed results, significant differ-

ence was found in the knowledge of Misuse Case diagrams and Secure Tropos diagrams (p = .009).

Most participants reported 3 semesters of IT studies, with a mean of 3.3 and a higher standard deviation 1.27 due to the outliers with 5, 7 semesters. They reported 0 - 2 man-month jobs or internships besides 2 outliers with 5, 7 months. Here the mean was 1.26.

Almost all of them were at the same study level - sophomore students, so they have limited chance to find a relevant ICT work in summer since Norwegian companies prefer the interns in or above their third year studies at the higher education level. According to their responses data and the analyzed results, it was believed that the participants had no significant difference between the study background and their knowledgeable about secure modeling approaches. This situation was not a surprise since the participants were chosen randomly whom study computer science at the IT department.

8.2.2 Performance

The data collecting was done through asking the participants to identify and list as many threats and mitigations as they can in the specific systems of the experiment. The Wilcoxon signed-rank tests[11] of two paired groups for the variables THR and MIT was used, using exact 2-tailed significances to compare how well the participants have identified threats when they have applied ST and MUC. The average of identified threats by applied both modeling techniques were 2.54 for Secure Tropos and 2.90 for Misuse Cases (see table 8.1). According to this result (Appendix B), there were no find significant differences in the number of identified threats between the two modeling approaches when they have applied techniques with the different cases order(p = .113). Hence, it derives a conclusion that there were no significant differences of the two modeling techniques to identify the threats when the participants have applied to the two modeling techniques with relevant cases.

However, on the contrary, it was found that a significant difference in the number of identified mitigation between the two modeling approaches (p = .005). Also there were significant differences in the number of identified threats and mitigations between the two modeling approaches (p = 0.004). Hence, to sum up, the conclusion can be derived that H1 was rejected and H2 and H3 were accepted.

Wilcoxon signed-ranked test[11] was also used to compare the participants' performance for the first technique-and-case pair the second technique-and-case they used. All the variables such as THR, MIT and THMI were tested here. Through analyzed the results, there were no significant differences in the number of identified threats between the first half of the experiment and the second half (p = .076). However, there were significant differences in the number identified mitigation between the first half of experiment and the second half experiment (P = 0.014). It may be due to frigates during the experiment because there were only 10 minutes for the participants to have a rest between the two parts of experiment.

Table 8.1: Comparison results for performance

Identification task	MUC Mean	MUC St. Dev value	ST Mean	ST St. Dev value
Threats(THR)	2.90	1.39	2.54	1.30
Mitigation (MIT)	2.94	2.03	2.27	1.51
Both(THMI)	5.84	3.02	4.80	2.32

8.2.3 Estimating of the usage of diagrams, textual description and memory

There was a completely blank sheet of responding on estimated used diagrams, which came from the group 2. Wilcoxon signed-rank tests[11] was performed for the two paired groups with exact significances and the two paired groups were applied for the variables concerning the estimated use of diagrams when solving the performance task. The results were presented in table 8.2. Except testing the variables of THR_DIAG and MIT_DIAG, all other variables were also tested too (see Appendix B). According to the results, there were significant differences between the use of textual description to identify threats (p = .001) and using memory to suggest mitigations (p= .025). When the participants have used textual description to suggest mitigations, it was obviously that there were significant differences between Secure Tropos diagrams and Misuse Cases diagrams (P = .013).

Meanwhile, two cases were compared in order to see if there were any significant differences between the two modeling techniques when the participants tried to estimate the use of diagrams, textual description and memory in the experiment. There were significant difference of using diagrams to identify threats between Secure Tropos and Misuse Case (p = .035) for the NS case but no significant differences were found between the modeling techniques for the HIS case (p = .065). When using textual description to identify threats, there were also no found that the participants have displayed significant differences in both two cases - NS and HIS. But they have used the same cases with different modeling techniques, there were significant differences between the two modeling techniques with the same cases (NS p= .050, HIS p = .011).

In order to increase the reliability of the assessment for two cases which were used in the experiment, from the opposite view of the estimated the usage of two modeling techniques, the cases with the same security modeling techniques were also tested in the stage of data analysis. For example the two cases with ST modeling technique were tested in order to see whether there were significant differences between the participants' performance for the cases. It was obviously found the significant difference between two cases with the modeling diagram to identify threats(p = .025) and signigicant difference between two cases with textual description to identfy the threats (p = .007). This may due to the complexity of the case, thus this was the reason why a further analyzed the data of participants' perception was need (see 8.2.4).

Table 8.2: Estimating the usage of diagrams, textual description and memory

Variables	MUC Mean	MUC St. Dev value	ST Mean	ST St. Dev value
THR_DIAG	0.54	0.02	0.29	0.26
THR_TXT	0.25	0.19	0.41	0.28
THR_MEM	0.31	0.25	0.30	0.23
MIT_DIAG	0.19	0.19	0.21	0.24
MIT_MEM	0.41	0.32	0.33	0.31
MIT_THR	0.24	0.26	0.21	0.27
MIT_TXT	0.15	0.16	0.23	0.24

8.2.4 Perception

Wilcoxon signed-rank tests of two paired groups were applied for all perception variables using 2-tailed significances to compare how the participants perceived the techniques in terms of usefulness (PU) and ease of use (PEOU) and whether they intended to use them again in the future (ITU). The results show that the participants perceived Secure Tropos significantly more positively than Misuse Case (see table 8.3), perceived usefulness (p= .000), perceived ease of use (p= .000) and intention to use (p = .000) and average for all perception questions (p = .000). Hence, hypotheses 6, 7, 8 and 9 were all confirmed.

Meanwhile, all individual question in this section was tested as well. A significant result for the Wilcoxon signed-rank test was obtained that the participants were in favor of ST, except question 5, this did not give a significant result in either direction for both modeling approaches either. This was related to the PU variable and stated that the technique "would be useless in finding system security issues" (p = .054). All other statements were found significant difference between the two modeling approaches when the participants applied during the experiment (see Appendix B). The acceptable experiment type I error probability was 0.05 and the type II error probability was 0.20. Since the experiment compared two modeling technique groups, thus N/2 units were needed in each group (Hopkins 2001)[36]. According to the calculated effect size in table 8.4, the relevant characteristics were confirmed, such as the pooled standard deviation[6], effect size[17]. Because there were two modeling approaches in the experiment, it was therefore that the means-Wilcoxcon-Mann-Whitney test [65](two groups) could be used (Faul 2009) (Faul 2007)[28] [29]under the t-test family to compute the determined sample size. Two groups construct the total sample size, so each of the group has a half number of total samples.

Wilcoxon signed-ranked test was also applied to compare the participants' average perceptions for the HIS case and NS case when compared the first technique they applied with the second one. Significant differences were found that the participants significantly favored Net Shopping case (P = .001) when they have applied Secure Tropos diagrams. Their responses were that it was easy to use MUC when they used case Net Shopping (p = .011). Combination of the analysis for estimating of usage of textual description, diagrams and memory where the participants' self-report in the experiment, the conclusion for this kind of issues can be explained that the Net Shopping case was closer to the participants' daily life and they may touched this kind of transaction frequently.

Table 8.3: Perceived usefulness (PU), ease of use (PEOU), intention to use (ITU)

TAM Variables	MUC Mean	MUC St. Dev	ST Mean	ST St. Dev	Sign. (exact)
THR_DIAG	0.45	0.02	0.29	0.26	p=.003
MIT_DIAG	0.19	0.19	0.21	0.24	p = .634
PU	2.50	0.36	3.01	0.42	p=.000
PEOU	2.18	0.54	3.06	0.70	p= .000
ITU	2.13	0.54	3.02	0.72	p = .000
AVE	2.34	0.51	3.02	0.62	p= .000

Table 8.4: Effect and sample size for the significant differences in perception (SM: samll, MED: medium, Lar: large)

Variables	Pooled st.dev.	Effect size	Cohen	Hopkins	Sample size required
PU	0.39	-1.30	Lar	Lar	22
PEOU	0.63	-1.40	Lar	Lar	20
ITU	-1.39	Lar	Lar	Lar	20
AVE	0.57	-1.22	Lar	Lar	26

Chapter 9

Experiment - Findings

This chapter discusses the results by summarizing and interpreting the main findings in chapter 8. A possible further development of techniques with classification of threat types are also suggested in this chpater.

9.1 Main findings

Table 9.1: Result of Hypothesis testing

ID	Description	Result
$H1_1$	There will be a significant difference in the number of identified threats between the two groups.	Reject
$H2_1$	There will be a significant difference in the number of identified mitigations between the two groups.	Accept
$H3_1$	There will be a significant difference in the number of identified threats and mitigations between the two groups.	Accept
$H4_1$	There will be a significant difference in the estimated use of the diagrams of the two techniques in identifying threats.	Accept
$H5_1$	There will be a significant difference in the estimated use of the diagrams of the two techniques in identifying mitigations.	Reject
$H6_1$	The two techniques will be perceived significantly differently regarding usefulness.	Accept
$H7_1$	The two techniques will be perceived significantly differently regarding ease of use.	Accept
$H8_1$	The two techniques will be intended to be used significantly differently.	Accept

Continued on next page

Table 9.1 – continued from previous page

ID	Description	Result
$H9_1$	The two techniques will be perceived significantly differently.	Accept

Table 11.1 summarizes the results for the main hypotheses. As it was mentioned in previous chapters, this experiment comparison was indicating that the two techniques were complementary in terms of comprehensive understanding.

The subjective was sophomore students at the department of Computer and Information Science, NTNU. The reason why they were chosen as the participants can be discussed into two parts.

Firstly, since they were sophomore students who have been taught software engineering, thus they had the knowledge of UML use case modeling technique, and they had experience on using this technique in their exercise or project in their course works. Because they were in the same organization in the department of XCom'13, according to the outline of bachelor education, they may had very similar education background in the university's level of studies. Thus, it was easier to control their background if they have been chosen as the experiment sample. Even through some participants had experience on relevant ICT internships, but according to results analysis there were no signification differences between their background in knowledge of determining the experiment modeling techniques.

The second reason why they were chosen as the participants since it was also easier for us to motivate them to participate in the experiment because they always secure founding to extrusion, we can offer some money to them, hence the participants had higher motivation to took in the experiment that was really good news for the experiment.

Participants prefer to use Secure Tropos other than Misuse Case. However, there were no significant differences between the two modeling techniques to identify threats (p= .113). On the contrary, the participants suggested significant differences in the number of identified mitigations between this two modeling techniques (p= .005). This might be caused by the cases complexity when they tried to use the two modeling techniques (see 8.2). As it was planned, the next step was data analyzing from the responses of the participants estimated usage of textual description, diagrams and memory. According to the analyzed results there were significant differences between the use of textual description (p=.001) and memory to identify threats (p=.025). There were also significant differences between Secure Tropos and Misuse Case when the participants using textual description to suggest mitigation (p=.013). However, when compared the cases where the participants have used textual description, diagram and memory to identify threats in the addition of the same modeling technique with the same case, it was obviously significant differences between Secure Tropos and Misuse Case diagrams for Net Shopping case (p= .025). There were also significant differences between ST and MUC when the participants have used textual description for Net Shopping case (p= .007).

Therefore, the conclusion can be derived that Net Shopping case might be easier for participants to identify mitigations since this case was closer to their daily life. And it was also can concluded that Secure Tropos was preferred by the participants in the following several reasons. Secure Tropos is a kind of goal-based modeling technique; the participants could identify their security goals if they read the textual description of cases carefully, they can find the system's goals, sub-goals and the constraints between them. Then they might suggest high level threats and mitigation even though they had not focus on details of modeling diagrams. However, for Misuse Case, after they read the diagrams, participants can identify most of the threats and mitigations since the diagrams were illustrating systems' threats in low level, they can master this technique refer to their experience on UML use case. They might less focus on the textual description when they have identified threats or mitigations. Hence, when the participants have suggested mitigations, they may impacted by the characteristics of the two modeling techniques. Therefore, there were significant differences in using the two modeling approaches and their perceptions on the two modeling techniques. And according to the analyzed results, ST was received more intent to use in the future, the mean value was 3.02 with the standard deviation was 0.72. The mean value of MUC was 2.13 with the standard deviation was 0.54.

9.2 Other findings

According to the participants' response, there was another reason that also might impact on the experiment results, such as the threats and mitigation types. After looking though all the experiment sheets the participants filled in. An interesting thing was found according to their responses on each modeling technique with relevant cases. Hence, it was worthy to categorize their answers in order to find some valuable suggestions for both modeling approaches.

A tree figure 9.1 illustrates that the responses can be divided into two main classifications based on the type of threats or mitigations. It was found that for Secure Tropos diagram, most of the participants suggested threats and mitigations were based on the non-technique aspect. On the contrary, the participants suggested most of threats and mitigation for Misuse Case diagrams were based on technique aspect. Thus, the definition here are used to classify the technique and non-technique responses:

Technique aspect response The student gives the response to identify the threats and mitigation based on technical description such as authorization, biometric technique, availability and confidentially. E.g., the server shall be able to identify the attacks and can protect it for avoiding crash, or when pay the payment to bank system, the technique shall use "https" to protect the data communication and so forth.

Non-technique aspect response The student suggests that those threats and mitigation based on integrity, human mistakes and physical errors. E.g., HIC makes mistake to update the database that will inform the patient to pay the payment.

However, the quality of the threats were hard to control since there was no standard template for the participants to use when they identify threats for both two cases, therefore,

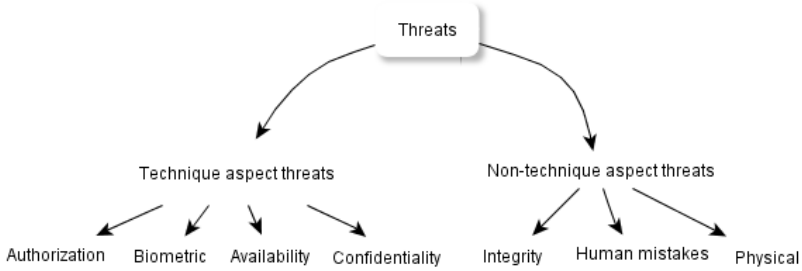

Figure 9.1: The main categories for threats of the experiment

it was difficult to evaluate their responses due to the vary evaluation criterion of different people. Hence, their responses were just classified in a high level category and could not go into the details or sub-categories. However, this finding was still sprit the idea that the further research direction will be correct according to the statistical analysis for the future experiment. Although the results were in many ways as expected, it still feels useful to have such suspicious confirmed by controlled experiments rather than letting them remain on the purely speculative level. For instance, the boilerplates [60] can be used for helping the participants to organize their response of threats and mitigations in a standard way, this will be easier for researchers to evaluate and classify their responses in details and also can be used to confirm more lower level threats or mitigation types of the two modeling techniques.

Furthermore, as shown in the table 9.1, that results of this experiment were not particularly surprising. Secure Tropos was better for the identify threats and mitigations of a system from the high level of pre-requirements and requirements for the system security analysis, and Misuse Cases was better for system's low level requirements analysis, especially the technique aspect before a real implementation such as the stage of system design as suspected. As for the participants' preference towards Secure Tropos, there were several possible issues of occurring of this situation, this may plausibly be explained by the fact that the students were like analyze the possible functional of a system in the stage of pre-requirements and high level requirements but no need to go further of the low levels security analysis. If they do not need to further analyze the cases but just finish an experiment, the MUC may just considered as non-prior for the analysis of a system due to the request by experimenters. However, both modeling techniques had no significant different to identify threats, it can be explained that the students had few experience on both modeling techniques, the experiment was the first time to let them know these modeling approaches for analyzing the security concerns. Hence,, when the students have identified the mitigations for threats they found, the mitigations might not only based on their feedback on the estimated the usage of diagrams, textual description and memory, but also their experience, such as some feedback suggested from biometric technique and non-technique areas.

Chapter 10

Experiment - Findings Discussion

In the previous project work, six dimensions were classified for the two modeling families-i*-based and Use Case-based modeling techniques[53]. A discussion of the results found in the experiment was compared with the previous classification (see chapter 3), to see whether the findings in this experiment were also claim the same idea in the previous project work.

10.1 Representation perspective and level of abstraction

As well known that ST is a Goal-based modeling, it is used for modeling and analysis of the system's security constraints as the attainment of security goals. On the contrary, the MUC is a Model-based or Problem-based modeling technique. MUC is a kind of modeling notation based on the UML or variants, or extensions of it, and probably allows modeling and analysis of different threats using formal or informal means to identify threats and possible mitigations to the threats[54].

Considering the number and type of the threats and mitigations identified by the participants when they have applied the two modeling approaches and their self-report background in the experiment, there were no significant differences found between the two modeling techniques when they have identified number of the threats (p= .113). However, an interesting finding was that there were significant differences in the number of identified mitigation between the two modeling approaches(p= .005). According to the answer of their responses (see section 9.2), it was found that the participants gave more high level suggestions for the case with ST modeling technique. But there were more technical suggestions for the case with MUC modeling technique. All of them had been taught the UML use case modeling technique in their software engineering course before they took part in the experiment. Thus, they may be more familiar with the Use Case-based modeling techniques for analyzing the system.

It was not surprise that they suggested technical solutions for the cases with MUC modeling technique. Hence, compared with the previous work, it can be confirmed that these findings claim as the same in the dimension of characterization of "representation perspective and level of abstraction" : the ST modeling technique is a Goal-based modeling technique, practitioners may more focus on the security goals of a system, but when they have to identify threats from technical view of a system or low level analysis of a system, Model-based or Problem-based modeling technique might be a good choice, such as MUC.

10.2 Kind of SRE tasks/activities

The activity classes were extracted from the recent 10 years work by several scholars[46][26] [59][31][47][41][42][45]. The following paragraphs illustrate each sub-class that whether the two modeling approaches' benefits were confirmed by the experiment or not. The table summarizes the characteristics were confirmed in the experiment.

Table 10.1: Confirmed characteristics in the dimension of "kind of security requirement engineering (SRE) tasks/activities"

Characteristics	**Confirmed**	**Not confirmed**
Security Objective	Both ST and MUC were confirmed	None
Identification and modeling of Assets (IMA)	ST: not support IMA, MUC: support IMA	None
Identification of Vulnerabilities(VT)	Both ST and MUC were confirmed	None
Elicitation and Analysis of SRs(EAS)	Both ST and MUC were confirmed	None
Specification of SRs (S)	Both ST and MUC were confirmed	None
Documentation of SRs (DS)	Both ST and MUC were confirmed	None
Verfication and Validation support (VVS)	Both ST and MUC were confirmed	None

1. *Security Objectives* presents the high-level requirements or goals that are most important to customers. According to the participants' performance in the experiment, even though there were no find significant differences between the two modeling techniques when they have identified threats, it was still noticed that the participants stand on the customers behalf to analyze the cases when they used ST, such as their responses of mitigations for the threats they found and its types. MUC was also fulfilling with this class since the threats and mitigations that the participants have found were also focus on the security goals for the cases.

2. *Identification and modeling of Assets (IMA)* presents the entities that are considered valuable, and a possible target of threat, should be protected. The participants behavior were the same as claimed in the previous work that ST was not support this feature but MUC supported, according to their non-technique threats suggestions for ST with two cases. But they gave technique threats and mitigation suggestion

for MUC with the two cases. Thus, MUC can fulfills with this requirements.

3. *Identification of Vulnerabilities and Threats (VT)* presents the ability of identification of software defects that can be exploited by an attacker to cause harm, and what could be an attacker's target. After analyzed the responses from the experiment sheets, it was realized that the participants can just identified threats when they have applied MUC but no potential threats were suggested when they have applied ST. This also confirms the previous claim that ST was not support VT but MUC just supported to identify threats.

4. *Elicitation and Analysis of SRs (EAS)* presents the ability to derive security requirements from other sources and to analyze the security requirements to achieve complete and unambiguous requirements. It was found that there were significant differences between the two modeling techniques when the participants' have applied diagrams to identify the threats (p=.003) but no significant differences when they have used diagrams to suggested mitigations (p= .643). Except the diagrams, the textual description and memory were also the extra sources for the participants to analyze the security requirements. When the participants have used textual description to identify threats, there were significant differences between two modelling approaches (p = .001). When the participants have used memory to identify mitigations there were significant differences between two modelling approaches (p = .025). When the participants have used textual description to identify mitigations, there were significant differences between two modelling approaches (p = .013). Although there were differences between the two modeling techniques during the experiment execution, both of the two modeling techniques can fulfill with this classification due to the extra sources to support analyze the security requirements to achieve complete and unambiguous criterion.

5. *Specification and Documentation of SRs (S and DS), Verification and validation support (VVS)* presents the use of informal or formal methods to document security requirements developed through elicitation and analysis, and its ability to document requirements in a way to ensure visibility, traceability. Formal methods or other verification and validation approaches guarantee that the security requirements are correct, consistent, complete and unambiguous. According the participants' responses (their self-assessed, number of identified threats or mitigations), it was obtained that the participants understood the two modeling approaches generally and they could gave their suggestion of threats and mitigations regarding to the modeling approach with cases that they have used in the experiment. As claimed in the previous work, it was verified that both modeling techniques can fulfilled with S and DS, but for the VVS the participant suggested more threats and mitigations from technique aspects, so the security requirements may just be able to understand by technique staffs instead of non-professional stakeholders of the system. Hence, it was confirmed that MUC can only support internal verification of VVS and ST cannot support the VVS. Since the identified high level and non-technique threats or mitigations cannot guarantee the requirements' correct, consistent, or complete and unambiguous.

10.3 Technical criteria and specification criteria

Technical criteria and specification criteria are closely related because in order to fulfill technical criteria all specification criteria related to that technical criterion must be fulfilled as well. Since further analyzing of type and quality of the threats and mitigations were not analyzed, it was too difficult to control the responses quality of the participants and the responses were range vary from non-technique aspects to technique aspects. For instance, human behavior mistakes in the HIS case like the doctor make a mistake to diagnose a patient health status, and biometric technique have been used in the case of NS to avoid salesperson sales goods to himself or herself. Hence, just a rough classification of types of threats and mitigations were given in the experiment results.

Furthermore, the experiment did not ask the participants to draw the diagram instead of asking them to self-assess the perception of the two modeling techniques. According to their responses and the experiment designer's (author's perception) during the design process of the experiment, the criteria in the dimension of "Technical and specification criteria" was partially verified. Others criteria might need a further experiment to test, such as "Standard integration", "Support for other development stages". The table 10.2 illustrates the confirmed criteria in the experiment.

Table 10.2: Confirmed characteristics in the dimension of "Specification and technical criteria"

Characteristics	**Confirmed**	**Not confirmed**
Internal verification support	Both ST and MUC were confirmed	None
External verification support	Both ST and MUC were confirmed	None
Support documentation generation	Both ST and MUC were confirmed	None
Standards integration	None	A further confirmation is needed
Requirements reuse	Both ST and MUC were confirmed	None
Support for other development stages	None	A further confirmation is needed

Continued on next page

Table 10.2 – continued from previous page

Characteristics	Confirmed	Not confirmed
Help support and easy to use	Both ST and MUC were confirmed	Participants did not state any easy to use model from the two modeling techniques

10.4 Modeling Language, Process and Method

This dimension is capable of promoting understanding of how an initiative can be applied. Through processing of the two case textual description and pre-drawn diagrams for the two cases, the following criteria were verified for both modeling techniques: *ability to formulate basic security requirements, usage scenarios, similarity with software specification languages and tool support*. As the dimension was cited in the previous work, MUC has no ability to represent security mechanisms and low level security requirements. For example when the buyers pay the payment to the system, the security criteria for password, account number and user name verification cannot illustrate in a detail diagram but the ST can fulfill this criteria. Due to the design constraints of the experiment, it was difficult to verify the rest criteria, such as "reuse of provided artifacts in later phase", "development resources", "reusable artifacts" and "usage in the industry". Certainly, these stage of points were also not the main concerns of this experiment either.

Table 10.3: Confirmed characteristics in the dimension of "modeling language, process and method"

Characteristics	Confirmed	Not confirmed
Ability to formulate basic security requirements	Both ST and MUC were confirmed	None
Ability to represent usage scenarios	ST: not support IMA, MUC: support IMA	None
Ability to represent security mechanisms and low level security requirements	Both ST and MUC were confirmed	None

Continued on next page

Table 10.3 – continued from previous page

Characteristics	Confirmed	Not confirmed
Similarity with software specification languages	Both ST and MUC were confirmed	None
Reuse of provided artifacts in later phase	None	Both ST and MUC were not confirmed
Tool support	Both ST and MUC were confirmed	None
Development resources	None	Both ST and MUC were not confirmed
Reusable artifacts	None	Both ST and MUC were not confirmed
Usage in the industry	None	Both ST and MUC were confirmed

10.5 Relevant SRE notation and Software Evolution and Other Perspectives

The case description and pre-drawn security modeling diagrams were illustrated fully of the SRE notations as classified in this dimension. Since the experiment did not ask the participants to identify the system's vulnerabilities but just asked them to identify the threat and its relevant mitigation. Thus, the vulnerabilities criteria cannot be investigated here. But others can be investigated by the experiment design and the participants' performance and perception in the dimension of relevant SRE notation (see table 10.4).

Meanwhile, according to the design of the case and its security modeling diagrams, the perception of the modeling techniques can help to make a judgment of software evolution and other perspective dimensions. For instance, the modularity is a measure of the separation of concerns and possibility of developing software components independently and its application. Both modeling techniques were not good at this feature, because ST diagram combined all security constraints together, so it is hard to only consider one constraint individually. MUC was better than ST but still cannot fully support developer only to consider some separately security goals.

The criteria of component architecture that is a level of support for a component-based structure. It allows software modules to be added or removed easily. It is impossible to add or remove any modules from the ST diagram since the ST diagram may lose security constraints, also cannot guarantee the secure entity if doing so. It will destroy the exits security constraints and the security logics of a system. The same with MUC, if trying to do the same behaviors in a MUC system diagram, although it is still possible to add or remove one individual misuse case diagram from the entire misuse case diagram of a system, it is still harm to entire security of the system.

As analyzed in the component architecture, ST modeling technique is low support to "change propagation", since it is hard to change or remove any constraint in an entire ST diagram. Therefore, the ST has low ability to keep track of changes made to system and only can limited guarantee the change is correctly propagated, such as no inconsistent dependency was left unresolved. For example, if we change one place in the system, it will impact on others in the system's security constraints or damage the whole system's security constraints. However, the MUC can low support component architecture, but there is a possible that the changes in the system cannot be tracked correctly.

Due to low support of modularity for a system, ST has low ability to evaluate the effect that changes made to specific artifacts. The same, MUC is medium support modularity of a system, thus analyzing the change impact of a system is a possible thing.

Table 10.4: Confirmed characteristics in the dimension of "Relevant SRE notions and Software evolution support"

Characteristics	**Confirmed**	**Not confirmed**
Security goal	Both ST and MUC were confirmed	None
Security requirements	Both ST and MUC were confirmed	None
Specification	Both ST and MUC were confirmed	None
Stakeholder	Both ST and MUC were confirmed	None
Asset	MUC was confirmed	ST was not confirmed
Threats	Both ST and MUC were confirmed	None
Vulnerabilities	None	Both ST and MUC were not confirmed
Risks/Mitigation	Both ST and MUC were not confirmed	None

Chapter 11

Suggestions based on the Results of the Experiment

11.1 Suggestions for the Two Modeling Techniques

The first suggestion based on the experiment results is to combine the Secure Tropos with other related modeling techniques. In this way it can provide complete security models for the system development. The entire system development duration can be divided into: identify requirement, identify scenarios of the system, design of the system, implementation, testing, and evaluation. As indicated, ST is used for the stage of analyzing the system, such as identify security requirements, and identify scenarios of the system. For instance analyst can through the textual description of the system to obtain the information and extract information from the meeting with the customers. As expressed in the experiment results, the participants preferred Secure Tropos, since the Secure Tropos expressed the system's security issues directly. As one of the i*-based modeling techniques, ST has good ability to involve all the stakeholders of the system. Non-professional staff can also participant in the system analysis, hence it can help to avoid the systems' non-usability if the system is just analyzed and developed by IT professionals.

Hence, if using Secure Tropos to analyze the system in the stage of system security requirements, boilerplate as a method to illustrate the security requirements in general natural language of the system should be a better way for the system designer to design the system other than only use the Use Case-based modeling techniques. And it is also useful when using ST to illustrate the systems' security constraints in the ST diagrams. Through this way, the development team can create systems' scenarios and model the scenarios that are face to both the stakeholders and the developers. After pre-analyzing the system's security concerns from multiple stakeholders of the system, it would be easier for developer to understand the requirements precisely and to do a further action, such as how many security constraints in the system, and how many security concerns should be considered

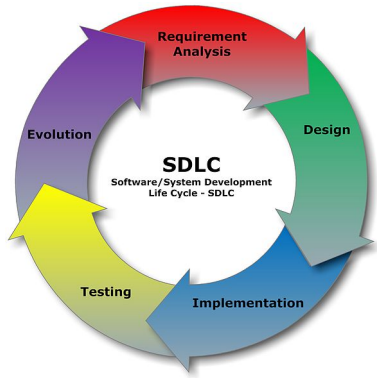

Figure 11.1: Life cyclic of system development[9]

before the project implementation.

The second suggestion is that because the UML sequence diagrams (USD) are very well known. Thus, if Secure Tropos can be involved into the model-driven development of system to generate relevant derivative framework, it would be also easier for developer to reduce the programming time. Furthermore, MUC is better than ST in low-level security requirements analysis, ST also medium support the criteria of modularity in the software evolution. Therefore, it is good for further analysis and design the system based on the first step by using ST method in the early and late stages of the requirements process, and just leave the rest things to MUC. MUC can be used to analyze and design the security constraints and its diagrams. It is also ensure that test-driven development (TDD) and agile development methods can be used in the software development.

Third, the main disadvantage of MUC are that MUC is based on informal analysis, hence MUC has no clear semantics; there is also no formal analysis of MUC. How to write a good quality MUC is also not clearly indicated. Contemporary, the technique is only focusses on the system-to-be, so it is not suitable for identify all kinds of threats, such as the non-technique aspect threats. The identifiable misuser and misuse case may not always consist of an identifiable sequence of actions. Consequently, a guideline for writing a good quality MUC is very usefulness. With good quality MUC, the ST can be used for help to illustrate the security mechanisms and low level security requirements. And other good features of ST, such as structural, exclusive and heavy weight development processes, and relevant formal approach, that can ensure the security concerns of a system could be more clearly. Therefore, they also indrectly ensure that the misuser and misuse cases may also can identifiable and accurate.

11.2 Suggestions for the future work

Yet, further investigation is needed to explore how well the Secure Tropos modeling technique combines with Misuse Cases modeling technique to develop the system/software, like a case study. The results still need to be tested in the future. Assuredly, there is still need a tool to support and realize the goal of the further project. Since there is already developed a tool can be used for analyzing the security requirements in the "ReqSec" project where researcher post-doctor Olawande Daramola preformed an experiment to test that tool. The method he used to identified security requirements was boilerplate[1]. Hence, If these "standard" security requirements can use in Secure Tropos diagrams or from Secure Tropos diagrams drive the relevant security requirements in natural language as the boilerplate template as a format for the threats identification. It will be a good idea to help the system designer and analysts to draw the related use-cases diagrams to help a good understanding of the system security concerns in the develop team. Also, it is a good opportunity for researcher to identify the types of threats or mitigations are found by the participants. Even through the software is still under construct, if i*-based modeling technique and Use Case-based modeling technique can be embedded, the system development could be more complete from analyze requirements to system design. That might be help to avoid both modeling approaches' disadvantages. Since all the stakeholders will be involved in the system development, it also can reduce the misunderstanding at the stage of analyzing the security requirement engineering. Hence, the security concerns would be more meticulous analyzed by the system analyst before the system is built.

Chapter 12

Experiment - Threats to Validity

This chapter discusses four kinds of validities, which may occur in the experiment. Since the experiment was constructing as controlled experiment, it was therefore that there may existed issues of which cannot avoid during the execution and the design of the experiment.

12.1 Conclusion validity

The conclusion validity is concerned with the relationship between the technique used in this experiment and the outcome in kind of score of the experiment[17]. One important issue here was the sample size of the experiment. It was very important to make sure that sample size was enough to help to justify the conclusions drawn in the significant cases. The main effect claimed from the performance data was a significant advantage for Secure Tropos regarding the perceived ease of use (p= .000). The results show that 20 participants are needed or more that can give the chosen probabilities for type I and type II errors. hence the following relationship is satisfied:

$$N = \frac{4(\mu_{\frac{\alpha}{2}} + \mu_{\beta})}{ES^2}$$

Since type I and type II errors had been already chosen, therefore, there is a formula shall be use:

$$N = \frac{32}{ES^2}$$

(Claes Wohin 2000) [18]. Thus, in this case the effect size of 1.40 yields N = 20 would be needed to observe the efficiency in the experiment. Hence, the sample size (50 participants) was enough to justify the main conclusions in the experiment. However it was noted that the effect sizes must be used with caution, because the data distribution were not in general normally. There were also existing random irrelevancies in the experiment setting, for example, one student sitting in the front of the room went to toilet during the

experiment other than the break time. It might impact other participants by distracting their attention from work. The participants may be heterogeneous being in same organization, having similar background. This might reduce external validity of the experiment too. However, choosing more non-heterogeneous groups were difficult at university since the students had big study pressure and lots of exercises, and participate the experiment at a fixed time were an overwhelming impossible task. Hence, these 50 participants had a superiority that they can take in the experiment at the fixed time.

12.2 Internal validity

The internal validity is concerned with whether some validities threat the conclusion about a possible relationship between treatment and outcome can be shown[17].

Grouping the participants bias might occur. For example, it was unknown that the participants in one group may more motivate and suited for a new task than the other groups previously. Latin-Squares experimental design of the experiment was a solution for this threat, where all participants tried both techniques. The Latin-Squares experimental design was also solved the problems such as learning effects, boredom or fatigue as the participants tried the pre-set sequences for each randomly group. Furthermore, the booked auditorium for executing the experiment was big enough, thus there was adequate spaces to distribute seats to participants, so all the participants were in the same auditorium with equal working conditions without interaction between them during the entire experiment. The experiment contexts were also selected without any bias due to the participants' response on the pre-questionnaire, this may avoid the social threats to internal validity, such as resentful demoralization and compensatory rivalry. Because both modeling approaches with cases were new to the participants, so they may not give up and had motivate to do a good job, while learning something new.

Moreover, existing previous knowledge threats to internal validity was also threat to the experiment. It was hard to control the previous knowledge and learns of relative experience before the participants took into the experiment, which were potentially affecting both performance and preference for the two modeling approaches. Previous knowledge was hard to control although there was a pre-experiment questionnaire was designed to control the participants' background. However, all the participants were all from the same organization where they had almost the same classes at the same the department during their computer science studies at IDI due to their study level and the department's education plan for sophomore student. It was also investigated that there was no previous courses relative to both modeling approaches have been taught to the participants before the experiment, especially the security modeling approaches. But according to the knowledge of the author, the participants have been taught about software engineering in their third semester studies at IDI, so they may familiar with use cases from the software engineering course which may impact on the participants' bias in favor of Misuse case diagrams, but the experiment results were proved this threat was invalid.

The introduction to the both two modeling approaches may also have bias if one technique

was presented in a better way or more positivity than the other. It was felt that an issue of presenting the experiment techniques should in balance but even the author made every possible attempt to write the introduction to the both modeling approaches in a balance way, such as a similar structure and the same complexity to understand, there was still exist threats from the bias of individual participant who prefer one introduction than the other.

Moreover, the expected time for the experiment was enough for each section. But there were still exist threats to the experiment if the participants felt the introduction to the modeling approach was too short in each part of experiment. The experiment only gave participants limited time to read the relevant technique with a case. The introduction to each modeling approach was only 8 minutes; therefore such a short time was likely to lead to participants' favor an easily learnt technique other than a harder one when they try to use the relevant modeling technique, such as the notations of the two modeling approaches and the way to use these notations in the example section of the experiment. Furthermore, 35 minutes for the rest parts in each modeling technique in the experiment may also too short. Participants may like to identify and express easy-to-find threats and mitigations. After analyzing the participants' responses, it was showed that even some of the participants had experience on MUC and no experience on ST; they still were able to use Secure Tropos correctly after they read the introduction to the ST and its examples in the experiment sheets, all of them were able to identify a small number of threats in both cases. Frankly, according to their responses, within a limited time the participants were prefer to use Secure Tropos because ST modeling technique seems more sociality and they can identify threats and mitigations by non-technical method to express. Thus, only further experiments with more introduction time can assess the impact of limited time on the results.

The analyzing process might also be a threat, for example if the author was more willing to count the identified threats. In order to reduce this kind of threats, a rule was set that all the identified threats was treated as one or more threats depending on their description. For instance if their description including several conjunctions such as "or", "and", the author counted them as several threats due to the word the participants used. The same as the threats, if mitigation was suggested to several threats; this mitigation was treated as several mitigations relate to which threat was paired. However, the better way to coding the scoring the threats by multiple external experts, but it was too hard to find a person to involve in this work since the work was the author's master thesis.

The two modeling approaches backgrounds were not introduced to the participant as well. Since Misuse Case was the author's professor whose invention, so the author and author's two post-doctor supervisors had more experience and knowledge of MUC. However, as mentioned, the author's two post-doctor supervisors were responsible to contract the participants and they did not mention any detail information of experiment, such as the technique name that used in the experiment before the execution of the experiment. In this way it avoids threat that if the participants tried to search some information on the internet in order to know the experiment techniques in advance.

12.3 Construct validity

Construct validity is concerned with the extent to which the measurement of the data were relevant to the hypotheses[17]. The main concerns were the threats and mitigations of the effectiveness and coverage of the respective techniques, also with estimated the usage of diagrams, textual description and memory to identify threat or mitigations. These were measured in terms of the identified number of threats and mitigation, and the estimated percentages of the usage of diagrams, textual description and memory.

The experiment only had a brief introduction for each method, and a short example was introduced to the participants to apply each threat-analysis modeling approach. Hence, the experiment may not explore the full breadth of either technique, then there might be a risk that the participants did not use the techniques but rather brainstormed threats and mitigations in an ad hoc manner. However, according to the results, these limitations would apply to both modeling techniques and could not explain the significant differences in effectiveness.

And yet, It was also uncertain that all the participants measured threats and mitigations in the condition of they already understood the concepts and ways to use both modeling techniques. Although their outputs were scored by quantity, the quality of threats and mitigations were also worth to analyze. However, to control the quality of threats and mitigations were a hard task since each person has his/her own criteria of the answers. Thus, it was a difficult issue to control the output of the quality of threats and mitigation and taking them into account of the experiment analysis. Therefore, it would be interesting to perform new experiments, e.g., with a template that can be used the participants, and where a distinction between major and minor threats and their mitigations have already been made, the participants' may just select which one they think is match their thinking of the threat analysis. Such as boilerplates [1][60] to identify threats and mitigation is a good method for researcher to analyze the quality of threats and mitigation. In this way, coding is less of a problem.

There was a threat concerns with issues related to behavior of the participants and experiment. When the participants took part in the experiment, they might try to figure out what the purpose and intend result of the experiment is. So they may guess the hypotheses, either positively or negatively, depending on their attitude to the anticipated hypotheses. For instance, they may give guess numbers on the estimating of usage tasks in each modeling technique of experiment, also may give oppose number in the post-questionnaire to realize their purpose.

12.4 External validity

The external validity is concerned with the conditions that may limit the researcher's ability to generalize the results of the experiment to other situations, mostly to industrial systems development[17].

The participants of this experiment were the sophomore students with the background of Computer Science or Informatics. As the study plan for bachelor study level, they may little focus on security concerns refer to their courses when they have developed software. However, the practitioner who severed as developer who may have a higher competence in security concerns. Therefore, there is a possibly bigger difference between students and practitioners. Apparently, the practitioner may more sensitive for security modeling techniques. But, for an information system development, the stakeholders shall be not only the technique staff but also non-IT professionals, such as system end-users and project managers. Thus, it was also a reason why students were chosen as the participants since they were believed fulfill with both rules of technique and non-technique staffs as the real life situation required.

Both systems were described by one page of documentation only, for each system the participants just had 20 minutes to read text and diagrams, and then they were required to find threats and their mitigation within 10 minutes. Even though the complexities of both cases were reduced in order to make the experiment more understandable for participants. However, the system maybe still very complex and more documentation and more time needed when the participants try to identify the threats and their mitigations. Also if they may cooperate together to reach the same goal, it might be easier to the participants to find more unique threats and mitigations. However, in further the controlled experiment shall involve more complex issues in domains and allowing the participants do a team work on identify threats and suggest mitigations. Through this method the experiment cases can be extend the problem more complexity with larger scope.

The motivation of the participants was also an important issue of the experiment. Even though their organization was paid to support their future excursion, it was still hard control their motivation to take in the experiment, specially to ask them to take in experiment seriously, but if the experiment can be assigned to the participants' class and ask them to finish it as a quiz in class. It can probably solve this kind of problems. Yet, with an extra inducement offered to the participant who can identify most threats might be the best way to solve this problem too.

Chapter 13

Experiment - Conclusion and Future work

13.1 Conclusion

The thesis has reported on a pair of controlled experiment that compared two modeling approaches for elicitation of security threats. Secure Tropos and Misuse Cases were introduced to the participants at the beginning of each part of the experiment. All the participants were sophomore students at IDI, NTNU. The purpose of the experiment was to introduce two modeling techniques to the participants. They were asked to try to apply the modeling techniques into the pre-setting cases to see the participants' perception and performance of these two modeling techniques. A Latin-Square experimental design was used to control the tasks and techniques order. The variables were design for investigating effectiveness of the techniques measured as the number of threats and mitigation they found, coverage of the techniques were measured in terms of estimating of usage of diagrams, textual description and memory, and perceptions of the techniques measured which using a post-experiment questionnaire based on the Technology Acceptance Model(Davis 1989)[20].

The results indicate that Secure Tropos and Misuse Cases were complementary techniques having their strengths on pre-requirements analysis and system design analysis aspects. By comparing with the analytical comparison level of the two modeling techniques, the experiment results derived that the previous findings cover most results of the experiment. Therefore, the experiment results can coincide to the previous analytical comparison. The participants have demonstrated that they prefer to use ST. Due to the results analysis, ST was a goal based modeling approach; the participant can just through giving high level suggestions to the threats and mitigation during the analysis of security concerns of a system. Net Shopping case was indicated that most mitigations were identified when they have applied ST modeling approach. The explanation of this situation was that NT was

close to the participants' daily life.

13.2 Future work

The experiment results enlighten the paths for future work; the first work for the future empirical study is to use a "standard" template (e.g., boilerplate template) for the participants to easily list down their ideas of threats and mitigations. It would be easier for researchers to evaluate their suggestion. The second work would be a case study to track the entire system development, through this way to make sure that the combination of ST and MUC can illustrate the ideas that suggested from this experiment results, and to investigate whether the complementary combination can significant support the two modeling techniques to avoid their weakness in order to enhance their strengths together for effective SR elicitation and modeling. The other possible future works are listed as follows:

- Future work should involve the types of threats and mitigations, through this experiment it was found that the participants' responses on the threats and mitigations might be come from their experience. They could suggest several possible and useful methods such as biometric technology in the security analysis of threats and mitigation. Hence, to give the threats and mitigations a kind of classification could help researcher to analysis the participants' performance and other aspects. It also could help researchers to further improve the modeling techniques themselves.

- The example modeling approach case shall different with the test modeling case. For example, after the participants read the modeling approaches with the cases, the testing case shall be different with previous one. The reason why it is need to change the cases because there were already some identified parts of the systems' threats in the experiment, so the participants might only response the identified threats but no other new threats identified by them even though they suggested mitigations for those threats. To solve this problem, boilerplate template is a possible way for the participants to identify threats, the method can avoid the participant have less motivation to finish the experiment and can test whether they understand the modeling approach or not.

- Training time and group work should be considered for the future experiments. A short time experiment might not fully express the participants' behaviors during the experiment when they work individually. Hence, a training time for the participants and allow them work in a group would be a solution to measure comprehensively of their performance in the experiment.

- Comparing Misuse Case and Secure Tropos was only a beginning. Additional experiments involving additional techniques for threat identification should also be performed in the future. The combination of the two modeling techniques would be a nice path for comprehensive understanding of the identified threats and mitigations at the stage of the analysis in a system development, also can prefect practiced into a real system design stage. The investigation of this kind of work shall be in an

environment of industry with case study involving real threats analysis during the requirements phases.

Chapter 14

Project summary

This chapter contains the evaluation of various aspects of the work done in this project, as well as the course TDT4506 itself.

14.1 Work process

The project began at autumn semester 2011. It was divided into two main parts, the first part of the project was initiated by reviewing the basic concepts of i*-based modeling techniques and use case-based modeling techniques. Meanwhile, a couple of requirements comparison frameworks by other authors in this academic area were reviewed. Since there was no analytical comparison of i*-based and Use Case-based modeling initiatives as the task description mentioned. Thus, the project's first work was to illustrate the similarities and complementarily and differences of these techniques.

Secondly, in order to investigate the previous assumption that was presented in the late first phase of comparative analysis. Two modeling techniques (ST and MUC) were chosen to design an experiment (ST is i*-based and MUC is use case-based) in the master thesis. The objective of this experiment was to determine the differences in software engineer's modeling performance, experience and their preferences in use of different techniques (ST and MUC) for specific needs.

At end of the first part of the project, we had a published conference paper on conference RCIS'12 (see Appendix D). And the second conference paper is still under construction and will be submitted to a international conference.

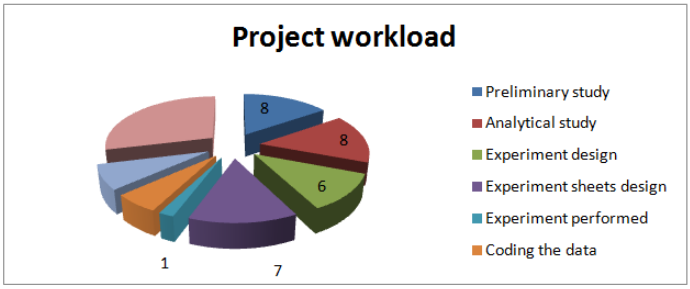

Figure 14.1: Project workload (week)

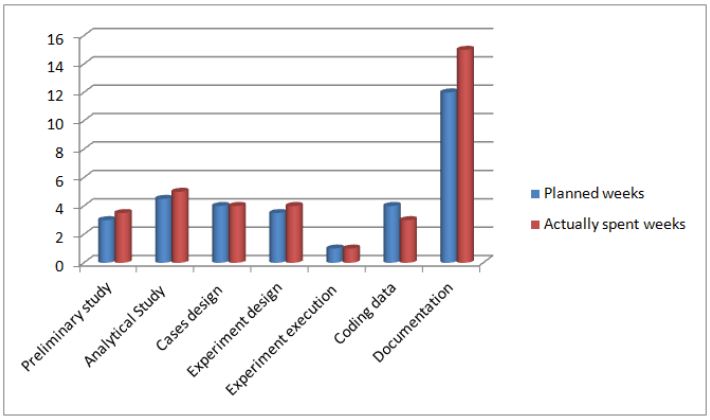

Figure 14.2: Planned weeks v.s. actually spend weeks for the project

14.2 Workload

The workload of the project was presented in figure 14.1 and 14.2. As the project was long term project from autumn 2011 to spring 2012, therefore a good time schedule was very important. The entire project costs ten months, the first four months were used for analytical comparison of i*-based and Use Case-based modeling initiatives. Since there was the first time in this academic area to compare Secure Tropos and Misuse Cases together, so the design of the experiment sheets cost 2 months until the final format of the experiment sheets were approval.

14.3 Supervisor relations

The supervisors and I had a lot of meetings during the project execution. The following table shows the meetings for each task of the project execution. Because we also had

meetings via Skype, so the table is just list the important dates of the project.

Table 14.1: Meeting dates with supervisor and co-supervisors

Meeting date	**Description**	**People**
April 15th, 2011	Kick off the project	Y. Pan and P. Karpati
June 7th, 2011	Identify a description of the search and the identified initiatives.	Y. Pan, P. Karpati and G. Sindre
July 10th, 2011	Preparation of analytical comparison for 8 modeling techniques.	Y. Pan and P. Karpati
August 29th, 2011	First draft of analytical comparison results of 8 modeling techniques.	Y. Pan, O. Daramola, P. Karpati
October 15th, 2011	Finally version of comparison results of the first part project.	Y. Pan, O. Daramola, P. Karpati
December 14th, 2011	Submitting the comparison results.	Y. Pan
January 10th, 2012	Confirm the context of the thesis and sign thesis contract.	Y. Pan and G. Sindre
Feburary 27th, 2012	Approved final version of experiment sheets.	Y. Pan, O. Daramola
March 13th, 2012	Experiment execution.	Y. Pan, O. Daramola, P. Karpati
March 24th, 2012	Discussing data extract rules.	Y. Pan and P. Karpati
April 15th, 2012	Confirming the results of the experiment.	Y. Pan and P. Karpati
May 5th, 2012	First thesis draft.	Y. Pan and P. Karpati
May 26th, 2012	Second thesis draft.	Y. Pan and P. Karpati
June 9th, 2012	Thesis submission.	Y. Pan

The supervisors often encouraged me during the project process when I misunderstood the concepts or sometimes the academic writing problems. They also provided valuable feedback on the work process, such as the analytical study and the experiment design; the communications between us were very nice, their academic spirits encouraging and enlightening for my future research work.

Bibliography

[1] Boilerplate. *http://en.wikipedia.org/wiki/Boilerplate_(text), 9:05am 5th May, 2012*.

[2] Characteristics of good requirements. http://en.wikipedia.org/wiki/Requirement.

[3] Kruskal-Wallis H tests. http://mathstat.carleton.ca/ smills/2009-10/STAT4504/k-wtest.ppt.

[4] Latin square experimental design. *http://en.wikipedia.org/wiki/Latin_square*.

[5] Likert scale. *http://en.wikipedia.org/wiki/Likert_scale, 9:35am 10th May, 2012*.

[6] pooled standard deviation. http://old.iupac.org/goldbook/P04758.pdf.

[7] Requirements engineering process. `http://www.cs.ucy.ac.cy/courses/EPL603/requirements_engineering_slides.ppt`.

[8] SecTro Beta Ver.2.0 User's Guide. `http://sectro.securetropos.org/resource/userGuide/useGuiSecTroVerTwo.pdf`.

[9] Systems development life-cycle. *http://en.wikipedia.org/wiki/Systems_development_life-cycle, 2:44pm 10th May, 2012*.

[10] Use Cases. http://en.wikipedia.org/wiki/Use_case.

[11] Wilcoxon signed-rank test. http://en.wikipedia.org/wiki/Wilcoxon_signed-rank_test.

[12] ReqSec-Requirements for Secure Information Systems. http://idi.ntnu.no/research/index.php?prosjekt=39, 2012.

[13] I. Alexander. Initial industrial experience of misuse cases in trade-off analysis. *Proceedings of the 10th Anniversary IEEE Joint International Requirements Engineering Conference (RE'02)*, 2002.

[14] Guttorm Sindre Andreas L. Opdahl. Experimental comparison of attack trees and misuse cases for security threat identification. *Information and Software Technology*, 51(5):916–932, 2009.

[15] A. Apvrille and M. Pourzandi. Secure software development by example. *IEEE security and privacy*, 3(4):10–17, 2005.

[16] I. K. Bray. An introduction to requirements engineering. 2002.

[17] Martin Höst Magnus C. Ohlsson Björn Regnell Anders Wesslén Claes Wohin, P. R. Experimentation in software engineering: An introduction. 2000.

[18] Martin Höst Magnus C. Ohlsson Björn Regnell Anders Wesslén Claes Wohin, P. R. Experimentation in software engineering: An introduction. 2000.

[19] Cohen P. West S.G. Cohen, J. and L.S. Aiken. Applied multiple regression/correlation analysis for the behavioral sciences. 2002.

[20] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 13(2):319–340, 1989.

[21] Yu E. Elahi, G. Modelling and Analysis of Security Trade-offs - A Goal Oriented Approach. *Data and Knowledge Engineering*, 68(7):579–598, 2009.

[22] Yu E Elahi, G. Trust trade-off analysis for security requirements engineering. 2009.

[23] C.B. Haley et al. Security requirements engineering: A framework for representation and analysis. *IEEE Trans. software engingeering*, http://doi.ieeecomputersociety.org/10.1109/TSE.2007.70754, 2007.

[24] G. Boström et al. Exteding XP pratices to support security requirements engineering. *Proc. int'l workshop software engineering for secure system (SESS)*, pages 11–18, 2006.

[25] J. Mylopoulus E.Yu. Why goal-oriented requirements engineering. *E.Dubois, A.L.Opdahl, K. Pohl(Eds.)*, Proceedings of fourth international workshop on requirements engineering: foundations of software quality(15-22), 1998.

[26] Mylopoulos E.Yu., L. L. A Social Ontology for Integrating Security and Software Engineering. *In Integrating Security and Software Engineering: Advance and Future Visions*, pages 743–772, 2006.

[27] Gürse S. Heisel M. Santen T. Schmidt H. Fabian, B. A comparison of security requirements engineering methods. *Requirements Engineering*, 15(1):7–40, 2010.

[28] Erdfelder E. Buchner A. Lang A. G. Faul, F. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(1149-1160), 2009.

[29] Erdfelder E. Lang A. G. Buchner A. Faul, F. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(175-191), 2007.

[30] Eduardo Fernandez. A methodology for secure software design. *International Symp, Web services and applications (ISWS)*, 2004.

[31] D Firemith. Security Use Case. *In Journal of Object Technology*, 2(3):53–64.

[32] D.G. Firesmith. Engineering security engineering. *J. Object Technology*, 2(1):53–68, 2003.

[33] Martin Fowler. UML Distilled: A Brief Guide to the Standard Object Modeling Language (3rd ed. ed.). *Addison-Wesley*.

[34] Cares C. Franch X. Navarret F Grau, G. A Comparative Analysis of i* Agent-Oriented Modelling Techniques. *The 18th international conference on software engineering and knowledge engineering (SEKE' 06)*, 2006.

[35] Andreas L. Opdahl Guttorm Sindre. Capturing Security Requirements through Misuse Cases. *In Proc. 14th Norwegian Informatics Conference*, 2001.

[36] W. G. Hopkins. A new view of statistics. 2001.

[37] Jonsson P. Övergaard G Jacobson Ivar, Christerson M. Object-Oriented Software Engineering - A Use Case Driven Approach. 1992.

[38] Dianxiang Xu Josh Pauli. Integrating Functional and Security Requirements with Use Case Decomposition. *Proceedings of the 11th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS'06)*, 2006.

[39] Sindre G Karpati, P. Comparing Misuse Case and Mal-activity Diagrams for Modelling Social Engineering Attacks. *2nd European Security Conference*, 2011.

[40] Sindre G. Karpati, P. and A Opdahl. Characterising and Analysis Security Requirements Modelling Initiatives. *Proceedings of the Sixth International Conference on Availability, Reliability and Security*, (710-715), 2011.

[41] Sindre G. Opdahl A. Karpati, P. Visualizing Cyber Attacks with Misuse Case Maps. *LNCS, REFSQ*, 6182:262–275, 2010.

[42] P. Opdahl A. Raspntnig C. Sindre G Katta V., K. Comparing two techniques for intrusion visualization. *Lecture Notes in Business Inforamtion Processing*, 68:1–15, 2010.

[43] Zulkernine M Khan, M. A survey on requirements and design methods for secure software development. *Technical report*, 562, 2009.

[44] S. Lipner and M. Howard. The trustworthy computing security development lifecycle. *Microsoft corp.*, http://msdn2.mirosoft.com/en-us/library/ms995349.aspx, 2005.

[45] E. Mylopoulos J. Liu, L. Yu. Security and Privacy Requirements Analysis within a Social Setting. *In Proceedings of the RE'03*, pages 151–161, 2003.

[46] Zannone N Massicci F., M. J. Ontologies for Bussiness Interaction. *Information Science Reference*, pages 188–207, 2007.

[47] Fox C McDermott, J. Using Abuse Case Models for Security Requirements Analysis. *Computer Security Application Conference (ACSAC'99)*, 1999.

[48] J. D. Meier. Web application security engineering. *IEEE security and privacy*, 4(4):16–24, 2006.

[49] Blanco C. Sanchez L. Fernández-Medina E Mellado, D. A systematic review of security requirements engineering. *Computer standards interfaces*, 32 (4)(153-165), 2010.

[50] Mouratidis. Secure Tropos notations. *http://www.securetropos.org/*.

[51] Giorgini P. Mouratidis, H. A Security-oriented Extension of Tropos Methodology. *International Jouranl of Software Engineering and Knowledge Engineering*, 12(2):285–309, 2004.

[52] Nuseibeh B. Yu J Nhlabatsi, A. Security requirements engineering for evolving software system: A survey. *International journal of secure software engineer (IJSSE)*, 1 (1)(54-73), 2011.

[53] Y Pan. P Karpati. G-Sindre. O, Daramola. A comparative review of i*-based and use case-based security modelling initiatives. *In Proceeding of International Conference on Research in Information Science (RCIS'12)*, pages 146–157, 2012.

[54] Yushan Pan. Comparison of i*-based and Use Case-based Modelling initiatives for Security Requirement Engineering. *Porject report*, http://idi.ntnu.no/research/index.php?prosjekt=39, 2011.

[55] Guttorm Sindre Peter Karpati, Andreas L. Opdahl. Experimental evalution of Misuse Case Maps for Eliciting Security Requirements. *1st Security Conference Europe*, 2010.

[56] G. Peterson. Collaboration in a secure development process part 1. *Information security bull*, pages 165–172, 2004.

[57] P.Torr. Demystifying the threat modelling process. *IEEE security and privacy*, 3(5):66–70, 2005.

[58] Lee L. Siau, K. Are use case and class diagrams complementary in requirements analysis? An experimental study on use case and class diagrams in UML. *Requirements engineering*, 9(4):229–237, 2004.

[59] Opdahl A.L Sindre, G. Eliciting Security Requirements with Misuse Cases. *In Requirements Engineering*, 10(1):34–44, 2005.

[60] I. O. a. T. Stålhane. Guided Natural Language and Requirement Boilerplates. *TDT4242 Requirements and testing*, 2012.

[61] Sindre G. Stålhane, T. Safety Hazard Identification by Misuse Cases: Experimental Comparison of Text and Diagrams. *MoDELS 2008*, 2008.

[62] Meland P. Tøndel I., Jaatun M. Security requirements for the rest of us: a survey. *IEEE Trans. software engingeering*, 25(1):20–27, 2008.

[63] K. R. van Wyk and G. McGraw. Bridging the gap between software development and information security. *IEEE security and privacy*, 3(5):75–79, 2005.

[64] Andreas L. Opdahl-Christian Taspotnig Guttorm Sindre Vikash Katta, Peter Karpati. Comparing two techniques for intrusion visualization. *Lecture Notes in Business Information Processing*, 68:1–15, 2010.

[65] Frank Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[66] E Yu. Agent-Oriented Modelling: Software Versus the World(2011). *Agent-oriented software engineering AOSE-2001 Workshop Proceedings, LNCS222*, pages 206–225.

[67] Eric Yu. Modelling strategic relationships for process reengineering. *Ph.D thesis, also Tech. Report DKBS-TR-94-6*, 1995.

[68] Eric S.K. Yu. Towards Modelling and Reasoning Support for Early-Phase Requirement Engineering. *Requirements engineering*, pages 226–235, 1997.

[69] Stian Veum Møllersen. Øystein Jaren Samuelsen. Dag Øyvind Tornes. Yushan Pan., Finn Robin Kåveland Hansen. ESUMSDroid: Handheld Client for Heart Monitoring. *Norwegian University of Science and Technology*, 2010.

Appendix A

Experiment Sheets

Introduction to the experiment

Thank you for your participation! In this experiment, you will learn about two different security modelling techniques, read two different scenarios illustrated by either of the modelling techniques and answer questions.

Estimated schedule in minutes:

1. Pre-experiment questionnaire: *5 minutes*

2. Reading the introduction to the experiment: *5 minutes*

a) Introduction to Secure Tropos (ST)/Misuse Use Cases (MUC) modelling technique: *8 minutes.*

b) Reading Net Shopping/ (Health Insurance System) case description and diagrams together: *20 minutes.*

c) Find threats and mitigations for Secure Tropos/Misuse Case: *10 minutes.*

d) Estimating the usage percentage of textual description and diagram:*5 minutes*

e) Post-experiment questionnaire: 5 minutes

f) Break *5 minutes*

g) Introduction to Misuse Case (MUC)/Secure Tropos (ST) modelling technique: *8 minutes*.

h) Reading Health Insurance System/Net Shopping case description and diagrams together: *20 minutes.*

i) Find threats and mitigation: *10 minutes*

j) Estimating the usage percentage of textual description and diagram:*5 minutes*

k) Post-experiment questionnaire: *5 minutes*

Pre-experiment questionnaire (5 minutes)

Please give an anonymous identifier which you will remember: _____

Please circle the number which you consider most fit.

	Never heard about it	Read about it	Tried it out	Used it a lot	Expert
System modelling	1	2	3	4	5
UML use case diagram	1	2	3	4	5
Misuse case diagram	1	2	3	4	5
UML activity diagrams	1	2	3	4	5
Mal-activity diagrams	1	2	3	4	5
Secure Tropos diagrams	1	2	3	4	5
i* modelling diagrams	1	2	3	4	5

Please indicate the following data:

1. Number of completed semesters of study after high school: _____

2. Number of months of study-related working experience, including summer jobs:__

(Re-calculated of full time, e.g., a 10 month 20% part job should be report as 2 month)

Introduction to Secure Tropos Modelling (8 minutes)

1. What is Secure Tropos?

Secure Tropos is a security-oriented extension of the Tropos methodology. Tropos is an agent-oriented software engineering methodology. Secure Tropos is used for analysing the security needs of the stakeholders and the system. It shows the security constraints imposed on the stakeholders and the system, identifies secure entities that guarantee the satisfaction of the security constraints, and assigns capabilities to the system to help towards the satisfaction of the secure entities.

The key concepts of the Tropos, includes the following:

1) **Security constraint**: This is defined as a restriction related to security issues, such as privacy, integrity and availability.

2) **Secure goal**: This represents the strategic interests of an actor with respect to security.

3) **Secure plan**: This is defined as a plan that represents a particular way for satisfying a secure goal.

4) **Secure dependency**: This introduces security constraint(s) that must be fulfilled for the dependency to be satisfied.

2. Notations used in Secure Tropos Diagram.

Figure 1 shows the Tropos and Secure Tropos notation and some Tropos concepts graphically.

Actor: entities that have strategic goals and intentionality.

Goal: an actor's strategic interests.

Soft-goals: goals without clear criteria whether they are satisfied or not.

Task: represents at an abstract level, a way of doing something.

Resource: represents a physical or information entity.

Dependencies: indicate that one actor depends on another in order to attain some goals, execute some tasks, or deliver a resource.

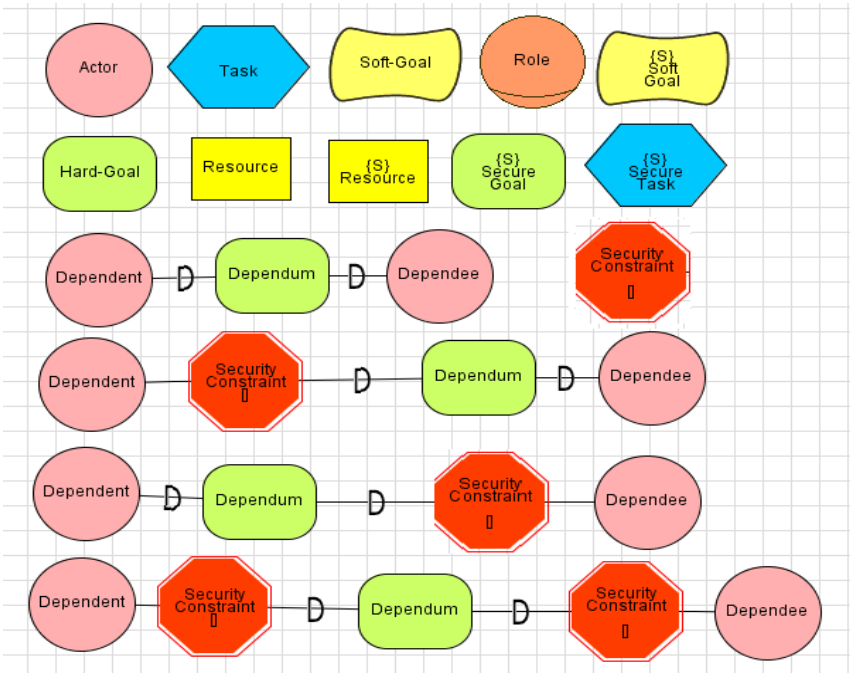

Figure 1: Tropos, Secure Tropos notation and Tropos concepts graphically

3. Example of Secure Tropos Diagram

Figure 2 shows the full dependency link with a security constraint. Customer requires DVD from the DVD store. Both customer and DVD store are actors in this case. Watch movie is a secure goal. Customer is a depender with constraints to the dependee DVD store. DVD store has the responsibility to

keep the customer's personal information private and just could share the information only when consent is available.

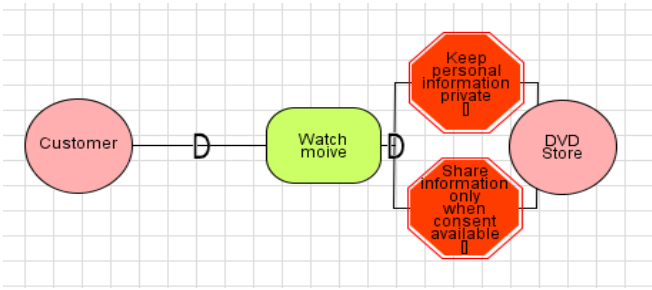

Figure 2 Full dependency links with a security constraint

Net Shopping (20 minutes)

The most known traditional way is that we browse goods, compare the goods provided by different salespersons, and decide to buy it, pay for it and wait for its delivery. Another way to buy stuff via the Internet is "trust rating". In trust rating, we pay the fees to pay-pal first, when we receive the goods, we check and use it within 3 days to ensure we are comfortable with the goods then we can confirm the deal with pay-pal to finish this transaction, pay-pal pays money to the seller after we confirm the deal.

"Trust rating" is a transaction for the salesperson. Each successful transaction allows buyers to mark a credit evaluation for it. Evaluation is divided into positive, neutral, and negative feedback categories. Each credit is equal to a credit score. After each transaction, if there is a good evaluation for the salesperson, they will get one point. No point for medium evaluation, one point is deducted from a salesperson for a negative evaluation. If a salesperson has no medium and negative evaluations within 300 continuous transactions, he will get a diamond to illustrate the high trust rating. The more diamonds he has, the higher the good quality and trust during the transaction. In this case there are three important vulnerabilities:

1. How to avoid salespersons selling goods to themselves in order to earn more diamonds.

2. The salespersons could sell the virtual products (e.g. illegal scanned eBook) in a very low price to their partners (themselves or other salespersons), after they get the diamonds, they may sell other real products (e.g. print copy of a real book).

3. Some third party platforms may help salespersons to get diamonds in a dubious way, such that all the sold goods are later refunded to the seller but they will get a fee after the transaction (when the salesperson has already earned the diamond).

Secure Tropos diagram for Net Shopping

Figure 3 illustrates Secure Tropos diagrams for Net Shopping.

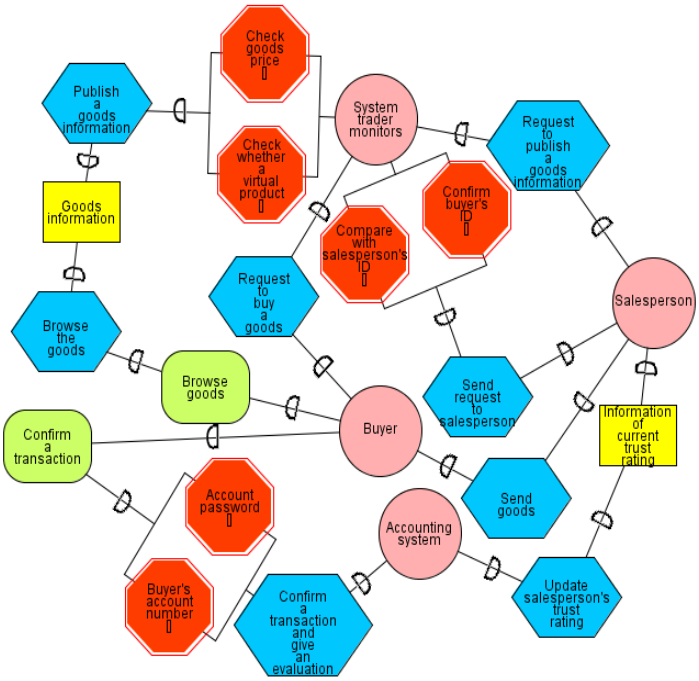

Figure 3 Secure Tropos diagrams for Net Shopping

Health Insurance System (20 minutes)

An application that runs on the Microsoft and Android Mobile platform could be used to collect an older person's health data everyday which includes heart beat rates, carbon dioxide in blood, body temperature, and blood pressure.

All of the data from an older person could be sent at once or one at a time to the Health Insurance Department (HID) server in the hospital. The most important and delicate matters for the health insurance department is the privacy of all older person's medical information and the sharing of it. When HID receives the data, a doctor analyses the data and gives a feedback, the feedback is sent to the older person by a system, and also a copy is sent to HID.

When HID receives the feedback form from the doctor, they will calculate fees and update the health insurance database in the hospital, then check the database of the status of older person's health insurance, to figure out whether the medical health fee is now above 1800NOK or not. According to the welfare and security system in Norway, if the medical health fee is above 1800NOK the older person will not need to pay the bills in a normal year rather, HID will request bank to send the bills to the insurance company, otherwise, HID will request bank to send bills to the older person that should be paid in a normal year.

When bank receives the message from the HID, they will create bills according to the messages received from HID, and send bills to the older person or insurance company. We are only concerned with the way the older person and insurance company pay the bills via Internet. Bank should guarantee that there is a secure interaction between the users and bank system.

Secure Tropos diagram for Health Insurance System

Figure 4 illustrates the Secure Tropos diagrams for Health Insurance System.

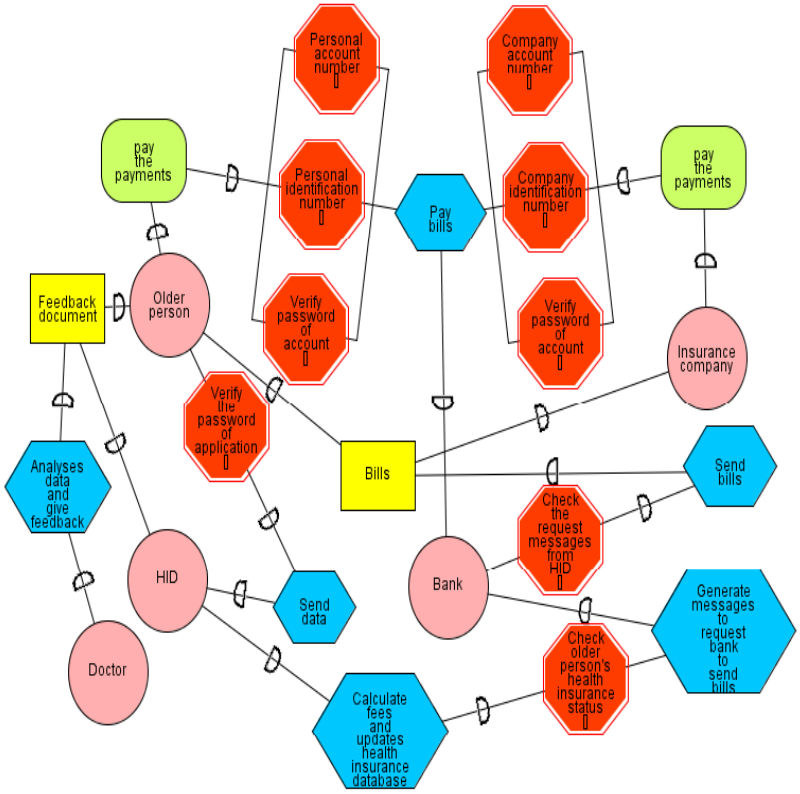

Figure 4 Overview of Health Insurance System

Threats and Mitigation (10 minutes)

Please list here as many potential threat(s) as you can **after reading.** Try to come up with mitigation for the threat(s) you find (in English), several possible mitigation can be suggested for one threat.

Threat	***Mitigation***
1.	1.1 1.2 ...

Estimating the usage percentage of textual description and diagram (5 minutes)

1) Please give an estimate, to what extent did you use:
   a) The Secure Tropos diagrams to identify threats? _____%
   b) The textual description of the scenario to identify threats? ____%
   c) Only memory to identify threats? ____%

The sum of the percentages should be 100%. Sum: 100%

2) Please give an estimate, to what extent did you use
   a) The Secure Tropos diagrams to find ideas for mitigations? ___%
   b) The textual description of the scenario to identify mitigations? ___%
   c) Only threats identified earlier to find ideas for mitigations? ____%
   d) Only memory to find ideas for mitigations? _____%

The sum of the percentages should be 100%. Sum: 100%

Post-experiment questionnaire for Secure Tropos (5 minutes)

Please fill in the number from 1-5 in front of each raw in the table which you feel most fit. "1" stands for strongly agree while "5" stands for strongly disagree.

	Strong Agree	Agree	Neutral	Disagree	Strongly Disagree
The textual description combined with the Secure Tropos diagram gave me a better understanding of the Secure Tropos than I would have gotten from the text alone.	1	2	3	4	5
If working as a freelance consultant for a customer who needs help, I would use Secure Tropos diagram to identify security concerns.	1	2	3	4	5
I found it easy to interpret the Secure Tropos diagram.	1	2	3	4	5
If I need to analyse a complex scenario, I would consider using Secure Tropos diagram.	1	2	3	4	5
Secure Tropos diagram would be useless in finding system security issues.	1	2	3	4	5
If I need to explain some security scenarios to my colleagues, I would consider using Secure Tropos diagram.	1	2	3	4	5

It would be easy to draw a Secure Tropos diagram based on a similar case description like in the experiment.	1	2	3	4	5
Secure Tropos diagram caused me to waste time on threats of minor importance	1	2	3	4	5
The connection between the Secure Tropos diagram and the textual description of the securities were confusing.	1	2	3	4	5
Secure Tropos diagram made me more productive in recognizing threats and finding security issues for the case.	1	2	3	4	5
Secure Tropos diagram could be a great help to discuss preventions for security scenarios with colleagues.	1	2	3	4	5
It would be easy to get used to the notation of Secure Tropos diagram.	1	2	3	4	5

Introduction to Misuse Case Modelling (8 minutes)

1. What is Misuse Case?

Misuse Case is an extension of Use Case where it is possible to model not only the normal functionality wanted in the system, but also negative functionality that is not wanted, with the purpose of addressing security concerns. In addition to normal use cases and normal actors, there would be malicious actors performing Misuse Case that cause harm to the system.

2. How to define the Misuse Case and Misuser?

We define Misuse Case and Misuser as follows:

Misuse Case: A sequence of actions, including variants, that a system or other entity can perform, interacting with Misusers of the entity and causing harm to some stakeholder if the sequence is allowed to complete.

Misuser: An actor that initiates Misuse Case, either intentionally or inadvertently.

3. What are the notations used in Misuse Case Diagram?

Compared to regular Use Cases, the inverted notation indicates both: similarity (because the same symbol shapes are used) and negation (because of the inverted graphics). Use Case and Misuse Case can, therefore, be shown in the same diagram without confusion.

Ordinary Use Case relationships such as include, extend, and generalize can be used between Misuse Case too, and ordinary association relationships can be

used between Misuser and their Misuse Case. Misuse Case also can threaten a Use Case when the Use Case is threatened by the "obtain info on marketing plans" Misuse Case.

Notation of Misuse Case diagram:

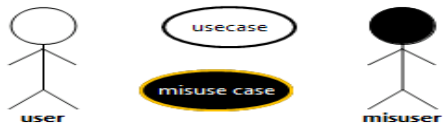

Figure 5 The notation of MUC

4. Example of Misuse Case diagram

Figure 6 uses inverted graphics to show Misuse Case together with regular Use Case in a high-level specification of part of Health Insurance System.

Figure 6 Threats of Health Insurance System

Net Shopping (20 minutes)

The most known traditional way is that we browse goods, compare the goods provided by different salespersons, and decide to buy it, pay for it and wait for its delivery. Another way to buy stuff via the Internet is "trust rating". In trust rating, we pay the fees to pay-pal first, when we receive the goods, we check and use it within 3 days to ensure we are comfortable with the goods then we can confirm the deal with pay-pal to finish this transaction, pay-pal pays money to the seller after we confirm the deal.

"Trust rating" is a transaction for the salesperson. Each successful transaction allows buyers to mark a credit evaluation for it. Evaluation is divided into positive, neutral, and negative feedback categories. Each credit is equal to a credit score. After each transaction, if there is a good evaluation for the salesperson, they will get one point. No point for medium evaluation, one point is deducted from a salesperson for a negative evaluation. If a salesperson has no medium and negative evaluations within 300 continuous transactions, he will get a diamond to illustrate the high trust rating. The more diamonds he has, the higher the good quality and trust during the transaction. In this case there are three important vulnerabilities:

1. How to avoid salespersons selling goods to themselves in order to earn more diamonds.

2. The salespersons could sell the virtual products (e.g. illegal scanned eBook) in a very low price to their partners (themselves or other salespersons), after

they get the diamonds, they may sell other real products (e.g. print copy of a real book).

3. Some third party platforms may help salespersons to get diamonds in a dubious way, such that all the sold goods are later refunded to the seller but they will get a fee after the transaction (when the salesperson has already earned the diamond).

Misuse Case diagrams for Net shopping case

Figure 7: shows salesperson pretends as a buyer and sales goods to himself or herself.

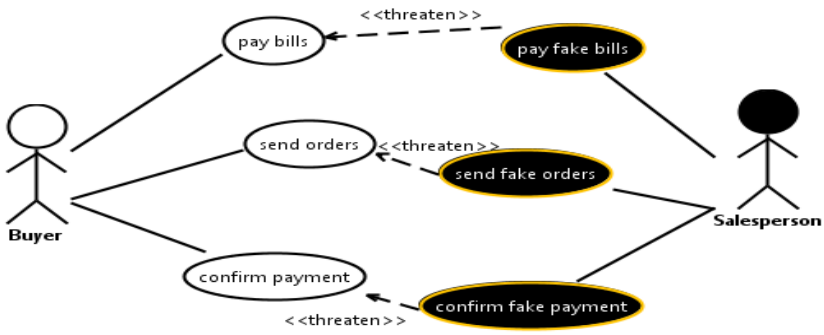

Figure 7 Salesperson pretends as a buyer

Figure 8 shows some third parties pretend as a buyer to buy goods to help salesperson to get more diamonds.

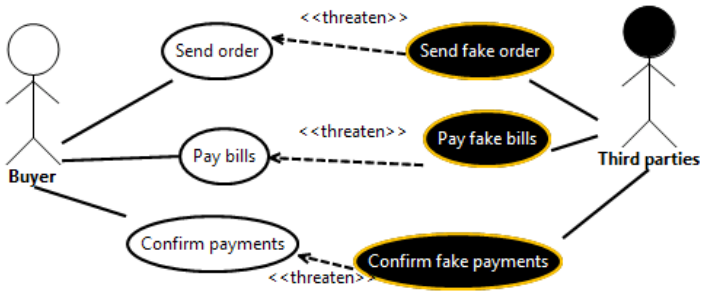

Figure 8 Third parties pretends as a buyer

Figure 9 shows how attacker can pretend as a system trade monitor to allow the virtual products transaction.

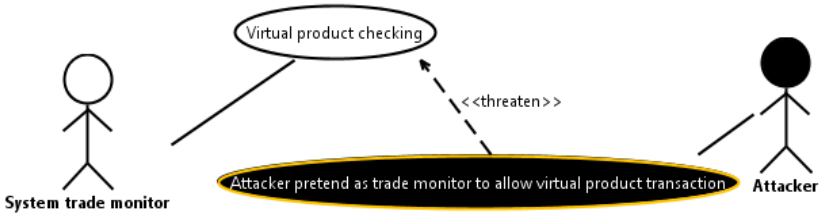

Figure 9 Attacker allows the virtual products trade.

Health Insurance System (20 minutes)

An application that runs on the Microsoft and Android Mobile platform could be used to collect an older person's health data everyday which includes heart beat rates, carbon dioxide in blood, body temperature, and blood pressure.

All of the data from an older person could be sent at once or one at a time to the Health Insurance Department (HID) server in the hospital. The most important and delicate matters for the health insurance department is the privacy of all older person's medical information and the sharing of it. When HID receives the data, a doctor analyses the data and gives a feedback, the feedback is sent to the older person by a system, and also a copy is sent to HID.

When HID receives the feedback form from the doctor, they will calculate fees and update the health insurance database in the hospital, then check the database of the status of older person's health insurance, to figure out whether the medical health fee is now above 1800NOK or not. According to the welfare and security system in Norway, if the medical health fee is above 1800NOK the older person will not need to pay the bills in a normal year rather, HID will request bank to send the bills to the insurance company, otherwise, HID will request bank to send bills to the older person that should be paid in a normal year.

When bank receives the message from the HID, they will create bills according to the messages received from HID, and send bills to the older person or insurance company. We are only concerned with the way the older person and

insurance company pay the bills via Internet. Bank should guarantee that there is a secure interaction between the users and bank system.

Misuse Case diagrams for Health Insurance System

Figure 10 and 11 are the examples of Health Insurance System's vulnerabilities.

Figure 10: Attacker can get the password and send the data package to DHI server without notice by older persons.

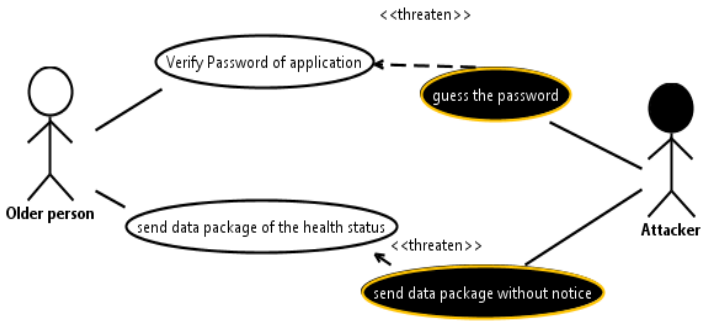

Figure 10 Attacker illegally get the password and send data to DHI

Figure 11 shows attacker modifies the database after he/she attacked the system.

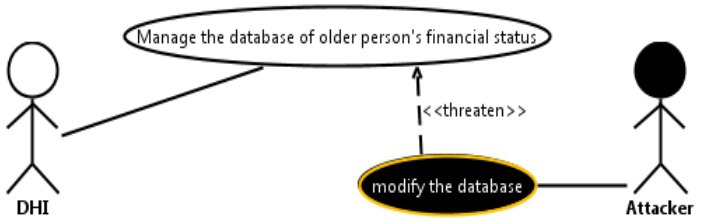

Figure 11 Attacker modifies the database after he/she attacked the system

Figure 12 shows attacker gets the password of the insurance company.

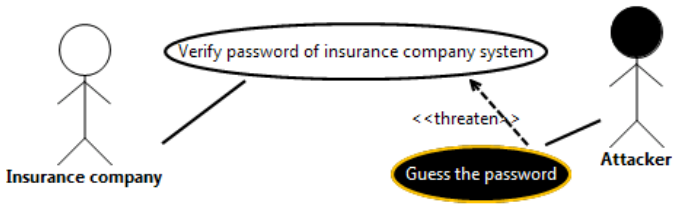

Figure 12 Attacker gets the password of the insurance company system

Figure 13 shows the attacker get the password of the bank system and login illegally. He can create fake bills and collect the payments through his own account.

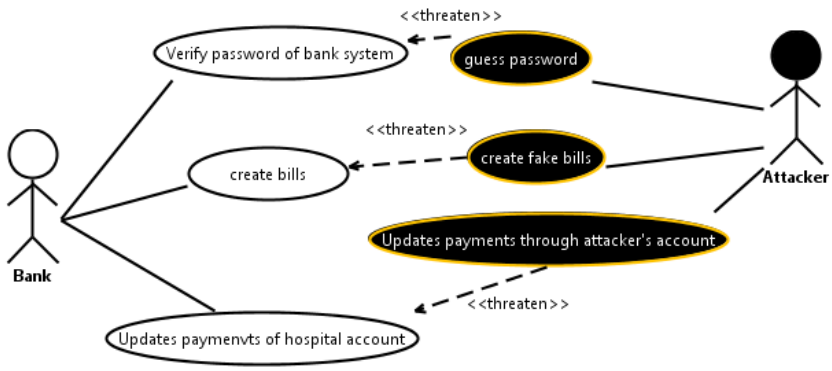

Figure 13 Attacker get the password of bank system and create fake bills, collect payments by his own account.

Threats and Mitigation (10 minutes)

Please list here as many potential threat(s) as you can **after reading.** Try to come up with mitigation for the threat(s) you find (in English), several possible mitigation can be suggested for one threat.

Threat	***Mitigation***
1.	1.1 1.2 ...

Estimating the usage percentage of textual description and diagram (5 minutes)

1) Please give an estimate, to what extent did you use:
   a) The Misuse Case diagrams to identify threats? _____%
   b) The textual description of the scenario to identify threats? ____%
   c) Only memory to identify threats? ____%

The sum of the percentages should be 100%. Sum: 100%

2) Please give an estimate, to what extent did you use

   a) The Misuse Case diagrams to find ideas for mitigations? ___%
   b) The textual description of the scenario to identify mitigations? ___%
   c) Only threats identified earlier to find ideas for mitigations? ____%
   d) Only memory to find ideas for mitigations? _____%

The sum of the percentages should be 100%. Sum: 100%

Post-experiment questionnaire for Misuse Case (5 minutes)

Please fill in the number from 1-5 in front of each raw in the table which you feel most fit. "1" stands for strongly agree while "5" stands for strongly disagree.

	Strong Agree	Agree	Neutral	Disagree	Strongly Disagree
The textual description combined with the Misuse Case diagram gave me a better understanding of the Misuse Case than I would have gotten from the text alone.	1	2	3	4	5
If working as a freelance consultant for a customer who needs help, I would use Misuse Case diagram to identify security concerns.	1	2	3	4	5
I found it easy to interpret the Misuse Case diagram.	1	2	3	4	5
If I need to analyse a complex scenario, I would consider using Misuse Case diagram.	1	2	3	4	5
Misuse Case diagram would be useless in finding system security issues.	1	2	3	4	5
If I need to explain some security scenarios to my colleagues, I would consider using Misuse Case diagram.	1	2	3	4	5
It would be easy to draw a Misuse Case diagram based on a similar case description like in the experiment.	1	2	3	4	5
Misuse Case diagram caused me to waste time on threats of minor	1	2	3	4	5

importance					
The connection between the Misuse Case diagram and the textual description of the securities were confusing.	1	2	3	4	5
Misuse Case diagram made me more productive in recognizing threats and finding security issues for the case.	1	2	3	4	5
Misuse Case diagram could be a great help to discuss preventions for security scenarios with colleagues.	1	2	3	4	5
It would be easy to get used to the notation of Misuse Case diagram.	1	2	3	4	5

Appendix B

Experiment Data

Experiment Data

1. Pre-questionnaire

Pre-questionniare		Never heard about it = 1, Read about it = 2, Tried it out = 3, Used it a lot = 4, Expert = 5								
Seq. ID	Student ID	ystem modellin	L use case diag	use case diagr	. activity diag	-activity diagre	Tropos diag	odelling diagr	Semesters	Jobs
1	winniuile	2	3	2	2	1		1	3	1.5
2	88	2	3	1	3	1	1	1	4	0
3	MJ14	2	3	1	1	1	1	1	7	0
4	VON	3	3	2	2	1	1	1	1	0
5	G2K	2	4	1	2	1	1	1	3	0
6	H33	1	3	1	3	1	1	1	3	0
7	202	3	3	1	3	1	1	1	3	0
8	JB24	2	3	2	3	1	1	1	3	0
9	6	3	3	2	4	2	1	1	3	0
10	1114k	2	3	1	2	1	1	1	5	0
11	tru	3	3	1	2	1	1	1	3	2
12	~~!!$	4	3	1	3	1	1	1	3	2
13	orange	2	3	1	1	1	1	1	5	1
1	248357	3	4	1	1	1	1	1	3	2
2	```````	3	3	1	3	1	1	2	3	1
3	i1648	3	3	1	3	1	1	1	7	0
4		3	3	2	2	1	1	2	3	0
5	SH	3	3	1	1	1	1	1	3	0
6	123	3	3	1	2	1	1	1	3	0
7	MHH	2	3	1	1	1	1	1	3	0
8	he	3	4	1	2	1	1	1	3	0
9	APE123	3	3	1	2	1	1	1	3	0
10	Trallala	3	3	1	1	1	1	1	3	0
11	test	3	4	1	2	1	1	2	3	3
1	45	2	3	1	1	1	1	1	3	0
2	AK	3	4	1	3	1	1	1	3	0
3	hdr	3	1	1	1	1	1	1	3	0
4	bo	3	3	1	1	1	1	1	3	0
5	294	2	3	2	3	2	2	2	3	0
6	4747	2	3	1	2	1	1	1	3	0
7	pfhoy	3	3	1	1	1	1	1	3	0
8		3	3	1	1	1	1	1	3	0
9	725	1	2	1	1	1	1	1	2	0
10	1024	3	3	2	2	1	1	3	3	2
11	1247	3	4	1	2	1	1	1	3	
12	#10	2	3	1	2	1	1	1	3	1
13	C9A1	3	3	1	3	1	1	1	2	1
1	333	3	3	2	3	1	1	1	9	2
2	1105	3	3	1	2	1	1	3	3	2
3	47900	3	3	1	1	1	1	1	3	0
4	A.N	3	3	2	3	1	1	1	3	0
5	vena	3	4	1	2	2	1	1	3	0
6	67	3	3	1	3	1	1	1	3	0
7	3647	1	3	1	2	1	1	1	3	0
8	HMO	2	3	1	1	1	1	1	3	0
9	lowe	3	3	1	2	1	1	3	3	0
10	T	2	3	1	1	1	1	1	3	0
11	1106	3	3	2	2	1	1	1	3	1
12	8526	3	3	1	2	1	1	1	3	5
13	CYI	3	3	2	3	1	1	1	3	7

2. Post-questionnaire –Secure Tropos

Secure Tropos	ST_PU 1	ST_PU2	ST_PU3	ST_PU4	T_Average PU	ST_ITU 1	ST_ITU2	ST_ITU3	ST_ITU4	_AVERAGE I	ST_PEOU1	ST_PEOU2	ST_PEOU3	ST_PEOU4	AVERAGE PE
Group 1 <1>	2	3	3	4	3	3	3	2	3	2.75	3	3	3	3	3
2	2	4	2	3	2.75	4	3	4	4	3.75	4	3	3	4	3.5
3	3	3	4	2	3	5	2	5	5	4.25	5	5	5	4	4.75
4	1	5	2	3	2.75	2	2	1	2	1.75	3	2	2	2	2.25
5	4	3	3	4	3.5	4	4	4	2	3.5	4	2	4	4	3.5
6	2	4	3	2	2.75	4	3	2	2	2.75	4	4	4	3	3.75
7	2	4	2	2	2.5	3	3	3	2	2.75	4	2	3	2	2.75
8	3	4	2	2	2.75	3	3	2	2	2.5	4	4	3	2	3.25
9	1	5	3	1	2.5	2	2	3	2	2.25	2	3	3	1	2.25
10	1	3	3	2	2.25	2	3	4	2	2.75	2	4	3	2	2.75
11	2	4	2	2	2.5	2	2	2	2	2	3	3	2	2	2.5
12	5	3	2	2	3	4	4	4	4	4	5	4	4	4	4.25
13	2	4	1	2	2.25	4	2	2	4	3	3	3	3	2	2.75
Group <2>1	1	3	3	5	3	2	2	3	1	2	2	2	3	3	2.5
2	1	4	3	4	3	1	4	2	2	2.25	2	2	2	2	2
3	1	4	3	4	3	3	4	3	2	3	1	2	3	3	2.25
4	2	3	4	3	3	3	2	2	2	2.25	2	2	4	3	2.75
5	2	4	4	4	3.5	3	3	3	4	3.25	2	3	2	3	2.5
6	2	3	3	4	3	3	2	2	2	2.25	2	3	3	3	2.75
7	1	4	4	5	3.5	3	3	4	2	3	2	3	2	3	2.5
8	1	4	3	5	3.25	2	2	2	2	2	2	4	3	2	2.75
9	2	4	4	4	3.5	3	3	2	2	2.5	2	2	3	2	2.25
10	2	3	3	4	3	3	3	3	3	3	4	3	4	3	3.5
11	1	3	4	4	3	2	3	3	3	2.75	3	2	4	3	3
Group <3> 1	3	4	3	2	3	3	4	3	4	3.5	3	4	4	3	3.5
2	2	3	3	2	2.5	4	4	4	3	3.75	4	4	3	3	3.5
3	3	4	3	2	3	4	3	4	3	3.5	4	4	4	4	4
4	4	3	3	2	3	4	4	4	2	3.5	3	3	4	4	3.5
5	2	3	3	4	3	3	3	3	3	3	2	3	3	3	2.75
6	2	4	3	4	3.25	3	3	4	4	3.5	4	4	3	3	3.5
7	3	4	4	3	3.5	4	5	5	3	4.25	4	4	4	4	4
8	1	5	5	4	3.75	1	2	2	1	1.5	2	2	2	2	2
9	2	4	4	4	3.5	3	3	2	2	2.5	3	3	3	3	3
10	2	3	4	3	3	3	3	2	4	3	3	2	3	2	2.5
11	4	3	3	4	3.5	4	4	4	2	3.5	2	2	3	4	2.75
12	2	5	3	4	3.5	2	2	3	4	2.75	4	4	2	1	2.75
13	2	4	4	4	3.5	3	3	4	3	3.25	3	3	3	3	3
Group <4> 1	2	5	3	1	2.75	3	2	2	2	2.25	3	3	2	2	2.5
2	1	4	2	2	2.25	4	2	3	2	2.75	3	2	3	2	2.5
3	2	4	2	2	2.5	3	2	3	2	2.5	3	2	3	2	2.5
4	4	2	5	4	3.75	5	5	5	3	4.5	5	5	5	5	5
5	4	3	2	3	3	4	4	4	3	3.75	3	2	4	4	3.25
6	3	3	4	3	3.25	4	4	4	3	3.75	4	4	3	4	3.75
7	2	3	3	4	3	3	4	4	4	3.75	3	3	2	3	2.75
8	2	3	2	2	2.25	2	1	2	3	2	3	2	2	2	2.25
9	2	3	5	4	3.5	5	4	4	3	4	4	5	5	3	4.25
10	2	4	4	5	3.75	3	4	3	4	3.5	4	4	4	3	3.75
11	1	2	4	4	2.75	4	4	5	3	4	3	4	4	3	3.5
12	2	4	4	2	3	4	4	4	2	3.5	3	4	3	4	3.5
13	2	4	2	2	2.5	3	4	3	2	3	3	2	3	2	2.5

3. Post-questionnaire –Misuse Case

Misuse Cases	MUC_PU 1	MUC_PU2	MUC_PU3	MUC_PU4	UC_Average I	MUC_ITU 1	MUC_ITU2	MUC_ITU3	MUC_ITU4	_AVERAGE	MUC_PEOU1	MUC_PEOU2	MUC_PEOU3	MUC_PEOU4	UC_AVERAG
Group 1 <1>	1	4	2	3	2.5	2	3	2	2	2.25	2	2	3	2	2.25
2	2	4	2	2	2.5	2	2	3	1	2	2	2	3	4	2.75
3	1	5	1	2	2.25	1	1	1	1	1	1	1	1	1	1
4	1	5	2	1	2.25	2	4	2	2	2.5	1	1	1	2	1.25
5	1	4	3	2	2.5	1	2	2	1	1.5	1	1	2	1	1.25
6	2	4	2	2	2.5	2	2	2	2	2	2	2	3	2	2.25
7	2	4	2	3	2.75	1	2	1	1	1.25	1	3	2	2	2
8	2	5	1	3	2.75	2	4	1	1	2	2	3	3	1	2.25
9	1	4	3	2	2.5	2	3	2	2	2.25	2	2	1	4	2.25
10	1	3	3	3	2.5	2	3	2	4	2.75	1	1	2	2	1.5
11	1	4	1	1	1.75	2	2	2	2	2	1	1	2	2	1.5
12	2	4	3	2	2.75	2	2	3	2	2.25	1	2	3	3	2.25
13	4	3	1	2	2.5	5	4	4	2	3.75	3	2	3	4	3
Group <2>1	1	4	2	1	2	1	1	3	2	1.75	1	2	2	3	2
2	1	4	3	2	2.5	4	4	2	2	3	2	2	2	2	2
3	1	4	2	1	2	1	1	1	1	1	1	1	3	1	1.5
4	1	5	2	2	2.5	2	2	2	1	1.75	1	2	3	2	2
5	1	4	2	2	2.25	2	2	2	1	1.75	2	2	2	2	2
6	2	2	3	2	2.25	2	2	3	1	2	1	2	4	4	2.75
7	2	3	3	3	2.75	4	4	4	2	3.5	4	2	4	4	3.5
8	2	4	3	2	2.75	2	2	2	2	2	2	3	3	3	2.75
9	2	3	2	1	2	2	2	2	2	2	2	2	3	2	2.25
10	2	3	2	2	2.25	3	3	3	2	2.75	3	2	2	2	2.25
11	3	4	2	1	2.5	2	2	3	2	2.25	1	2	2	2	1.75
Group <3> 1	2	4	3	3	3	2	2	2	2	2	2	3	2	1	2
2	4	3	3	2	3	2	2	2	2	2	1	3	2	2	2
3	2	4	3	1	2.5	2	4	2	2	2.5	2	3	3	2	2.5
4	1	2	3	2	2	3	2	2	2	2.25	2	2	4	3	2.75
5	2	3	3	3	2.75	3	3	2	3	2.75	3	3	4	3	3.25
6	1	4	3	3	2.75	2	3	2	2	2.25	3	3	2	2	2.5
7	3	4	1	1	2.25	3	2	2	1	2	1	2	3	1	1.75
8	1	4	2	2	2.25	3	3	3	1	2.5	2	3	3	3	2.75
9	1	3	2	2	2	2	3	2	2	2.25	2	2	3	3	2.5
10	3	4	3	3	3.25	2	2	2	2	2	2	3	3	3	2.75
11	2	4	3	2	2.75	3	2	2	2	2.25	2	2	3	3	2.5
12	2	4	3	4	3.25	2	2	2	1	1.75	1	2	3	3	2.25
13	4	4	3	4	3.75	3	3	4	2	3	4	2	3	3	3
Group <4> 1	1	5	2	2	2.5	2	1	2	1	1.5	1	1	1	1	1
2	1	4	2	2	2.25	3	2	3	2	2.5	2	2	4	2	2.5
3	2	4	3	1	2.5	2	2	2	1	1.75	1	2	3	2	2
4	1	4	3	2	2.5	2	3	3	2	2.5	2	3	2	2	2.25
5	1	4	3	2	2.5	2	2	2	2	2	1	3	3	2	2.25
6	2	4	2	2	2.5	3	2	3	2	2.5	2	2	3	3	2.5
7	1	4	2	2	2.25	2	2	1	1	1.5	2	2	2	1	1.75
8	1	3	2	2	2	2	2	2	1	1.75	1	2	2	1	1.5
9	1	5	2	2	2.5	2	3	2	1	2	1	3	2	1	1.75
10	1	5	2	2	2.5	1	2	2	1	1.5	2	2	3	2	2.25
11	1	4	3	1	2.25	2	2	3	2	2.25	2	2	2	3	2.25
12	2	4	3	2	2.75	3	2	2	2	2.25	2	3	3	2	2.5
13	1	4	3	3	2.75	2	2	2	2	2	2	1	2	2	1.75

Appendix C

Analysed Data

Appendix C – Analysed Data

1. Results of Kruskal-Wallis H tests - four independent groups background variables.

Knowledgeable about 7 modelling approaches (pre-questionnaire)

Kruskal-Wallis Test: System modelling versus Group

Kruskal-Wallis Test on System modelling

Group	N	Median	Ave Rank	Z
1	13	2,000	20,0	-1,59
2	11	3,000	31,4	1,52
3	13	3,000	24,0	-0,43
4	13	3,000	27,5	0,59
Overall	50		25,5	

H = 4,08 DF = 3 P = 0,253

H = 5,69 DF = 3 P = 0,128 (adjusted for ties)

Kruskal-Wallis Test: UML use case diagram versus Group

Kruskal-Wallis Test on UML use case diagram

Group	N	Median	Ave Rank	Z
1	13	3,000	24,8	-0,19
2	11	3,000	29,5	1,04
3	13	3,000	23,4	-0,61
4	13	3,000	24,8	-0,19
Overall	50		25,5	

H = 1,17 DF = 3 P = 0,759

H = 2,63 DF = 3 P = 0,452 (adjusted for ties)

Kruskal-Wallis Test: Miuse case diagram versus Group

Kruskal-Wallis Test on Miuse case diagram

Group	N	Median	Ave Rank	Z
1	13	1,000	27,7	0,63
2	11	1,000	22,3	-0,83
3	13	1,000	23,8	-0,48
4	13	1,000	27,7	0,63
Overall	50		25,5	

H = 1,29 DF = 3 P = 0,730

H = 2,51 DF = 3 P = 0,473 (adjusted for ties)

Kruskal-Wallis Test: UML activity diagrams versus Group

Kruskal-Wallis Test on UML activity diagrams

Group	N	Median	Ave Rank	Z
1	13	2,000	31,2	1,65
2	11	2,000	22,2	-0,84
3	13	2,000	21,3	-1,19
4	13	2,000	26,7	0,34
Overall	50		25,5	

H = 3,71 DF = 3 P = 0,295

H = 4,18 DF = 3 P = 0,243 (adjusted for ties)

Kruskal-Wallis Test: Mal-activity diagrams versus Group

Kruskal-Wallis Test on Mal-activity diagrams

Group	N	Median	Ave Rank	Z
1	13	1,000	25,9	0,12
2	11	1,000	24,0	-0,39
3	13	1,000	25,9	0,12
4	13	1,000	25,9	0,12
Overall	50		25,5	

H = 0,15 DF = 3 P = 0,985

H = 0,88 DF = 3 P = 0,830 (adjusted for ties)

Kruskal-Wallis Test: Secure Tropos diagrams versus Group

49 cases were used

1 cases contained missing values

Kruskal-Wallis Test on Secure Tropos diagrams

Group	N	Median	Ave Rank	Z
1	12	1,000	24,5	-0,14
2	11	1,000	24,5	-0,13
3	13	1,000	26,4	0,41
4	13	1,000	24,5	-0,15
Overall	49		25,0	

H = 0,17 DF = 3 P = 0,983

H = 2,77 DF = 3 P = 0,429 (adjusted for ties)

Kruskal-Wallis Test: i* modelling diagrams versus Group

Kruskal-Wallis Test on i* modelling diagrams

Group	N	Median	Ave Rank	Z
1	13	1,000	22,0	-1,01
2	11	1,000	28,4	0,75
3	13	1,000	25,9	0,11
4	13	1,000	26,2	0,19
Overall	50		25,5	

H = 1,22 DF = 3 P = 0,748

H = 3,36 DF = 3 P = 0,339 (adjusted for ties)

Kruskal-Wallis Test: Semesters versus Group

Kruskal-Wallis Test on Semesters

Group	N	Median	Ave Rank	Z
1	13	3,000	29,2	1,06
2	11	3,000	26,2	0,19
3	13	3,000	20,7	-1,38
4	13	3,000	26,0	0,14
Overall	50		25,5	

H = 2,29 DF = 3 P = 0,514

H = 5,11 DF = 3 P = 0,164 (adjusted for ties)

Test the work experience of the groups.

Kruskal-Wallis Test: Jobs versus Group

47 cases were used

3 cases contained missing values

Kruskal-Wallis Test on Jobs

Group	N	Median	Ave Rank	Z
1	13	0,000000000	24,7	0,23
2	11	0,000000000	24,3	0,09
3	12	0,000000000	22,9	-0,32
4	11	0,000000000	24,0	0,00
Overall	47		24,0	

H = 0,12 DF = 3 P = 0,990

H = 0,19 DF = 3 P = 0,979 (adjusted for ties)

Comparison of paired modelling approach

Wilcoxon Signed Rank Test: Diff UseCase v.s. i* modelling

```
Test of median = 0.000000 versus median not = 0.000000

                                  N for   Wilcoxon         Estimated
                             N   Test  Statistic      P     Median
Diff UseCase v.s. i* modelling  50     46     1081.0  0.000      2.000
```

Wilcoxon Signed Rank Test: Diff. MUC v.s. ST

```
Test of median = 0.000000 versus median not = 0.000000

                            N for   Wilcoxon            Estimated
                     N  N*   Test  Statistic      P        Median
Diff. MUC v.s. ST   49   1      9       45.0  0.009  0.000000000
```

Wilcoxon Signed Rank Test: Diff Activity v.s. Mal-activity

```
Test of median = 0.000000 versus median not = 0.000000

                                    N for   Wilcoxon         Estimated
                               N   Test  Statistic      P     Median
Diff Activity v.s. Mal-activity  50     34      595.0  0.000      1.000
```

2. Performance

Compare the number of identified threats and mitigations for the two modelling approaches.

Wilcoxon Signed Rank Test: Diff THR(ST) V.S. THR(MUC)

```
Test of median = 0.000000 versus median not = 0.000000

                              N for   Wilcoxon         Estimated
                         N   Test  Statistic      P     Median
Diff THR(ST) V.S. THR(MUC)  50     36      231.5  0.113    -0.5000
```

Wilcoxon Signed Rank Test: Diff MIT(ST) V.S. MIT(MUC)

```
Test of median = 0.000000 versus median not = 0.000000

                              N for   Wilcoxon         Estimated
                         N   Test  Statistic      P     Median
Diff MIT(ST) V.S. MIT(MUC)  50     40      200.0  0.005    -0.5000
```

Wilcoxon Signed Rank Test: Diff THMI(ST) V.S. THMI(MUC)

```
Test of median = 0.000000 versus median not = 0.000000
```

		N for	Wilcoxon		Estimated
	N	Test	Statistic	P	Median
Diff THMI(ST) V.S. THMI(MUC)	50	39	184.0	0.004	-1.000

Compare the number of identified threats and mitigation between the first modelling technique with the second modelling technique.

Wilcoxon Signed Rank Test: Diff First v.s. Second (THR)

Test of median = 0.000000 versus median not = 0.000000

		N for	Wilcoxon		Estimated
	N	Test	Statistic	P	Median
Diff First v.s. Second (THR)	50	36	446.5	0.076	0.5000

Wilcoxon Signed Rank Test: Diff First v.s. Second (MIT)

Test of median = 0.000000 versus median not = 0.000000

		N for	Wilcoxon		Estimated
	N	Test	Statistic	P	Median
Diff First v.s. Second (MIT)	50	40	593.0	0.014	0.5000

Wilcoxon Signed Rank Test: Diff First v.s. Second (THMI)

Test of median = 0.000000 versus median not = 0.000000

		N for	Wilcoxon		Estimated
	N	Test	Statistic	P	Median
Diff First v.s. Second (THMI)	50	39	572.5	0.011	1.000

3. Estimated use of diagrams, textual description and memory

Descriptive Statistics: THR_DIAG(ST)

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
THR_DIAG(ST)	50	0	29.00	3.69	26.07	0.00	8.75	25.00	42.50

Variable	Maximum
THR_DIAG(ST)	100.00

Descriptive Statistics: MIT_DIAG(ST)

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
MIT_DIAG(ST)	50	0	20.70	3.44	24.33	0.00	0.00	10.00	30.00

Variable	Maximum
MIT_DIAG(ST)	100.00

Descriptive Statistics: THR_DIAG (MUC)

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
THR_DIAG (MUC)	50	0	44.54	3.86	27.27	0.00	20.00	50.00	70.00

Variable	Maximum
THR_DIAG (MUC)	90.00

Descriptive Statistics: MIT_DIAG (MUC)

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
MIT_DIAG (MUC)	49	1	19.23	2.75	19.24	0.00	1.25	15.00	30.00

Variable	Maximum
MIT_DIAG (MUC)	80.00

Compare the different between using THR and MIT with relevant methods (diagrams, textual description and memory) to identify threats and mitigation.

Wilcoxon Signed Rank Test: Diff THR_TXT ST v.s. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff THR_TXT ST v.s. MUC	50	45	816.5	0.001	15.00

Wilcoxon Signed Rank Test: Diff THR_MEM ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff THR_MEM ST V.S. MUC	50	38	372.0	0.988	0.000000000

Wilcoxon Signed Rank Test: Diff MIT_MEM ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff MIT_MEM ST V.S. MUC	49	1	37	202.5	0.025	-5.000

Wilcoxon Signed Rank Test: Diff MIT_VUL ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff MIT_VUL ST V.S. MUC	49	1	35	302.5	0.844	0.000000000

Wilcoxon Signed Rank Test: Diff MIT_TXT ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000
                                    N for   Wilcoxon         Estimated
                           N  N*   Test  Statistic      P     Median
Diff MIT_TXT ST V.S. MUC  49   1     34      444.0  0.013      7.500
```

Wilcoxon Signed Rank Test: Diff THR_DIAG (ST V.S. MUC)

```
Test of median = 0.000000 versus median not = 0.000000
                                  N for   Wilcoxon         Estimated
                             N   Test  Statistic      P     Median
Diff THR_DIAG (ST V.S. MUC)  50     47      284.5  0.003     -15.00
```

Wilcoxon Signed Rank Test: Diff MIT_DIAG (ST V.S. MUC)

```
Test of median = 0.000000 versus median not = 0.000000
                                     N for   Wilcoxon         Estimated
                             N  N*   Test  Statistic      P     Median
Diff MIT_DIAG (ST V.S. MUC)  49   1     40      375.0  0.643     -2.500
```

Wilcoxon Signed Rank Test: THR_DIAG Diff NS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000
                                    N for   Wilcoxon         Estimated
                            N  N*   Test  Statistic      P     Median
THR_DIAG Diff NS ST V.S. MUC  24   2     21       54.5  0.035     -15.00
```

Wilcoxon Signed Rank Test: THR_DIAG Diff HIS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000
                                      N for   Wilcoxon         Estimated
                              N  N*   Test  Statistic      P     Median
THR_DIAG Diff HIS ST V.S. MUC  24   2     21       62.0  0.065     -15.00
```

Wilcoxon Signed Rank Test: THR_TXT Diff NS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000
                                    N for   Wilcoxon         Estimated
                            N  N*   Test  Statistic      P     Median
THR_TXT Diff NS ST V.S. MUC  24   2     21      172.5  0.050      15.00
```

Wilcoxon Signed Rank Test: THR_TXT Diff HIS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000
                                    N for   Wilcoxon         Estimated
                            N  N*   Test  Statistic      P     Median
```

```
THR_TXT Diff HIS ST V.S. MUC  24   2     23      222.0  0.011      17.50
```

Wilcoxon Signed Rank Test: THR_MEM NS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000

                                N for   Wilcoxon           Estimated
                          N  N*   Test  Statistic      P       Median
THR_MEM NS ST V.S. MUC   24   2     22      128.5  0.961  0.000000000
```

Wilcoxon Signed Rank Test: THR_MEM HIS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000

                                  N for   Wilcoxon           Estimated
                            N  N*   Test  Statistic      P       Median
THR_MEM HIS ST V.S. MUC   24   2     21      114.5  0.986  0.000000000
```

Wilcoxon Signed Rank Test: MIT_DIAG NS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000

                                  N for   Wilcoxon         Estimated
                            N  N*   Test  Statistic      P     Median
MIT_DIAG NS ST V.S. MUC   23   3     19      108.5  0.601      2.500
```

Wilcoxon Signed Rank Test: MIT_DIAG Diff HIS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000

                                       N for   Wilcoxon         Estimated
                                 N  N*   Test  Statistic      P     Median
MIT_DIAG Diff HIS ST V.S. MUC   24   2     22      110.5  0.615     -5.000
```

Wilcoxon Signed Rank Test: MIT_TXT Diff NS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000

                                    N for   Wilcoxon         Estimated
                              N  N*   Test  Statistic      P     Median
MIT_TXT Diff NS ST V.S. MUC  23   3     22      155.0  0.363      5.000
```

Wilcoxon Signed Rank Test: MIT_TXT Diff HIS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000

                                    N for   Wilcoxon         Estimated
                              N  N*   Test  Statistic      P     Median
MIT_TXT Diff HIS ST V.S. MUC  24   2     20      153.0  0.076      7.500
```

Wilcoxon Signed Rank Test: MIT_THR NS ST V.S. MUC

```
Test of median = 0.000000 versus median not = 0.000000
```

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
MIT_THR NS ST V.S. MUC	23	3	21	68.0	0.102	-10.00

Wilcoxon Signed Rank Test: MIT_THR HIS ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
MIT_THR HIS ST V.S. MUC	24	2	19	113.0	0.481	5.000

Wilcoxon Signed Rank Test: MIT_MEM Diff NS ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
MIT_MEM Diff NS ST V.S. MUC	23	3	20	82.0	0.401	-7.500

Wilcoxon Signed Rank Test: MIT_MEM Diff HIS ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
MIT_MEM Diff HIS ST V.S. MUC	24	2	20	87.0	0.514	-5.000

Compare the different between the Cases with same modelling technique when estimating the usage of diagrams, textual description and memory.

*** Wilcoxon Signed Rank Test: Diff THR_DIAG ST NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff THR_DIAG ST NS V.S. HIS	24	2	21	50.5	0.025	-15.00

*** Wilcoxon Signed Rank Test: Diff THR_DIAG MUC NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff THR_DIAG MUC NS V.S. HIS	24	2	21	74.0	0.154	-15.00

*** Wilcoxon Signed Rank Test: Diff THR_TXT ST NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

			N for	Wilcoxon		Estimated

```
                                N  N*   Test  Statistic      P     Median
Diff THR_TXT ST NS V.S. HIS    24   2     24      197.5  0.179      10.00
```

*** Wilcoxon Signed Rank Test: Diff THR_TXT MUC NS V.S. HIS**

```
Test of median = 0.000000 versus median not = 0.000000

                                      N for   Wilcoxon         Estimated
                                N  N*   Test  Statistic      P     Median
Diff THR_TXT MUC NS V.S. HIS   24   2     22      209.5  0.007      15.00
```

*** Wilcoxon Signed Rank Test: Diff THR_MEM ST NS V.S. HIS**

```
Test of median = 0.000000 versus median not = 0.000000

                                    N for   Wilcoxon          Estimated
                                N  N*   Test  Statistic      P     Median
Diff THR_MEM ST NS V.S. HIS    24   2     20      116.5  0.681      5.000
```

*** Wilcoxon Signed Rank Test: Diff THR_MEM MUC NS V.S. HIS**

```
Test of median = 0.000000 versus median not = 0.000000

                                   N for   Wilcoxon           Estimated
                               N  N*   Test  Statistic      P       Median
Diff THR_MEM MUC NS V.S. HIS  24   2     22      130.0  0.922  0.000000000
```

*** Wilcoxon Signed Rank Test: Diff MIT_DIAG ST NS V.S. HIS**

```
Test of median = 0.000000 versus median not = 0.000000

                                      N for   Wilcoxon         Estimated
                                N  N*   Test  Statistic      P     Median
Diff MIT_DIAG ST NS V.S. HIS   24   2     20       85.0  0.467     -5.000
```

*** Wilcoxon Signed Rank Test: Diff MIT_DIAG MUC NS V.S. HIS**

```
Test of median = 0.000000 versus median not = 0.000000

                                       N for   Wilcoxon         Estimated
                                 N  N*   Test  Statistic      P     Median
Diff MIT_DIAG MUC NS V.S. HIS   23   3     19       47.0  0.056     -10.00
```

*** Wilcoxon Signed Rank Test: Diff MIT_TXT ST NS V.S. HIS**

```
Test of median = 0.000000 versus median not = 0.000000

                                    N for   Wilcoxon           Estimated
                                N  N*   Test  Statistic      P       Median
Diff MIT_TXT ST NS V.S. HIS    24   2     20      108.0  0.926  0.000000000
```

*** Wilcoxon Signed Rank Test: Diff MIT_TXT MUC NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

			N for	Wilcoxon		Estimated
	N	N*	Test	Statistic	P	Median
Diff MIT_TXT MUC NS V.S. HIS	23	3	20	130.0	0.360	5.000

*** Wilcoxon Signed Rank Test: Diff MIT_THR ST NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

			N for	Wilcoxon		Estimated
	N	N*	Test	Statistic	P	Median
Diff MIT_THR ST NS V.S. HIS	24	2	21	108.5	0.821	0.000000000

*** Wilcoxon Signed Rank Test: Diff MIT_THR MUC NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

			N for	Wilcoxon		Estimated
	N	N*	Test	Statistic	P	Median
Diff MIT_THR MUC NS V.S. HIS	23	3	19	126.5	0.212	10.00

*** Wilcoxon Signed Rank Test: Diff MIT_MEM ST NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

			N for	Wilcoxon		Estimated
	N	N*	Test	Statistic	P	Median
Diff MIT_MEM ST NS V.S. HIS	24	2	21	110.0	0.862	-5.000

*** Wilcoxon Signed Rank Test: Diff MIT_MEM MUC NS V.S. HIS**

Test of median = 0.000000 versus median not = 0.000000

			N for	Wilcoxon		Estimated
	N	N*	Test	Statistic	P	Median
Diff MIT_MEM MUC NS V.S. HIS	23	3	22	116.5	0.758	-5.000

4. Perception – post-questionnaire

Descriptive Statistics: PU_1, PU_2, PU_3, PU_4 (ST)

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
PU_1	50	0	2.060	0.126	0.890	1.000	1.000	2.000	2.000	4.000
PU_2	50	0	3.220	0.132	0.932	1.000	3.000	3.000	4.000	4.000
PU_3	50	0	3.300	0.104	0.735	1.000	3.000	3.000	4.000	4.000

PU_4	50	0	3.040	0.162	1.142	1.000	2.000	4.000	4.000	4.000

Descriptive Statistics: ITU_1, ITU_2, ITU_3, ITU_4 (ST)

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
ITU_1	50	0	2.940	0.129	0.913	1.000	2.000	3.000	4.000	4.000
ITU_2	50	0	2.880	0.130	0.918	1.000	2.000	3.000	4.000	4.000
ITU_3	50	0	2.840	0.141	0.997	1.000	2.000	3.000	4.000	4.000
ITU_4	50	0	2.640	0.124	0.875	1.000	2.000	2.500	3.000	4.000

Descriptive Statistics: PEOU_1, PEOU_2, PEOU_3, PEOU_4 (ST)

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
PEOU_1	50	0	2.880	0.130	0.918	1.000	2.000	3.000	4.000	4.000
PEOU_2	50	0	2.880	0.133	0.940	1.000	2.000	3.000	4.000	4.000
PEOU_3	50	0	2.960	0.118	0.832	1.000	2.750	3.000	4.000	4.000
PEOU_4	50	0	2.800	0.121	0.857	1.000	2.000	3.000	3.000	4.000

Compare the different of perceived usefulness between ST and MUC.

Wilcoxon Signed Rank Test: Diff PU1 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PU1 ST V.S. MUC	50	30	345.0	0.021	0.5000

Wilcoxon Signed Rank Test: Diff PU2 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PU2 ST V.S. MUC	50	26	99.0	0.054	0.000000000

Wilcoxon Signed Rank Test: Diff PU3 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PU3 ST V.S. MUC	50	30	427.0	0.000	0.5000

Wilcoxon Signed Rank Test: Diff PU4 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

		N for	Wilcoxon		Estimated

	N	Test	Statistic	P	Median
Diff PU4 ST V.S. MUC	50	38	678.0	0.000	1.000

Compare the different of intended to use between ST and MUC.

Wilcoxon Signed Rank Test: Diff ITU1 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff ITU1 ST V.S. MUC	50	39	690.5	0.000	1.000

Wilcoxon Signed Rank Test: Diff ITU2 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff ITU2 ST V.S. MUC	50	35	525.0	0.001	0.5000

Wilcoxon Signed Rank Test: Diff ITU3 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff ITU3 ST V.S. MUC	50	33	514.0	0.000	1.000

Wilcoxon Signed Rank Test: Diff ITU4 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff ITU4 ST V.S. MUC	50	34	561.0	0.000	1.000

Compare the different of perceive ease of use between ST and MUC.

Wilcoxon Signed Rank Test: Diff PEOU1 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PEOU1 ST V.S. MUC	50	41	816.5	0.000	1.500

Wilcoxon Signed Rank Test: Diff PEOU2 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

		N for	Wilcoxon		Estimated

	N	Test	Statistic	P	Median
Diff PEOU2 ST V.S. MUC	50	40	764.0	0.000	1.000

Wilcoxon Signed Rank Test: Diff PEOU3 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PEOU3 ST V.S. MUC	50	30	388.5	0.001	0.5000

Wilcoxon Signed Rank Test: Diff PEOU4 ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PEOU4 ST V.S. MUC	50	35	490.0	0.004	0.5000

Compare the different of average perceived perceptions between ST and MUC.

Wilcoxon Signed Rank Test: Diff Average_PU ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PU ST V.S. MUC	50	45	970.0	0.000	0.5000

Wilcoxon Signed Rank Test: Diff Average_ITU ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff ITU ST V.S. MUC	50	46	994.0	0.000	0.8750

Wilcoxon Signed Rank Test: Diff Average_PEOU ST V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff PEOU ST V.S. MUC	50	43	899.0	0.000	0.8750

The Pooled standard deviation computing

MUC			ST			pooled standard deviation	Pooled Mean (that took Ns into consideration)	Average of the two means
mean	n	SD	mean	n	SD			
2.5	50	0.36	3.01	50	0.42	0.39	2.75500	2.755
2.18	50	0.54	3.06	50	0.70	0.63	2.62000	2.62
2.13	50	0.54	3.02	50	0.72	0.64	2.57500	2.575
2.34	50	0.51	3.03	50	0.62	0.57	2.68500	2.685

Sample Size Calculation:

$\alpha = 0.05, 1-\beta = 0.80$.

$$N = \frac{(sd_1^2 + sd_2^2) * (\alpha + \beta)^2}{(Mean_1 - Mean_2)^2}$$

Compare the different of perception between the cases with the same modelling technique.

ST:

Wilcoxon Signed Rank Test: Diff AVER.PU.ST NS V.S. HIS

```
Test of median = 0.000000 versus median not = 0.000000
                                      N for   Wilcoxon         Estimated
                              N  N*   Test  Statistic      P     Median
Diff AVER.PUST NS V.S. HIS   24   2     19       13.5  0.001    -0.3750
```

Wilcoxon Signed Rank Test: Diff ST.AVER. ITU NS V.S. HIS

```
Test of median = 0.000000 versus median not = 0.000000
                                        N for   Wilcoxon         Estimated
```

	N	N*	Test	Statistic	P	Median
Diff ST.AVER. ITU NS V.S. HIS	24	2	24	185.5	0.317	0.1250

Wilcoxon Signed Rank Test: Diff ST.AVER. PEOU NS V.S. MUC

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff ST.AVER. PEOU NS V.S. MUC	24	2	23	185.0	0.157	0.3750

MUC:

Wilcoxon Signed Rank Test: Diff AVER.MUC.PU NS V.S. HIS

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff AVER.MUC.PU NS V.S. HIS	24	2	20	124.0	0.490	0.000000000

Wilcoxon Signed Rank Test: Diff AVER.MUC.ITU NS V.S. HIS

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff AVER.MUC.ITU NS V.S. HIS	24	2	18	112.0	0.258	0.1250

Wilcoxon Signed Rank Test: Diff AVER.MUC.PEOU NS V.S. HIS

Test of median = 0.000000 versus median not = 0.000000

	N	N*	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff AVER.MUC.PEOU NS V.S. HIS	24	2	22	205.5	0.011	0.3750

Appendix D

RCIS 2012 - Conference Paper

A Comparative Review of i*-based and Use Case-based Security Modelling Initiatives

Abstract Security requirements elicitation and modelling are integral for the successful development of secure systems. However, there are a lot of similar yet not identical approaches that currently exist for security requirements modelling, which is confusing for researchers and practitioners hence some characterisation will be useful to give better overview and understand advantages and disadvantages of various approaches. This paper provides a comparative review of i*-based and use caseŰbased security modelling initiatives, using a characterisation framework with several dimensions. Our findings show that both categories of initiatives have significant conceptual similarities in the aspect of modelling language and method process, and coverage of security requirements modelling notions They have conceptual differences in terms of their capacity to support: formal and informal approaches, threats identification and mitigation, identification of security goals, different software engineering development methods, and software evolution.

Keywords: security requirements, security requirements modelling, i*-based modelling, use-case based modelling.

A Comparative Review of i*-based and Use Case-based Security Modelling Initiatives

Olawande Daramola, Yushan Pan
wande@idi.ntnu.no, yushan@stud.ntnu.no
Department of Computer and Information Science
Norwegian University of Science and Technology, Norway

Peter Karpati, Guttorm Sindre
{Kpeter,guttors}@idi.ntnu.no
Department of Computer and Information Science
Norwegian University of Science and Technology, Norway

Abstract - Security requirements elicitation and modelling are integral for the successful development of secure systems. However, there are a lot of similar yet not identical approaches that currently exist for security requirements modelling, which is confusing for researchers and practitioners hence some characterisation will be useful to give a better overview and understanding of advantages and disadvantages of various approaches. This paper provides a comparative review of i*-based and use case–based security modelling initiatives, using a characterisation framework with several dimensions. Our findings show that both categories of initiatives have significant conceptual similarities in the aspect of modelling language and method process, and coverage of security requirements modelling notions. They have conceptual differences in the aspect of: representation perspective, kind of security requirements engineering activities that are supported, the quality of specification that is generated and the specification techniques used, and the degree of support for software evolution.

Keywords- security requirements, security requirements modelling, i*-based modelling, use-case based modelling.

I. INTRODUCTION

Increased scope of connectivity, interoperability, extendibility, and complexity of software systems has also amplified the threats to their security. In recent times, security requirements (SR) elicitation and modelling has been gaining increasing importance as an integral part of the development of secure software systems [1]. A positive trend in the field of security requirements engineering is the existence of many SR elicitation and modelling initiatives. Some of these approaches also appear similar but not identical, hence researchers and practitioners must find adequate basis to reason about them, apply them, combine them, and make decision for their adoption when necessary. The question is: "*are there conceptual differences and similarities between these SR modelling initiatives that could provide a better understanding of their capabilities, advantages and disadvantages, and whether they should be considered competitors (i.e., different approaches for the same modelling needs) or complementary (i.e., both approaches together covering a modelling need better than any of the approaches alone).*"

Two of the more prominent categories of SR modelling approaches are the i*-based, and the use case-based SR modelling initiatives. In pursuit of our motivation for this work, we selected to study these two categories of SR modelling initiatives. This is because they are based on well established modelling frameworks (i* Agent modelling framework [2] and Use Cases [3]), and have attracted appreciable interest in the literature, and also in practise [4][1].

Comparisons between different modelling languages can be performed in several different ways. The main distinction is between *empirical comparisons*, e.g., trying both (groups of) techniques in controlled experiments or case studies, and *analytical comparisons* where the techniques are evaluated theoretically, preferably according to some pre-established framework. As argued in [5] these two types of evaluations go hand in hand. Empirical evaluations would have the advantage of showing how techniques perform in practice; however, direct comparability between two techniques is only achieved with a controlled experiment. The limited duration of such experiments means that the experimental tasks must be fairly simple, and only a few aspects of the candidate modelling languages can be assessed in each experiment. Moreover, empirical studies tend to be more costly than analytical comparisons. It therefore makes sense to start out with an analytical comparison, which can compare more broadly a number of aspects of the candidate languages. This could then give insights into essential differences from which to make hypotheses for empirical investigations. The decision in this paper is therefore to go for an analytical comparison.

The next question then becomes what framework to use for the comparison. Often used in analytical evaluations of modelling languages is the BWW ontology [6], as was for instance applied for evaluating UML in [7]. This ontology focuses on the concepts found in a modelling language. Another effort, rather focussing on the visual representation of concepts, is [8]. However, both these approaches focus only on the modelling language (either conceptually or visually), not on other issues like development method, integration with other approaches, etc., and both of them are for evaluating modelling approaches in general, not specifically tailored for security-related modelling. A framework directly oriented

[1] iStar Showcase '11, Exploring the Goals of your Systems and Businesses - Practical experiences with i* modelling, http://www.cs.toronto.edu/km/istar/iStarShowcase_Proceedings.pdf