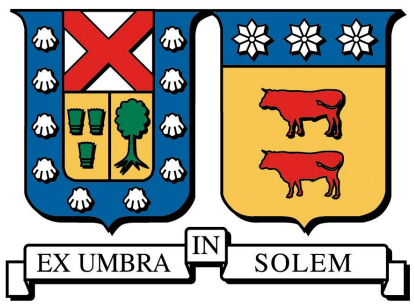
Universidad Técnica Federico Santa María
Departamento de Informática
Valparaíso, Chile



Tactics Selection Poker, TaSPer Technique for selecting tactics of software architecture by consensus (a security approach).

Felipe Arturo Osses Leinenweber

Tesis para optar al Grado de
Magíster en Ciencias de la Ingeniería Informática

Profesor Guía: Hernán Astudillo R, Ph.D.

10 de junio de 2021

Acknowledgements

I want to express my sincere gratitude to the Navy especially to Mr. RGR and Mr. PIA; to the CSIRT in a special way to Claudio; to the UTFSM especially Pabla, TOESKA and all those who helped me from day one; to my professors especially Dr. Hernán Astudillo who was always available to support my decisions, Dr. Marcello Visconti, Dr. Claudio Torres, Dr. Marcelo Mendoza and Dr. Raúl Monge; to my very big friends Juan Pablo and Ingrid and with a special mention to Gastón, without your support I would not have arrived here; Also and with all my love to my parents Arturo and Ingrid and grandparents Arturo, Juana, Alberto and Ruth; Last, but not least to my family, especially to Javiera, Trinidad and Santiago, I love you with all my heart. All of you were part of this thesis, thank you.

Abstract

Building secure software architectures requires taking several design decisions to achieve security requirements; these decisions must be revised carefully before agreement given their impact on system vulnerability and mission-readiness. Architects customarily take these resolutions, drawing upon specialized knowledge like architectural tactics for security; developers also have key information on platforms and tools actual performance, but their input may not be systematically considered to this end. This thesis presents Security Tactics Selection Poker (TaSPer), a consensual technique that extends the Planning Poker allowing development team members to identify, argue for, and choose among architectural security tactics according to objectives and priorities. First we conducted a case study to a group of nine Chilean Navy professional and then we conducted an experimental study establishing a process considering the development in five stages: definition, planning, evaluating, execution and analysis. For this, two pre-experiments were carried out with undergraduate and graduate students in different universities. Finally, we execute an experiment with twenty practitioners from IT Master Program, to assess the technique effectiveness in several scenarios.

Results show that TaSPeR (1) does support collaborative architectural decision-making, (2) encourages stakeholders participation, and (3) starts a group dynamics on how to act against threats. At the same time, it was possible to determine that the proposed technique allows decisions closer to Ground Truth (established by experts) compared to individual decisions made by the participants.

Thus, the use of a consensual technique for architectures evaluation seems to be a promising approach to establish secure software using architectural tactics.

Resumen

Construir software seguro requiere tomar decisiones de diseño que permitan cumplir los requerimientos de seguridad deseados; estas decisiones deben realizarse detenidamente antes de ser establecidas, lo anterior debido a que un mal diseño podría impactar gravemente el resultado final de un sistema, afectando los objetivos esperados.

Respecto al punto anterior, son los arquitectos quienes suelen tomar estas resoluciones, basándose tanto en su experiencias como en el uso de diferentes técnicas de arquitectura, como por ejemplo el uso de tácticas de arquitectura de software; por otro lado, los desarrolladores también ejecutan acciones claves al desarrollar el software, sin embargo es posible que estos no consideren de forma sistemática y primordial la importancia de un diseño seguro desde el punto de vista de las decisiones de diseño.

El trabajo realizado en esta tesis presenta la Técnica TaSPER (Security Tactics Selection Poker), técnica consensuada basada en Planning Poker, la cual permite que todas las personas involucradas en un desarrollo de software puedan identificar, argumentar, discutir y seleccionar las tácticas de seguridad para el desarrollo de arquitectura de acuerdo a los objetivos, requerimientos y prioridades establecidas.

Para lograr esto, primero se realizó un estudio de caso a un grupo de nueve profesionales de la Armada de Chile, para luego desarrollar un estudio experimental el cual consideró cinco etapas: definición, planificación, evaluación, ejecución y análisis. Para ello, se realizaron dos pre-experimentos con estudiantes de pregrado y postgrado en diferentes universidades y finalmente, se ejecutó un experimento con veinte profesionales del Programa del Magíster en TI con el objeto de evaluar la efectividad de la técnica en varios escenarios.

Los resultados muestran que la técnica TaSPeR (1) apoya la toma de decisiones arquitectónicas colaborativas, (2) fomenta la participación de los diferentes involucrados y (3) genera una dinámica grupal sobre cómo actuar contra las amenazas. Al mismo tiempo, se pudo determinar que la técnica propuesta permite decisiones más cercanas a Ground Truth (establecido por expertos) en comparación con decisiones individuales tomadas por los participantes.

Por tanto, el uso de una técnica consensuada para la evaluación de arquitecturas parece ser un enfoque prometedor para establecer de forma grupal la seguridad en el desarrollo de software.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter aims to describe the principal elements and concepts of this master thesis. Section 1.1 describes the main concepts used in this thesis; Section 1.2 details the general research question and hypothesis addressed; Section 1.3 describes the solution that addresses the research problem; Section 1.4 describes the main research objectives; Section 1.5 describes the main research questions; Section 1.6 introduces the main contributions; Section 1.7 describes the published works; And Section 1.8 describes the general structure of the thesis.

1.1 Background

This section introduces the most relevant concepts related to the proposal: software architecture; Quality attributes and non-functional requirements; Architectural tactics; Security tactics; Stakeholders and Consensus-based technique.

1.1.1 Software architecture

The architecture of a software system represents the connection between the business objectives, defined by the stakeholders, and the system's final result. Likewise, the system's software architecture is the set of structures needed to reason about the system, which comprises software elements, relations among them, and properties of both. At the same time, software architecture is a manifestation of the earliest design decisions about a system, and these early bindings carry enormous weight with respect to the system's remaining development, its deployment, and its maintenance life. It is also the earliest point at which these important design decisions affecting the system can be scrutinized. [5].

Additionally, it is important to consider that software architecture comprises the central structure of a system and the essential design decisions [44] related to components and structures that allow providing the required functionalities of a developing system. This is why design decisions fulfill a fundamental role in the

development, integration, evolution, and reuse of the architecture of a system [25], so that the dissociation between design decisions and architecture has improved costs related to changes, erosion of design, and limitations in its reuse, impacting the final result of a project.

1.1.2 Quality attributes and non-functional requirements

A quality attribute (QA) is a measurable or testable property of a system that is used to indicate how well the system satisfies the needs of its stakeholders. In other words, are characteristics that the system has, as opposed to what the system does, such as usability, maintainability, performance, reliability, and security [23]. Architects have no shortage of lists of quality attributes for software systems at their disposal. The standard with the pause-and-take-a-breath title of "ISO/IEC FCD 25010: Systems and software engineering—Systems and software product Quality Requirements and Evaluation (SQuaRE)—System and software quality models," is a good example [5] (see figure 1.1).

Related to quality attributes are the non-functional requirements, where in most cases, requirements and stakeholder's concerns are included. These non-functional requirements (or NFRs) play a critical role during system development, serving as selection criteria for choosing among alternatives designs and final implementation [31].

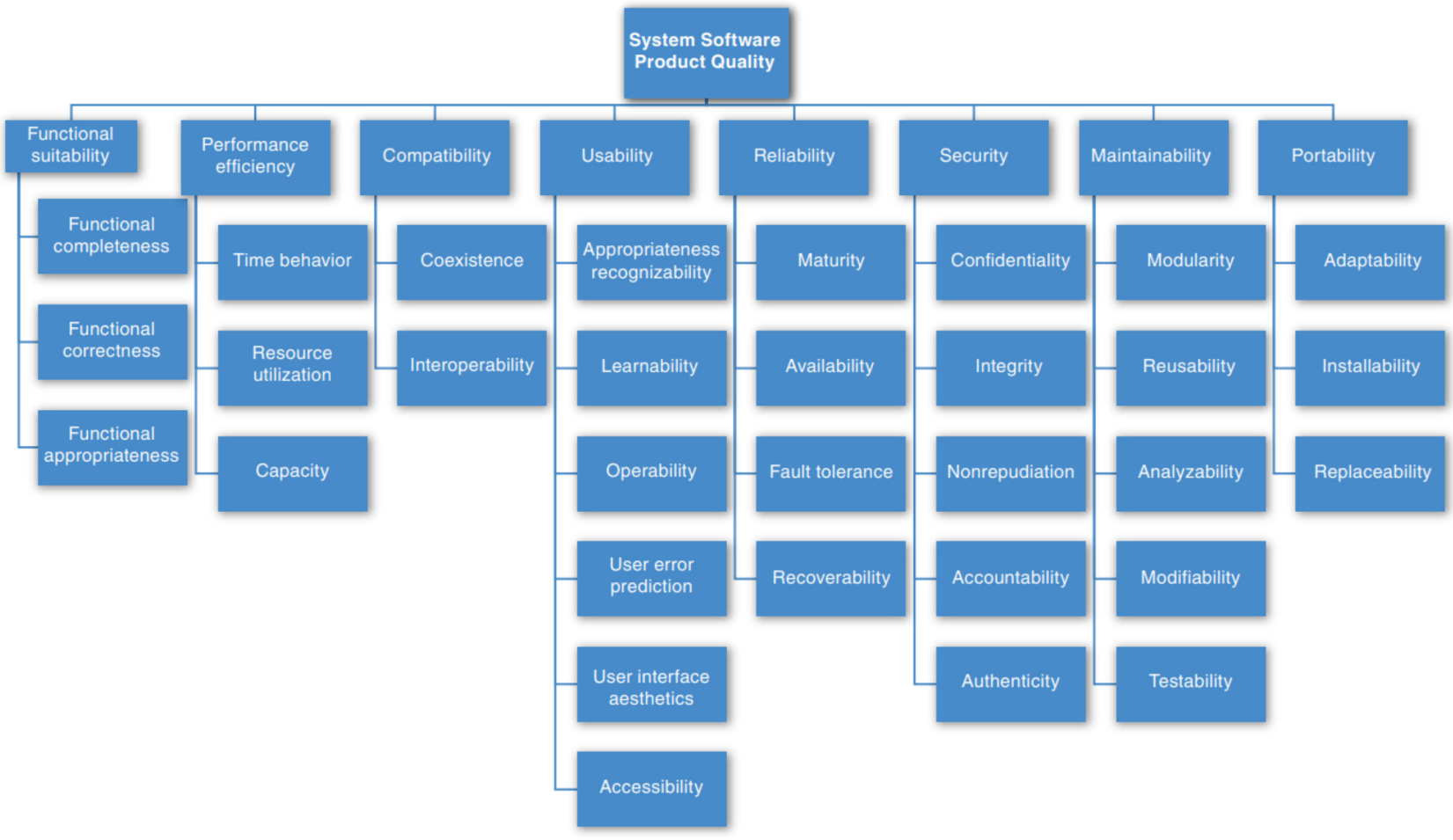


Figure 1.1: The ISO/IEC FCD 25010 product quality standard [5]

1.1.3 Architectural tactics

The architecture of a software system is the consequence of a succession of architectural design decisions. Among the design decisions that can be selected are architectural tactics, a reusable solution that describes which design alternatives help to achieve a quality attribute influencing the achievement of a response to the system, in other words, a tactic is a design decision that influences the achievement of a quality attribute response—tactics directly affect the system's response to some stimulus [5] (see figure 1.2). Software architects must understand the tactics and select proper design alternatives according to the type of applications or technologies they are reusing and the quality attributes they strive for [10]. Other authors suggest that a tactic is an architectural building block that provides a generic solution to guide issues related to quality attributes. The common point is that they provide solutions for compliance of quality attributes in a software system [28] [37].

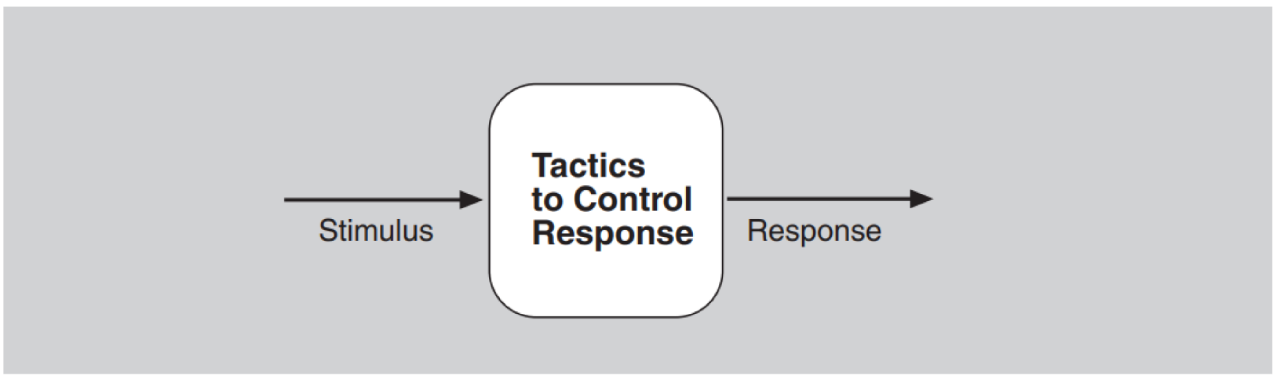


Figure 1.2: Tactics are intended to control responses to stimuli [5]

It is important to highlight that tactics are related to the different quality attributes that a system could possess, according to the work of Bass, Clement and Kazman [5], these can be classified as *availability*, *interoperability*, *modifiability*, *performance*, *testability*, *usability* and *security*.

In addition, it should be considered that tactics selection is influenced by the viewpoints of the project stakeholders [35] [36] [38], since each stakeholder has a different vision for requirements, this perspective can open other edges that the architect can analyze.

1.1.4 Security tactics

Regarding security, software architects face constant pressure to build secure software from their design on, where they must identify security requirements and adopt appropriate architectural solution, since a deficient architecture can generate opportunities and flexibility for malicious users to exploit system vulnerabilities affecting properties like confidentiality, integrity and availability.

The demand for secure software development has led to propose tactics for secure software architectures, initially by Bass et al. [4] and later on refined by themselves [5], Ryoo et al. [39], Fernandez et al. [17], and others.

Security tactics (and architectural tactics, in general) composed for a QA are organized in a taxonomy with categories for decisions to be taken, which group the tactics as options for that decision (see Figure 1.3). The usual categories for security tactics are: detect attacks, stop or mitigate attacks, react to attacks, and recover from attacks.

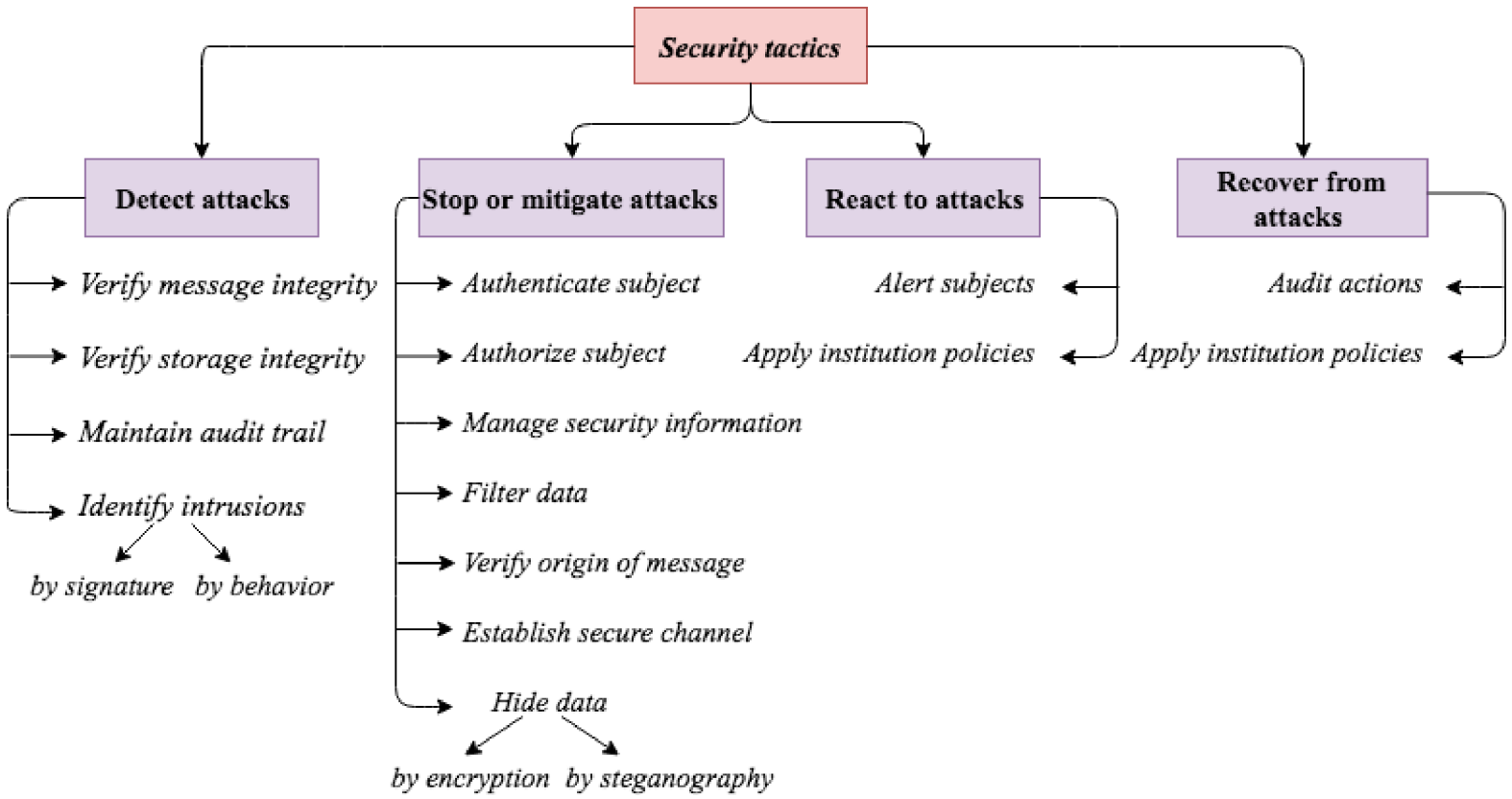


Figure 1.3: Security tactics proposed by Fernandez et al in [17]

1.1.5 Stakeholders

A stakeholder is anyone who has a stake in the success of the system: the customer, the end users, the developers, the project manager, the maintainers, and even those who market the system, for example. But stakeholders, despite all having a shared stake in the success of the system, typically have different specific concerns that they wish the system to guarantee or optimize. These concerns are as diverse as providing a certain behavior at runtime, performing well on a particular piece of hardware, being easy to customize, achieving short time to market or low cost of development, gainfully employing programmers who have a particular specialty, or providing a broad range of functions. Early engagement of stakeholders allows you to understand the constraints of the task, manage expectations, negotiate priorities, and make tradeoffs [5].

1.2 Problem and general hypothesis statement

Design decisions are relevant statements in software systems that impact the determination of the rationale of the software architecture, allowing capturing its functional and quality requirements [41]. Moreover, architectural design decisions play a significant role in the design, development, integration, evolution, and reuse of software architectures [25]. Capturing design decisions and rationale has several advantages [43], and the acquisition and record of this information is an interesting research topic.

In [31], they comprehended that to use architectural tactics, it is necessary to create a condition that needs to be evaluated when it arrives to the system. These conditions are obtained by the business analysis done by the stakeholders. So, we have realized that architectural tactics can be used to offer design decisions alternatives to the software architect, but depend on NFRs that stakeholders provide. These alternatives allow to software architects to compose more informed decisions to satisfy stakeholder's needs.

Nevertheless, incorrect designs related to security can provoke architecture weakness [13]. Software architecture design is the first and the fundamental step to address quality goals surrounding attributes such as security, privacy, safety, reliability, dependability, and performance. According to [33], estimations indicate that roughly 50% of security problems are the result of software design flaws such as miss-understanding architecturally important requirements, poor architectural implementation, violation of design principles in the source code and degradation of the security architecture, leading us to establish the main problem:

Main research problem

Bad architecture design decisions in software systems can have several impacts on various security concerns, giving malicious actors more opportunities and flexibility to execute a cyber attacks.

The problem described is one of the main challenges that a software architect must face is to consider the most appropriate secure design decisions. Therefore, we believe that a collaborative technique that can involve stakeholders (with different profiles and with or without knowledge of security tactics) in decision-making, can help to better satisfy security requirements. In this way, the software architect will have additional information provided by stakeholders, to select the final decision regarding security concerns.

What we described above allows us to specify what we want to achieve in the research and determine the intermediate results require to direct the conclusions of our experiment; for this, we established the following general hypothesis and corresponding sub-hypothesis of this thesis:

General hypothesis

The interaction between stakeholders produces more pragmatic architectural decision-making concerning the selection of security tactics in a given scenario.

Derived from the general hypothesis, we develop two sub-hypotheses that we address through an experimental study to validate the proposed technique. The first sub-hypothesis is related to the analysis of individual decisions versus group decisions using the proposed technique. The second attempt to compare the results of using the technique versus not using it.

Sub - hypothesis 1

Consensual architectural decision-making through the interaction between stakeholders generates better decisions alternatives than the decision taken by a stakeholder individually.

Sub - hypothesis 2

A decision-making technique based on the interaction between stakeholders generates more practical decisions instead of an ad-hoc procedure.

1.3 Proposed solution: TaSPer

TaSPer

This thesis proposes Tactic Selection Poker (TaSPer), a technique that extends the Planning Poker to allow a development team to agree on what architectural tactics should be used to build a secure architecture.

For this, the proposal adopts and combines the concepts related to architectural security tactics, thus presenting an original technique in the field of design decisions in the area of software engineering.

More specific, we design TaSPer to select security tactics to satisfy security requirements and later, in a future work could expand the technique to others quality attributes.

1.4 Research objectives

Main research objective

The main objective of this thesis is to design and validate a technique to select software architecture tactics by consensus.

The above, with the premise that stakeholders achieve better resolutions regarding the project, a greater understanding, and commitment to what must be developed, thus making better implementation and robust design decisions from the project beginning and during its evolution, helping to acquire a better understanding of architectural tactics. Therefore, a series of activities were carried out, which consisted of a pilot, two pre-experiments, and an experiment focused on the quality of the safety attributes to assess TaSPer. In order to achieve this, four specific objectives were established:

- Explore the state of the art regarding security design decisions.
- Design a consensus-based technique to select software architecture tactics.
- Evaluate and verify the technique by conducting an empirical study.
- Validate the proposed technique through the steps of the experimental process.

1.5 Research questions

Main research question

How does the interaction between people in the selection of architectural security tactics on specific security scenarios impact?

The research question was established to know and understand the effects of making design decisions by consensus based on security tactics in the area of software architecture. Thus, it is essential to verify the interaction, exchange, and points of view of the different stakeholders in the moment of decision-making. With this, establish the final result based on an agreement by consensus between all the participants.

For this, a case study was developed in order to evaluate and verify the proposed technique, defining 4 research questions. The questions are related to knowing the operation and process of TaSPer and thus obtaining first impressions regarding the technique, allowing to know the existing strengths and weaknesses of the process and continue its development.

- RQ1: Can the TaSPer technique be used for design decisions by consensus?
- RQ2: Can the prioritized selection of tactics allow determining the importance of the selected tactic?
- RQ3: Does the design of the TaSPer technique allow the interaction of the different participants?
- RQ4: Can TaSPer allow the collection of valuable data to be used by software architects?

Finally, we establish a experimental study to validate the TaSPer technique through the experimental process considering the stages: definition, planning, evaluating, execution and analysis. Thus, the following research questions was established:

RQ5:

Does TaSPer improve individual design decisions?

RQ6:

Having an established technique allows us to obtain results closer to the ground truth than not having it?

1.6 Research contributions

The main contributions of this master thesis are described below:

- Develop and establish a technique that allows the selection of software architecture security tactics through decision-making by consensus allowing the different project stakeholders' participation.
- Describe the experimental process carried out with TaSPer technique, defining each stage in a way that allows its replication and generating the opportunity to continue with its development.
- Establish a set of security tactics cards in software architecture, considering the most relevant information known about a tactic and the ability to integrate the most prominent information from a card.

1.7 Published work

Based on the work performed in this Master thesis, the following articles have been published:

Articles directly related to this thesis:

- Felipe Osses, Gastón Márquez, Mónica Villegas, Cristian Orellana, Marcello Visconti, Hernán Astudillo: Security tactics selection poker (TaSPer): a card game select security tactics to satisfy security requirements. ECSA 1.4 (Companion) 2018: 54:1-54:7. Madrid, Spain.

Figure 1.4: ECSA 2018

- Felipe Osses, Gastón Márquez, Cristian Orellana, Hernán Astudillo: Towards the selection of security tactics based on non-functional requirements: Security tactics planning poker. Security Tactic Planning Poker. SCCC 2017: 1-8 Arica, Chile.

Articles developed about Patterns and Tactics of Architecture:

- Gastón Márquez, Felipe Osses, Hernán Astudillo: Review of Architectural Patterns and Tactics for Microservices in Academic and Industrial Literature. IEEE Latin America Transactions, 16(9), 2321-2327, 2017.

- Felipe Osses, Gastón Márquez, Hernán Astudillo: Exploration of academic and industrial evidence about architectural tactics and patterns in microservices. ICSE (Companion Volume) 2018: 256-257. Gothenburg, Sweden.

- Felipe Osses, Gastón Márquez, Hernán Astudillo:: An Exploratory Study of Academic Architectural Tactics and Patterns in Microservices: A systematic literature review. CIbSE 2018. Bogota, Colombia.

Articles related to this master thesis:

- Juan Brito, Felipe Beroiza, Gastón Márquez, Marcello Visconti, Hernán Astudillo: Evaluating Impact of Experience in Architectural Design Decision-Making Techniques: An Experimental Study, SCCC 2019, Concepción, Chile

1.8 Structure

This thesis is structured by six Chapters as follows: Chapter 1 aims to present the background related to this thesis, give the problem statement, the research question, the main objectives, and its contribution. Chapter 2 describes the state of the art. Chapter 3 describes the proposed technique, introducing TaSPer and its operation. Chapter 4 presents a case study carried out to evaluate and verify TaSPer. Chapter 5 describes the experimental process considering the following stages: definition, planning, evaluating, execution, and analysis. Finally, Chapter 6 summarizes the research done in this thesis, mentions the main findings, and describes future work. To better understand the structure of the thesis, we developed a diagram (see figure 1.5) considering the different stages that were established to carry out this work, as well as emphasizing the proposal and its relationship with TaSPer and emphasizing the importance of the experimental study for this thesis.

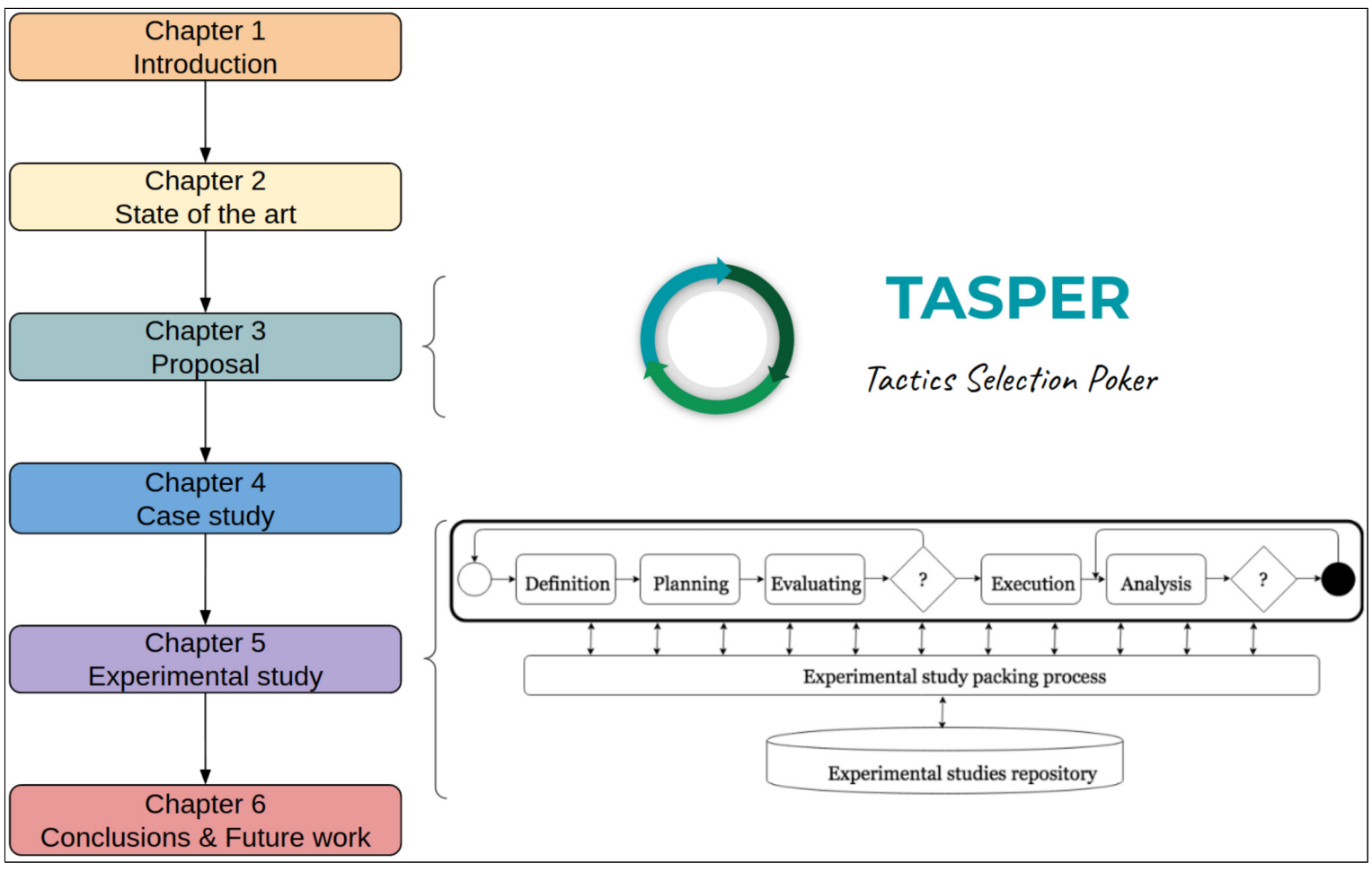


Figure 1.5: Thesis structure

Chapter 2

State of the Art

This chapter aims to describe the state of the art based on software architecture domains related to non-functional requirements, design decisions, software architecture tactics, and consensual decisions, considering as main focus the search for a technique that allows the selection of architectural tactics in a consensual way. Thus section 2.1 focuses on techniques used in selecting architectural tactics and section 2.2 describes techniques related to security concepts and consensual decision making within the field of software engineering. Finally, section 2.3 summarize the state of the art.

2.1 Selecting architectural tactics

Chavarriaga et al. [9] present a technique for selecting architectural tactics in the Cloud context. The scope of this technique considers vital aspects to support the appropriate selection of tactics for a specific scenario as it models tactics through architectural constructions based on feature models, models design alternatives and configuration options, relates architectural tactics with the design alternatives through the use of feature-solution graphs, and finally, allows the identification of possible trade-offs generated by conflicts in the design alternatives through the use of a framework called FAMA [7]. In a later work, Chavarriaga et al. [11] detail its proposal in-depth, presenting an approach for detecting trade-offs between tactics called FaMoSA. This framework allows obtaining a higher precision in the selection of tactics for a specific system from the early stages of design through three ways: (1) The specification of architectural tactics and the relationships between them; (2) the specification of possible configurations of architectural tactics identifying the existence of conflict between them; and (3) the specification of trade-offs decisions based on specific formats.

Kim et al. [28] presented an approach for systematically embodying NFRs into software architecture using architectural tactics. The feature modeling of tactics provides maneuverability for configuring tactics for a given set of NFRs. The RBML

specifications of tactics with the rigorous composition rules facilitate the mechanical composition of tactics. Related to this, Kim et al. [27] define tactics as reusable architectural building blocks that provide generic solutions to address quality attributes. This definition implies a plug-in or template approach, contrary to the idea of patterns, which include several sections indicating the conditions to apply the pattern and its consequences [16]. Kim [26] additionally presents an article that uses a quantitative approach to choose tactics based on the weighting of multiple factors of complexity and effort whose result allows the derivation of appropriate tactics for a specific problem.

Pedraza-Garcia et al. [38] proposed a workflow approach to support architects in selecting architectural security tactics for system design. They present a methodological approach to address and specify the quality attribute of security in architecture design, applying security tactics. They describe and use a set of activities and tools to implement a set of security tactics in designing a software system. The applicability of the tactics of security was achieved with the case study: the tsunami warning-prevention system.

Koziolek et al. [29] propose an approach guided by architectural tactics that use evolutionary optimization algorithms to model an architecture problem and identify trade-offs to subsequently derive candidate design decisions automatically based on constraints of the interactions between quality attributes and the design space.

Alashqar et al. [1] present a framework for selecting the most appropriate architectural tactics according to their best achievement of the required levels of quality attributes when developing transaction processing systems is proposed. The framework is base on fuzzy measures using the Choquet Integral approach. It considers the impact of architectural tactics on quality attributes, the preferences of quality attributes, and their interactions. It can also be used to compare different potential architectures in terms of their support of quality attributes. The abilities and the advantages of the proposed framework are clarified via practical experiments utilizing a case study.

2.2 Secure software engineering using gamification

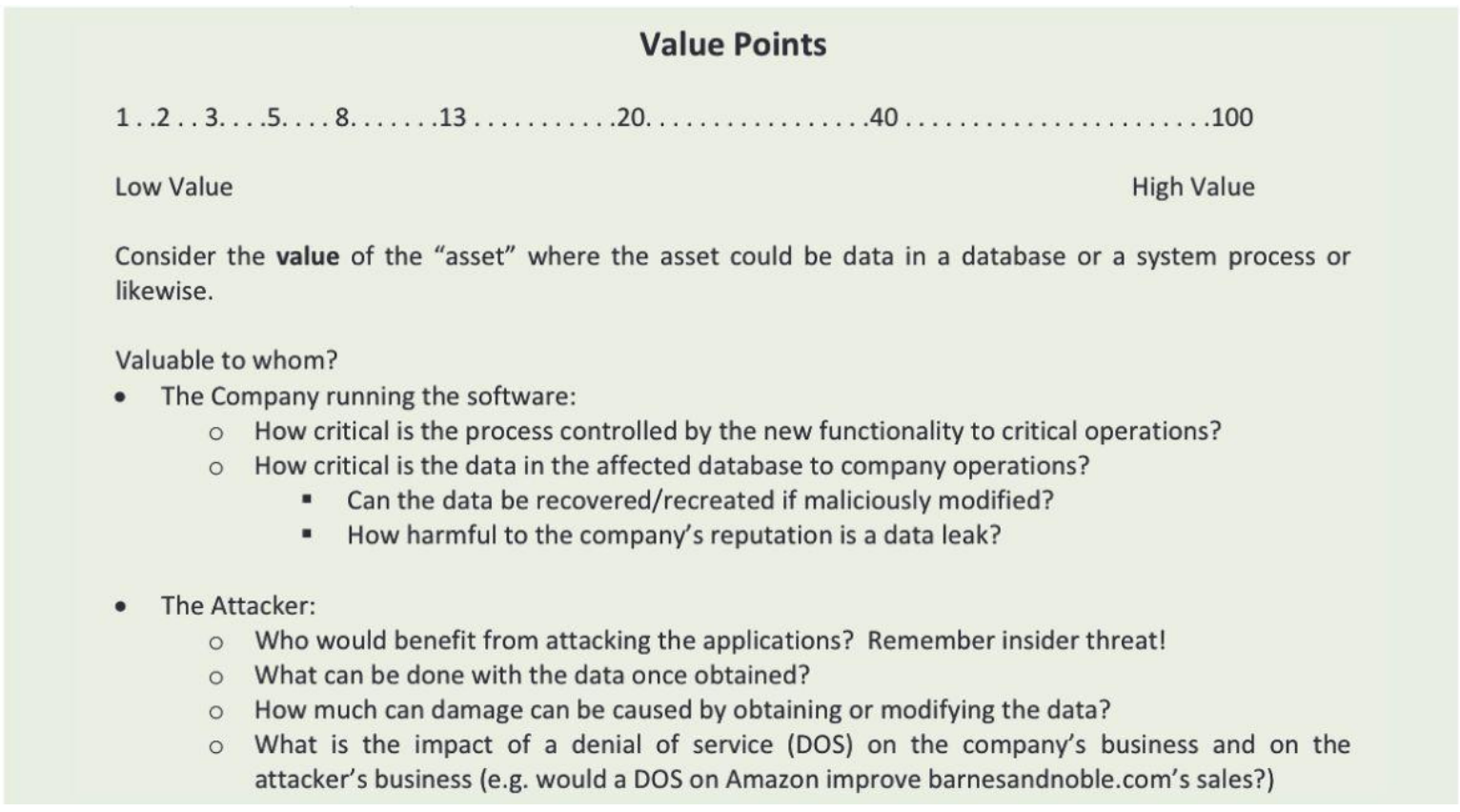
During the research, it was also possible to determine a significant relationship between security concepts and consensus decision-making under the idea of gamification within the scope of software engineering. The works presented below allowed us to determine that there are no works related to gamification focus on selecting security tactics, what is summarized in the table 2.1.

Williams et al. [45] proposes a technique called *Protection Poker*, which is a software security game where its result is a list of the security risk related to each requirement. The team can use this relative risk to determine the type and intensity of design and validation and verification effort the development team must include in the iteration for each requirement. The unit can then use this list to prioritize secu-

Table 2.1: Consensus, security and gamification approach's

	Security	Gamification	Cards	Game	Teaching/Learning	Design Decision	Tactics
Protection Poker	x	x				x	
Control-Alt-Hack	x	x	x	x	x		
d0x3d!	x	x	x	x	x		
Smells Phishy?	x	x	x	x	x		
Cyber Realm	x	x	x	x	x		
Smart Decisions Game	x	x	x	x	x	x	
OWASP Cornucopia	x	x	x	x	x		
TaSPer	x	x	x		x	x	x

rity engineering resources toward software areas with the highest risk of attack based on the ease of attack of the new functionality and the value of the data to which it is accessed through functionality. This enables the team to calculate necessary to securely implement the requirement and plan what resources are required for a secure implementation proactively. With more excellent knowledge, this prioritization should allow the team to develop more secure software, reducing its vulnerabilities. The four significant benefits of using this technique would be quantifying the software security risk, which a team can use to classify the requirements, adapt the criteria, and thereby reduce the security risk proactively, fortify security to minimize security risk and the exchange of software security knowledge among participants.



Value Points

1 . .2 . . 3. . . .5. . . . 8. . . . . .13 . . . . . .20. . . . . .40 100

Low Value High Value

Consider the **value** of the "asset" where the asset could be data in a database or a system process or likewise.

Valuable to whom?

- The Company running the software:
  - How critical is the process controlled by the new functionality to critical operations?
  - How critical is the data in the affected database to company operations?
    - Can the data be recovered/recreated if maliciously modified?
    - How harmful to the company's reputation is a data leak?
- The Attacker:
  - Who would benefit from attacking the applications? Remember insider threat!
  - What can be done with the data once obtained?
  - How much can damage can be caused by obtaining or modifying the data?
  - What is the impact of a denial of service (DOS) on the company's business and on the attacker's business (e.g. would a DOS on Amazon improve barnesandnoble.com's sales?)

Figure 2.1: Sample "cheat sheet" of security issues, Protection Poker

Denning et al. [14] create *"Control-Alt-Hack: White Hat Hacking for Fun and Profit"*: a recreational, tabletop card game about computer security. The authors' goal has generated awareness of security issues and improved people's perception

of computer security as a discipline and career choice. They traded some technical complexity in the topics discussed in exchange for increased engagement: put another way, they set out to create a game that players could find inherently fun. They might learn incidentally in the course of enjoying the gameplay.

Figure 2.2: Sample of Control-Alt Hack game cards

Following the same idea, Gondree et al. present *[d0x3d!]* [20] [21], a collaborative game as white-hat hackers, where they describe some opportunities for exposing young audiences to cybersecurity via informal lessons, leveraging play for education and outreach. The authors reveal the experience in using the proposal of [14], where they conclude that games inspire players to challenge the limits of play by exploring the meaning and interpretation of rules. Similarly, rule testing, rule interpretation, and rule-breaking are prerequisite Red Team skills. They argue that such adversarial thinking is foundational to both strategic games and security engineering. Also, the game provides an artificial context for discussing real ideas in network security. When designing it, we made sure to introduce and use appropriate security terminology, for example, "administrators," "intrusion detection," "compromise," "patch," "0-day," and "forensics"—in ways consistent with their real-world interpretations. They must infiltrate and navigate an adversarial network, retrieve a set of valuable digital assets, and escape. The adversary is encoded in the game's mechanics, as the system

Figure 2.3: d0x3d! security game

periodically adjusts its state, either patching or decommissioning servers for forensic investigation.

On the other side, Beckers et al. [6] propose to use a card game to elicit security requirements, which all employees of a company can play to understand the threat and document security requirements. The game considers the particular context of a company and presents underlying principles of human behavior that social engineers exploit and concrete attack patterns. The authors evaluate their approach with several researchers, IT administrators, and professionals from the industry. Another alternative is described by Baslyman et al. [3]. The authors propose *Smells Phishy?*, a board game that contributes to increasing users' awareness of online phishing scams, a proposal that was tested with 21 participants. The results showed that after playing the game, participants understood phishing scams and learned how to protect themselves better. Regarding the participants, they pointed out that it was a pleasant, fun, and exciting activity. According to the authors, the game increased knowledge and awareness and encouraged discussion.

Figure 2.4: Sample of Smells phishy game

Another proposal found was the *Cyber Realm* card game by Nestler [34][1], who presents a card game where players learn the ten principles of cybersecurity, by recognizing hand gestures, playing with others, or playing alone. The card decks are divided into three: Principles deck, information array y question cards. The game's development is under the initiative of the GenCyber program[2], which has the mission to increase the number of students studying cybersecurity in the United States. The program includes summer camps across the nation designed for elementary, middle, and high school students and teachers. The camps focus on engaging the learners with sound cybersecurity principles and teaching techniques.

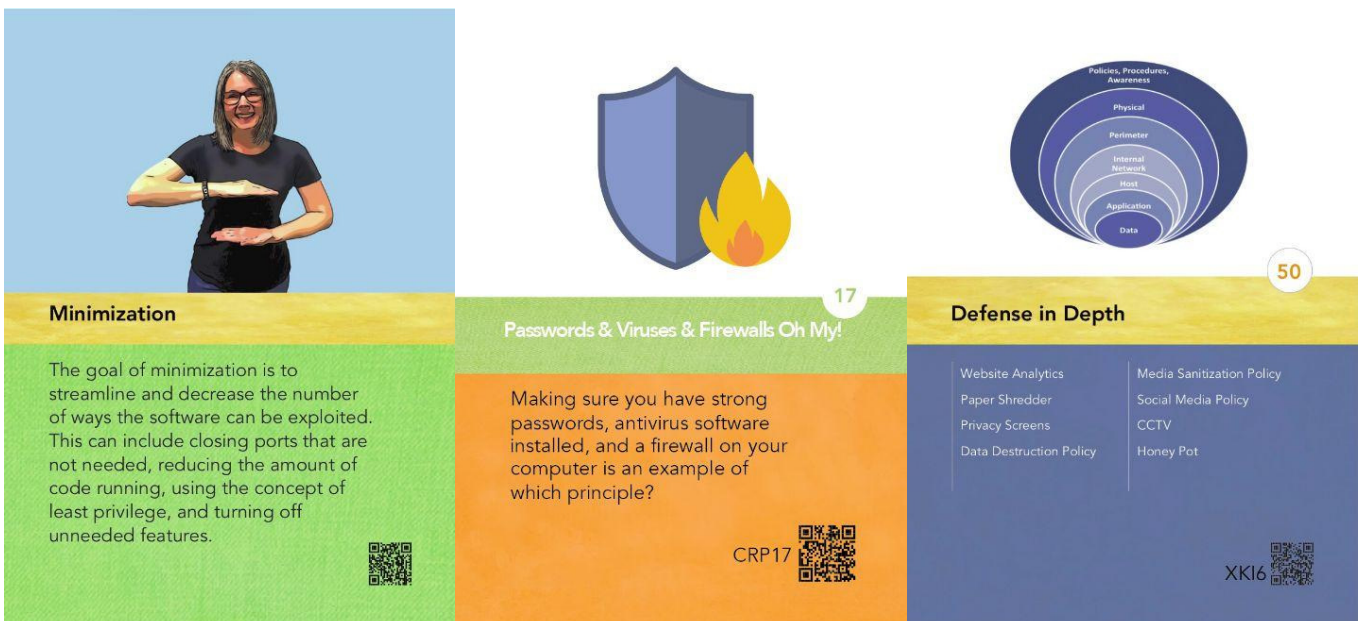


Figure 2.5: Sample of Cyber Realm card game

It was also possible to find a technique called *Smart Decision Game* (cards games for decision making) Cervantes et al. [8], a game that seeks to teach the use of

[1]https://gencybercards.com/
[2]https://gen-cyber.com/

architecture and technology patterns in a fun and short time using cards, dice, points and a scoreboard. The game is based on Attribute-Driven Design (ADD), a method from the Software Engineering Institute that allows complex architectures to be designed predictably and efficiently. Smart Decisions can be played by software architects, software and embedded engineers, data scientists, students, and anyone interested in learning about software architecture design techniques and how they can be applied in the design of software systems with modern technologies. The game's benefits are: Boost architectural skills, Learn Big Data, learn Machine Learning or Internet of Things technologies, build a team, and gamify the work.

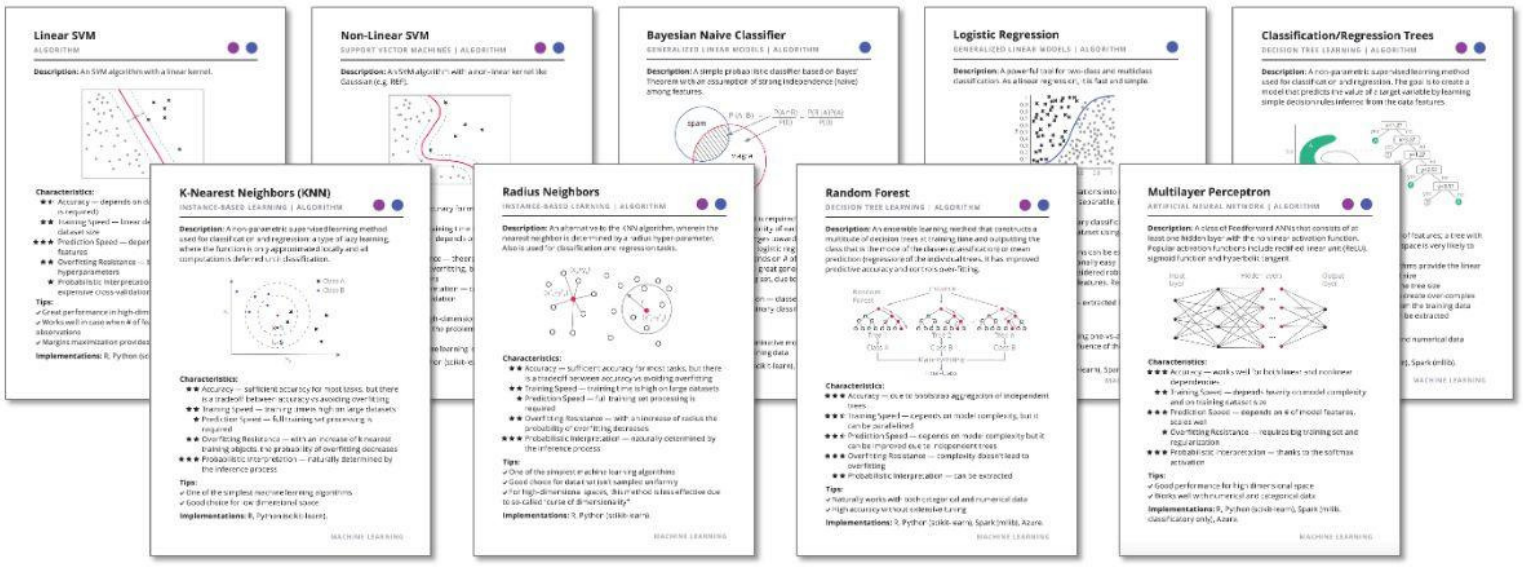


Figure 2.6: Sample of Smart Decisions cards

Finally, Thompson et al. [40] evaluate effectiveness of OWASP Cornucopia, a card game which is designed to assist software development teams, identify security requirements in agile, conventional and formal development processes. They performed an experiment where sections of graduate students and undergraduate students in a security related course at University of North Texas were split into two groups, one of which played the Cornucopia card game, and one of which did not. Quizzes were administered both before and after the activity, and a survey was taken to measure student attitudes toward the exercise. The results show that while students found the activity useful and would like to see this activity and more similar exercises integrated into the classroom, the game was not easy to understand.

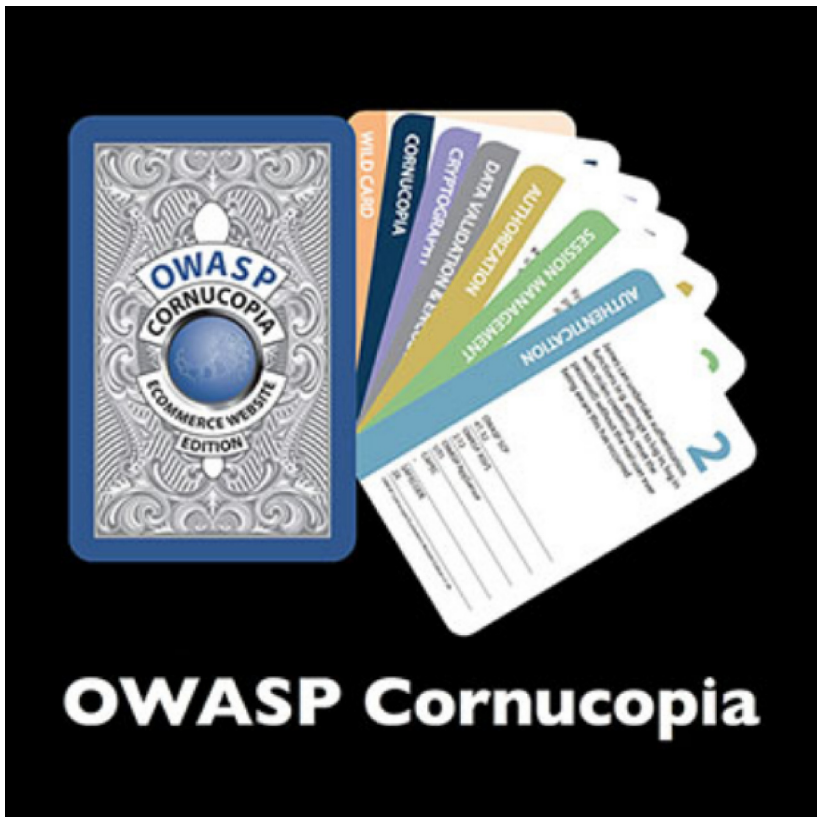


Figure 2.7: Sample of Cornucopias cards

2.3 Summary

The analysis of state of the art, whether focused on the decision-making process on architecture tactics or the strong relationship between consensual decisions, security, and the concept of gamification, allows us to visualize the importance of decision-making in the field of Software Engineering.

The articles cited in this section stand out for their originality when addressing such a complex issue as tactics or design decisions related to security. However, none of them deploys card game techniques to support the selection of architectural security tactics considering the project's stakeholders' points of view. Finally, the evidence shows insufficient information about proposals that use security tactics to address security requirements and support decision-making.

Chapter 3

Proposal

This chapter aims to describe the proposal technique Tactics Selection Poker, TaSPer. Section 3.1 is the introduction of the proposal chapter; Section 3.2 describes the technique and its objectives, the creation of the cards, the setting parameters, and how to use TaSPer. Finally, Section 3.3 summarizes the general structure of the thesis.

3.1 Introduction

The thesis's central idea was related to the work of the "Toeska"3.1 Software Engineering Group, under the Department of Computer Science of the Technical University Federico Santa María, who has worked with an emphasis on software architecture tactics [35] [31] [32] [17], [19]. It was possible to establish the need for a technique that allows guiding the selection of architectural tactics in a group and consensual manner, allowing us to obtain the stakeholders' different points of view, thus establishing the bases that will allow the development of the desired architecture.

Figure 3.1: Toeska Software Engineering Group

3.2 TaSPer technique

The TaSPer technique is based on Planning Poker [22] [12], a well-known card-based estimation technique developed for planning and effort estimation of the agile development field; However, instead of sharing effort estimations, participants share preferences among architectural tactics. TaSPer's main idea is to use Planning Poker as a mirror technique, seeking to extrapolate and replicate most of its possible characteristics and qualities and share its philosophy.

This decision was taken considering the most important attributes presented by the planning poker technique, which are:

- Achieve the active participation of those involved.
- It manages to establish trust among the participants.
- It allows the transfer of information between the participants.
- It manages to level and balance the knowledge of the people who participate thanks to the argumentation process regarding the estimate made.

On the other hand, the planning poker steps were essential for the creation and development of TaSPer. The steps described in the below list illustrate the procedure for using Planning Poker to estimate based on consensus.

1. To start a poker planning session, the product owner or customer reads an agile user story or describes a feature to the estimators.
2. Each estimator is holding a deck of Planning Poker cards with values like 0, 1, 2, 3, 5, 8, 13, 20, 40, and 100, which is the sequence recommended.
3. The estimators discuss the feature, asking questions of the product owner as needed. When the feature has been fully discussed, each estimator privately selects one card to represent their estimate. All cards are then revealed at the same time.
4. If all estimators selected the same value, that becomes the estimate. If not, the estimators discuss their estimates. The high and low estimators should especially share their reasons. After further discussion, each estimator re-selects an estimate card, and all cards are again revealed simultaneously.
5. The planning poker process is repeated until consensus is achieved or until the estimators decide that agile estimating and planning of a particular item needs to be deferred until additional information can be acquired.

3.2.1 Main objective

TaSPer is a technique that supports architectural tactics selection and design decisions from the project beginning and during its evolution. Its purpose is to help participants acquire a better understanding of tactics. It was designed to carry out collaborative work among the participants at the time of decision making, where each of them must prioritize tactics from a standard set of architectural tactics (shown in Figure 1.3) to satisfy the project goals, and shares them with other participants to converge on the most appropriate tactics. Within the analysis made to Planning Poker, some characteristics unavoidable to the technique were established, which had to be replicated in the process of creating TaSPer:

- Use of cards for decision making (estimation)
- Combination of opinions by all stakeholders
- Inclusion and participation of all involved
- Must have a moderator (can be any subject)
- There must be a discussion process, in which decisions can be argued and thereby achieve convergence in a collaborative decision.

Considering the characteristics mentioned above, it was established that the TaSPer technique should have five standard parameters that allow us to know how the technique works, and thereby guide those who use it. For this, the parameters determine the context of the project to be used and scenarios to be evaluated, and specific essential guidelines for the correct operation of TaSPer.

3.2.2 Card creation

One of the essential characteristics in Planning Poker is the use of cards that allow estimation to be carried out within the Gamification concept (Chapter 2). Because of this, for TaSPer, a design was sought that would allow decisions to be made regarding a specific condition, and with it would allow generating a consequence in the software architecture design. Therefore, we establish that the cards should represent each of the tactics since each one affects the architecture.

The cards were designed considering the different attributes of the architecture tactics [5] and looking to contribute with the use of the Cards to a greater dynamism, easy learning, and excellent usability for those interested.

Initial design

In order to adapt Planning Poker to TaSPer, the initial design of the proposed cards was as follows (See Figure 3.2):.

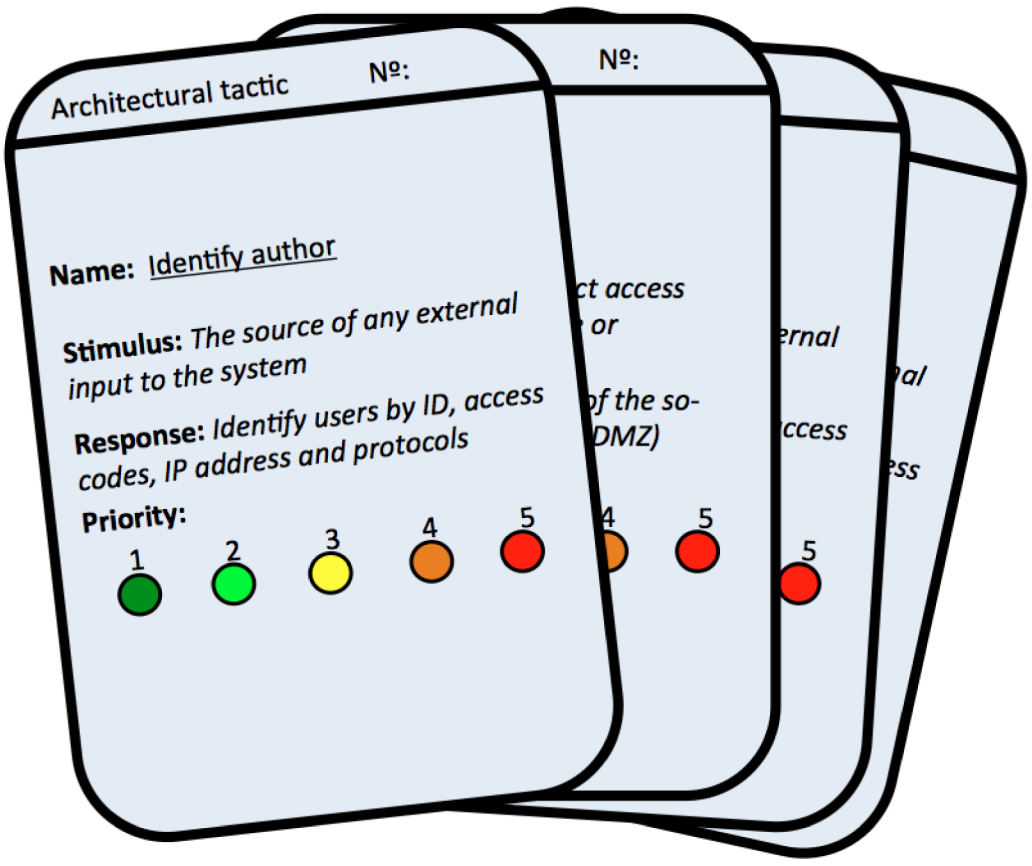


Figure 3.2: Initial design of the TaSPer cards

The initial design of the TaSPer cards contains the following fields:

- *Number:* Describes the number of security tactics.
- *Name:* Describes the name of the security tactic.
- *Stimulus:* Illustrates the incentive to use the corresponding security tactic.
- *Response:* Describes the response associated with the corresponding security tactic.
- *Priority:* Estimates the priority, according to the stakeholder, of the importance of the tactic to satisfy the NFR. The priority range corresponds to 1: Very low, 2: Low, 3: medium, 4: High, and 5: Very high.

Final Design

The final design differs from the initial one in some key aspects based on the experience obtained from both the pilot evaluation and the pre-experiment (see chapter, allowing its evolution and the necessary modifications to be used in the experimentation process. In addition to the aesthetic change that the cards had, the most crucial change is that the new cards do not have the possibility of choosing the stakeholders' priority. Below is the final design of the cards used.

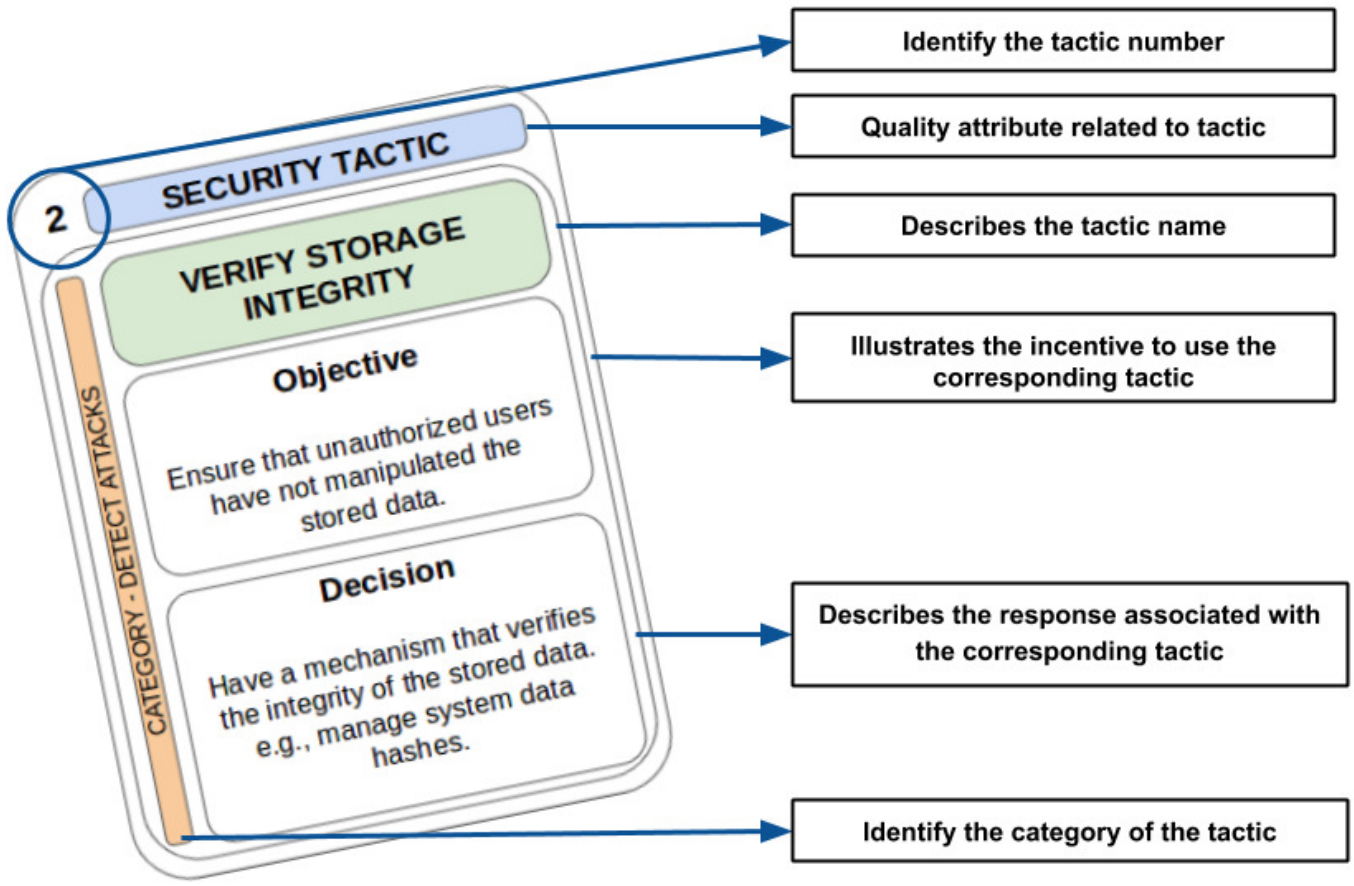


Figure 3.3: Final design - TaSPer Card

The final design of the TaSPer cards contains the following fields:

- *Number*: the tactic id.
- *Quality attribute*: the quality attribute related to the tactic (in this case, only security).
- *Name*: the tactic long name.
- *Objective*: the goal pursued with this tactic.
- *Decision*: the response associated with this tactic.
- *Category*: the taxonomy category of the tactic.

3.2.3 Setting parameters

Before using the technique, it is necessary to set the base parameters to execute it (see figure 3.4):

- **Project selection:** Proposal, task, development, or requirement in the field of software architecture that must be developed by a group of people. The project represents the general objective of the work to be carried out, for which it must be correctly defined.

Figure 3.4: Parameters configuration sequence for TaSPer

- **Scenario definition:** Context where the project is developed, being necessary to formally define the parts of the project that will be discussed, which could be carried out through use cases non-functional requirements (NFR).

- **Assume as moderator:** The technique requires that there be a person in charge of moderation, and it can be anyone involved and related to decision-making.

- **Subjects selection:** It is necessary that during the use of the technique, all those involved in the project or development to be carried out participate, this to obtain a more significant commitment of the decisions made from the beginning, and at the same time the exchange of opinions between different stakeholders.

- **Tactic selection:** The type of tactics that will be used should be selected based on the quality attributes that want to be evaluated. These will be presented through cards, as explained in the previous point 3.2.2.

3.2.4 TaSPer steps

An essential point in the development of the technique was establishing the sequence to be followed in choosing tactics framed in the development of Planning Poker. In this sense, we determine that TaSPer should have three steps that must be carried out in correlative order: ***Discussion, Choice of tactics, and Consensus***. These three steps aim to produce a workspace that allows interaction between the participants; then, they can finally select tactics to present the result. It is important to note that TaSPer represents a continuous cycle of three steps until reaching consensus in decision-making (see ref fig: steps). The TaSPer steps were refined in meetings with local experts in agility and architecture, which also yield new ideas included in the technique final version, being these:

1. **Discussion:** Allows achieving the necessary workspace for all participants.

   (a) One of the subjects assumes as *the moderator* and takes control of the meeting.

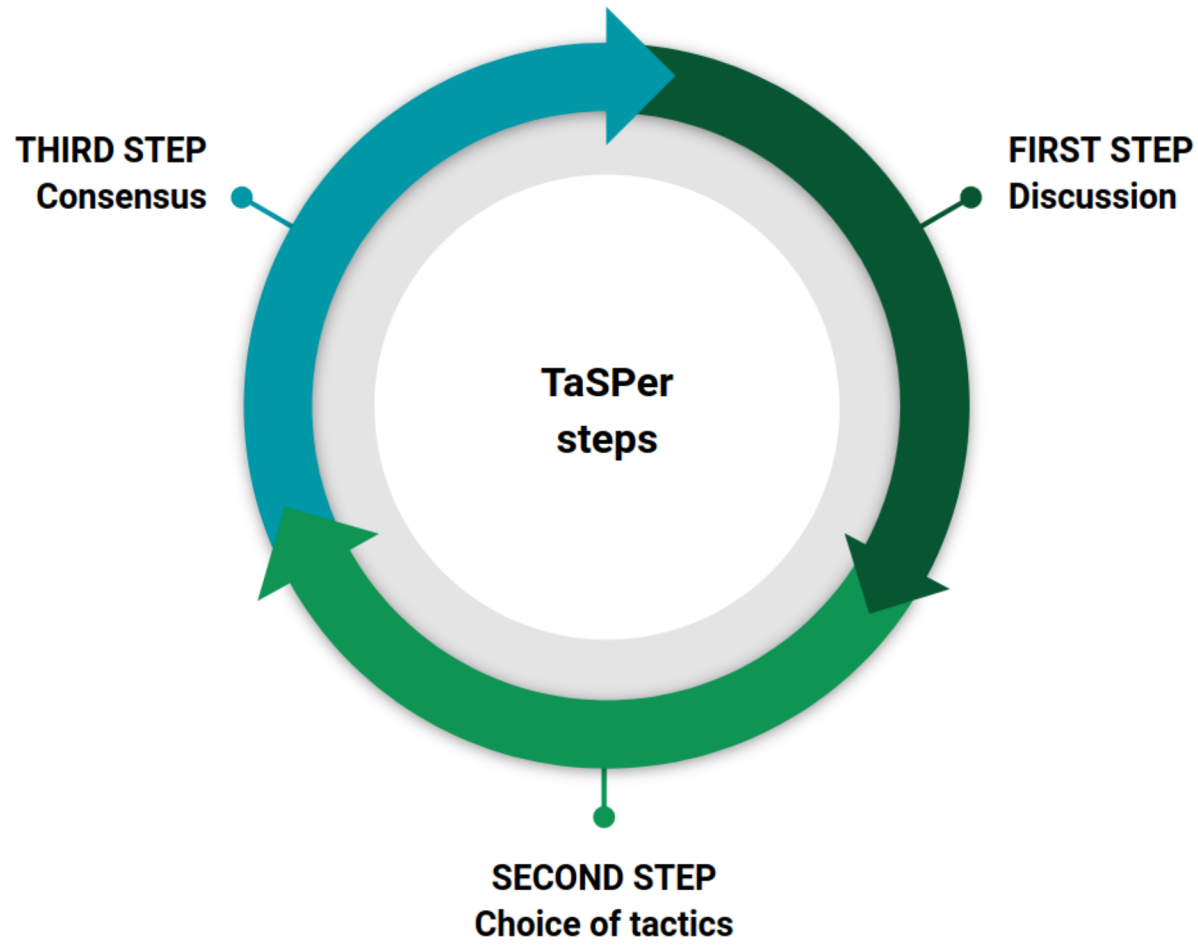


Figure 3.5: TaSPer cycle steps

(b) The moderator is the guide to achieve the consensual decision of security tactics.

(c) The first moderator task is to present the context and scenario to be discussed, and distribute the security tactics cards.

(d) Each participant receives a deck of security tactics cards.

(e) The context, goals, and scenario are discussed among all participants.

2. **Choice of tactics:** Participants, according to the presented scenarios, must choose the appropriate architecture tactics

   (a) Each participant privately selects the most appropriate card(s) for the situation and goals

   (b) The cards and their choices are revealed by all participants

3. **Consensus:** Participants must agree on what are the tactics that must be selected for a particular scenario, having as a final result, the consensus of the chosen tactics

   (a) Each participant argues their choice

   (b) The moderator records the rationales

(c) If all the participant selected one or more security tactics, they became the selected tactics

(d) If there was no consensus, participants can discuss immediately and try to make a new common choice on the spot; if this is difficult, the moderator can start everything again from the step **"1.b"**.

3.3 Summary

The proposed technique, TaSPer, was designed considering Planning Poker as a base, having as its primary objective the selection of software architecture tactics in a consensual way by the stakeholders participating in a project. Considering what has been described, rules were established that allow determining the main parameters and the steps that must be taken to use TaSPer, the latter being: Discussion, Choice of tactics, and Consensus.

Chapter 4

Case study: Innovation Management Project

The purpose of this chapter is to evaluate and verify the TaSPer technique considering its improvements and finally allow the TaSPer technique to mature and thus achieve its final design. For this we present the case study "Innovation Management Project" to a group of nine Chilean Navy professionals with IT security experience who are not familiar with architectural tactics.

First, in section 4.1, we describe the case study with its respective NFR. Then, in section 4.2 we describes the Design and planning for the case of study with the objectives, the research questions and the parameter of the study. Section 4.3 describes the preparation and collection of the data. Section 4.4 shows the data analysis made of the results; and finally Section 4.5 is the summary of the chapter.

4.1 Case study: Innovation Management Project

At the Federico Santa María Technical University (UTFSM from now on), the life cycle of a project has critical stages involving various actors, whether internal or external. These stages are: (1) the creation of the initiative or challenge, whether of the company or academic, (2) the development of new ideas or the use of a portfolio of initiatives that could solve this problem, (3) the preliminary draft or feasibility study, (4) the investigation and execution of the chosen solution and, finally, the possible patents, licenses and spin-offs, which could result in finalizing this project.

However, the current situation does not allow agile communication to coordinate support for potential or developing projects in different aspects. In the UTFSM there are systems of financial assistance for projects, management, application of patents, and various consultancies. But, these systems are limited to the use of specific units, and they use own systems that hinder effective coordination. The above also restricts the access to the information to the stakeholders on some of the potential projects, in development or finalized stages. The same situation is manifested when somebody

wants to access records of entrepreneurship initiatives that emerged as the product of some result or project.

Given the above description, we are currently working on the proposal of architecture with the aim of showing all the potential work of the UTFSM to the community. Because the objectives of the project require that there is a communication between the UTFSM and external systems, there are security requirements that must be satisfied as a requirement of the project. For the above and this comparative study, we select the three most important security NFRs to know what security tactics can help us to make better design decisions in the architecture in which we are currently working. Next, the three security NFRs descriptions are detailed:

- **NFR1:** The platform will use a publish/subscribe architecture, where the messages to be posted must be transformed into the desired format and addressed to one or more subscribers to communicate the initiatives of the UTFSM. For this, this communication should be based on the following aspects: confidentiality of information, integrity, authentication and access management under UTFSM standards.

- **NFR2:** The platform will have an SOA architecture that will communicate with the services offered by UTFSM systems through web services. This communication is required to use security mechanisms, such as WS-Security.

- **NFR3:** The platform must guarantee mechanisms to promote confidentiality and integrity in communications with internal systems of the UTFSM and with external platforms.

4.2 Design and Planning

Research objective

For this case study, the objective is related to knowing the operation and process of TaSPer and thus obtaining first impressions regarding the technique. This will allow to know the existing strengths and weaknesses of the process and continue its development. More precisely we want to evaluate 3 NFRs related to Security Tactics and the decision made by the different subjects.

Research questions

Based on the objective of the case study, we established the following research questions:

- **RQ1:** Can the TaSPer technique be used for design decisions by consensus? This RQ tries to determine if the design developed for TaSPer allows consensus

decision-making, considering the selection that each participant will establish depending on the NFR under discussion.

- **RQ2:** Can the prioritized selection of tactics allow determining the importance of the selected tactic? This research question seeks to assess the importance that the selection of each tactic can be assigned a priority, considering a scale of 1 to 5.

- **RQ3:** Does the design of the TaSPer technique allow the interaction of the different participants? This research question seeks to determine that TaSPer allows all those involved in decision-making to express their opinions and get involved, thus generating an enriching dialogue.

- **RQ4:** Can TaSPer allow the collection of valuable data to be used by software architects? This research question aims to assess the importance of giving architects a broader vision of the project and thus be able to make decisions with more information.

Parameters of the study

To carry out the pilot evaluation, the following base parameters were considered:

- **Project selection:** The design of the activity was based on a case study of an academic innovation project (described later) in which it was have analyzed the most relevant security requirements. This study includes addressing the most important NFRs for stakeholders.

- **Scenario definition:** Three security NFRs were used based on an academic project.

- **Assume as Moderator:** The moderator is in charge of guiding each phase of the activity, assuming one of the subjects in the room.

- **Subjects Selection:** This pilot evaluation was presented to a group of nine professionals dedicated to IT security, who were not familiar with architectural tactics.The participants are part of the computer security incident response team (CSIRT) of the Chilean Navy.

- **Tactics selection:** Regarding the tactics used, in Tables 4.1 and 4.2 we will describe the 17 TaSPer cards. For each security tactic, we have summarized them using the fields described in both tables for easy understanding of stakeholders.

Table 4.1: TaSPer cards list (Part I)

	Name	*Identifying author*
C1	**Stimulus**	The source of any external input to the system
	Response	Identify users by ID, access codes, IP address and protocols
	Name	*Detect intrusion*
C2	**Stimulus**	Identify malicious behavior stored in systems
	Response	Detect Malicious Behavior in Protocols, Applications, and others
	Name	*Detect Denial of Service*
C3	**Stimulus**	Denial of service
	Response	Check the configuration of routers and Firewalls to stop invalid IPs
	Name	*Verify Message Integrity*
C4	**Stimulus**	Procedures whose purpose is to alter the integrity of the data of a message
	Response	Procedure to ensure verification of data integrity
	Name	*Detect delay in message*
C5	**Stimulus**	Detect a potential attack by man-in-the-middle
	Response	Detect suspicious behavior
	Name	*Authenticate actors*
C6	**Stimulus**	Ensure that an authorized user has the rights to access and modify data or services
	Response	Define user groups, roles, or individual listings
	Name	*Authorize actors*
C7	**Stimulus**	Ensure that an authorized user has the rights to access/modify data or services
	Response	Define individual user groups, roles, or lists
	Name	*Limit access*
C8	**Stimulus**	Firewalls that restrict access based on source messages or destination ports
	Response	DMZ configuration (demilitarized zone). Providing Internet services but not a private network

Table 4.2: TaSPer cards list (Part II)

C9	**Name**	*Limit Exposure*
	Stimulus	Exploit a single weakness to attack all data and services
	Response	Host services mapping design for limited services to be available on each host
C10	**Name**	*Encrypt data*
	Stimulus	Data must be protected from unauthorized access
	Response	Confidentiality, data protection, Virtual Private Network (VPN), Secure Sockets Layers (SSL)
C11	**Name**	*Separate identities*
	Stimulus	Separation of different servers
	Response	Reduces the chances of attack of those who have access to non-sensitive data
C12	**Name**	*Change default settings*
	Stimulus	The default settings that are on a system
	Response	Prevents attackers from accessing the system through settings that are publicly available
C13	**Name**	*Revoke access*
	Stimulus	Access must be severely limited to sensitive resources
	Response	Protection of sensitive resources
C14	**Name**	*Lock computer*
	Stimulus	Many failed login attempts
	Response	Mechanisms to prevent potential attacks from non-legitimate users
C15	**Name**	*Inform actors*
	Stimulus	Notify a certain actor
	Response	Report when the system has detected an attack
C16	**Name**	*Maintain audit*
	Stimulus	Collect, group and evaluate evidence of attacks
	Response	Trace actions of an attacker
C17	**Name**	*Restoration*
	Stimulus	Restoration of services
	Response	Recovery of an attack

4.3 Preparation and Collection of Data

1. To start a TaSPer session, the moderator reads an NFR.
2. Each participants is holding a deck of TaSPer cards with 17 cards.
3. The participants discuss the NFR, asking questions to the moderator if is necessary. When the NFR has been fully discussed, each participant privately selects one or more cards to represent his or her option. All cards are then revealed at the same time.
4. If all participants selected the same card, that becomes the most appropriate security tactic. If not, the participants discuss their cards selected. After further discussion, each participants re-selects an card, and all cards are again revealed at the same time.
5. The TaSPer process is repeated until consensus is achieved or until the participants decide that NFR needs to be deferred until additional information can be acquired.

The evaluation process to carry out the pilot study was carried out together with the experts of the toeska group, who tested the use of the tactics proposed, supported the definition of the steps to be followed and proposed improvements to the technique, sharing their experience in the development of TaSPer. The improvements that were implemented in the technique gradually changed the design, allowing the initial design of TaSPer to be generated, and with this, the planning process described above was developed.

This phase was conducted and monitored by two people, the director of the study and an assistant/record. The assistant must support the director at all times and ensure the greatest possible number of registrations.

- **Preparation and Training**:
  - In 1 hour it was introduce TaSPer technique. A presentation was made to the 9 subjects of the exercise, which was spent 20 minutes explaining the security tactics, 20 minutes in the TaSPer process, 10 minutes in the operation of the cards and 10 minutes in all the details regarding the evaluation. Subsequently, there was a process to answer questions.
  - It was appreciated from the subjects a great motivation of participating in an academic experience uncommon to their daily activities and thus be part of the empirical development of a new technique. In addition, the possibility of continuing the process of knowledge exchange in the future was opened.

- **Experimentation**:
  - In this phase it was performed the exercise spending 1.5 hr and it was the next week from the previous phase. In this phase, the case study was presented together with the three selected NFRs. Each NFR was read by the director and then, were given 10 minutes to perform their selection of cards with the corresponding security tactics. After 10 minutes, each person had to argue the reason for each card selected, a process that spent 20 minutes and was guided by the moderator.
- **Post-experimentation**:
  - In this phase it was collected data and it was performed an analysis of the results. The collection, storage and analysis of the obtained information was carried out. This phase was the longest and we carefully reviewed each card selected by each subject according to the NFR. In addition, the audio obtained in the previous phase was analyzed in detail and a subsequent meeting was held with the subjects who participated in the evaluation to obtain their feedback.

4.4 Data Analysis

1. **Selection of cards according to percentage.**

   The objective of this section is to analyze the selection made by each participant, in order to evaluate the design of TaSPer to be used as a technique for making decisions by consensus.

   (a) **Tactics selected by all subjects – 100%.**

   About the selection of tactics made by the subjects, the table 4.3 illustrates those security tactics that were selected by the 9 subjects for each of the NFRs.

Table 4.3: Security tactics selected by all subjects - 100%

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
NFR 1						x	x										
NFR 2		x								x							
NFR 3		x				x	x	x		x						x	

With this, is possible to appreciate (see figure 4.1), that the most observed category of tactics was *resistance to attacks* (reflected in tactics C6, C7, C8 and C10). In this phase the subjects located their analysis based on the textual description of the NFR. At this point, the moderator intervened rarely in the discussion of the subjects. It should be noted that there is no intersection between the three NFRs.

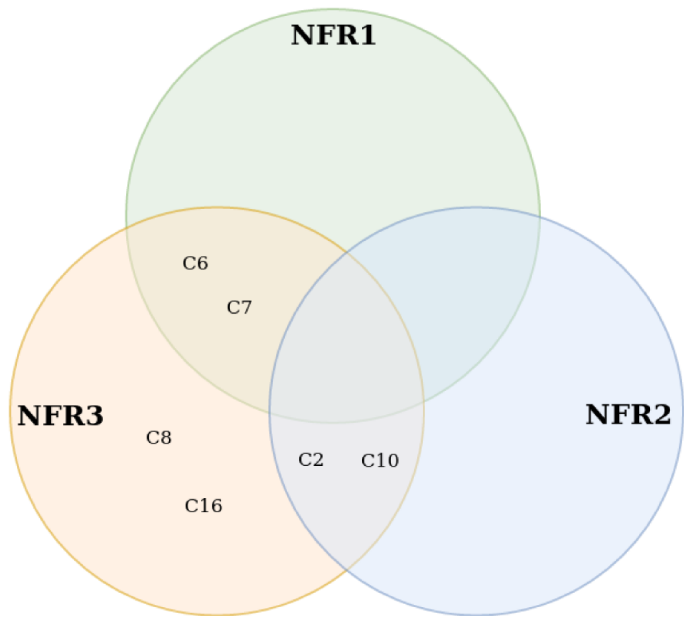


Figure 4.1: Visualization of security tactics selected by all subjects 100%

(b) **Tactics selected by more than 50% of the subjects and less than 99%.**

Regarding the selection of tactics made by the subjects, the table 4.4 describes those security tactics that, on average between a range of 50% and 99%, were selected by for each of the NFRs.

Table 4.4: Security tactics selected by 50% or more subjects and less than 99%

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
NFR 1	x	x						x		x	x	x	x		x	x	x
NFR 2			x					x								x	x
NFR 3	x		x	x							x	x	x		x		x

With this, is possible to appreciate that Figure 4.2) represent the most recurrent decisions made by the subjects. It can be observed that the category *recover from attacks* represents a greater relevance for the subjects, where tactics C17 (*restore services*) predominates in all three NFRs, as well as tactic C16 (*maintain audit*) selected for NFR1 and NFR2.

In addition, it is important to note that the subjects recognize that for the region that intersects NFR1 and NFR3, tactics C13 and C15 are chords to react to attacks.

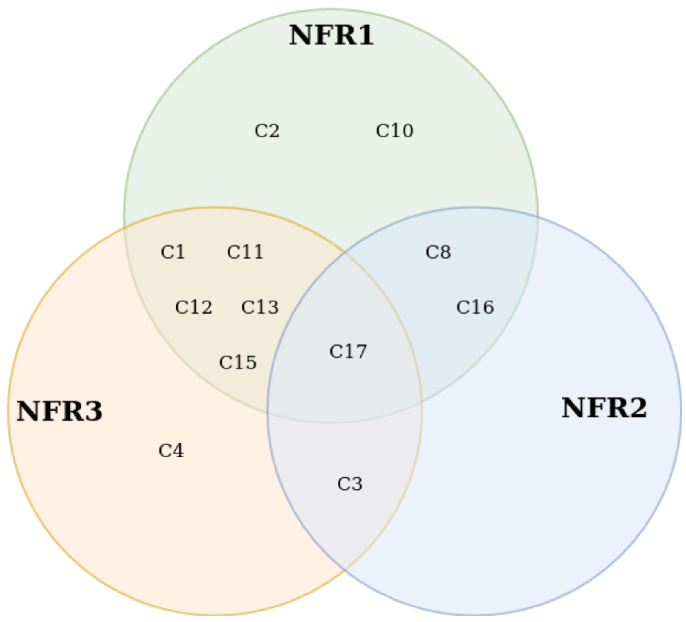


Figure 4.2: Visualization of security tactics selected by more than 50% of the subjects and less than 99%

(c) **Summary of selected security tactics by more than 50% of the subjects.**

Figure 4.3, illustrate the consolidation of Figure 4.1 and 4.2. The subjects revealed that for the proposed NFRs, it is necessary to consider all categories of security tactics. For example, we highlight the tactics C2, C8, C10, C16 and C17, which are transversal to all NFRs. Also, it can be seen that the subjects considered a great similarity between NFR1 and NFR3, where they share 12 security tactics.

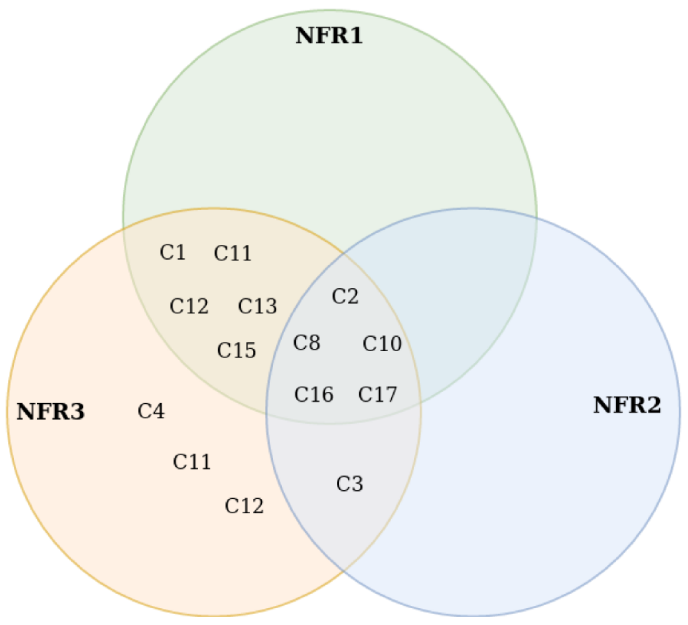


Figure 4.3: Visualization of summary of selected security tactics by more than 50% of the subjects

2. **Selection of cards according to NFRs priorities.**

   TaSPer allow subjects to select not only security tactics, but also a priority associated with each tactic. In this sense, it was observed that for each tactics selected, priorities used were between 3 and 5, excluding priorities 1 and 2. Based on this evidence, the subjects were consulted about their criteria. They responded that they felt that priorities 1 and 2 were similar to discarding the card, which is why they prefer not to use them.

   With respect to the prioritization of tactics, it was possible to see that there was a trend towards a higher prioritization (4.79) (see table 4.5) in the segment of security tactics selected by all subjects, whereas in security tactics selected by more than 50% of the subjects and less than 99 %, the average priorities were 4.38 (see table 4.6). From the above reason, it is possible to mention that when all the subjects agreed with a particular security tactic, they also did so with the selected prioritization level.

   To obtain the average of the priorities in this segment, a priority ratio was used. This ratio considers the priority established by tactics versus the number of subjects that selected that tactic.

Table 4.5: Priority of security tactics selected by all subjects - 100%

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	$\overline{X}$
NFR 1						4.88	4.66											4.78
NFR 2		4.77								4.89								4.83
NFR 3		4.67				4.78	4.89	4.78		5						4.22		4.76
																		$\overline{X}$ priority= 4.79

Table 4.6: Priority of security tactics selected by 50% or more subjects and less than 99%

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	$\overline{X}$
NFR 1	4.75	4.4						4.67		4.3	3.8	4.8	4.2		4.5	4	4.6	4.4
NFR 2			4					4.7								4.3	4.2	4.3
NFR 3	5		4.3	3.7							4.5	4.25	4.5		4.75		4.6	4.44
																		$\overline{X}$ priority= 4.38

3. **Interventions made by subjects.**

   The table 4.7 shows the number of interventions per subject. Interventions are those suggestions performed by the subjects at the moment of give an opinion for each NFR.

   The interventions were counted manually after the exercise. The average number of interventions observed was 6.9 interventions per subject. By observing each intervention in detail, it is possible to determine that the TaSPer process allows the interaction and intervention of all participants. It also allowed us to verify that there was a quick familiarization for most of the issues related

Table 4.7: Number of interventions per subject

Interventions	
Subject	Nº interventions
1	12
2	6
3	8
4	8
5	6
6	4
7	5
8	7
9	6
Total	62
Average	6.89

to NFRs, which allowed them to participate in a group form, giving different points of view.

Key finding (1)

TaSPer allows the subjects to establish trade-offs between security tactics.

An interesting finding that we observed at the moment of TaSPer was executed is that the process allowed the subjects to discuss which security tactic is most appropriate for each NFR. For example, subjects with more experience corrected novice subjects because they did not know the value of the meaning of security tactics. Experts could see that certain security tactics *shut down* (they used that term) to other security tactics when analyzing NFRs. This type of analysis allowed the experts to select a low amount of security tactics and the novices, a high amount.

Key finding (2)

Subjects with work experience select security tactics more accurately than subjects with no experience.

Clearly, experience in selecting security tactics plays a fundamental role. The expert subjects selected security tactics not only thinking about the design

decision regarding security, but also in the context of the business. They appreciated that the selection of security tactics has a close relationship with the objectives of the business that wants to achieve the project under study.

4. **General view of project's software architects.**

   The experience with TaSPer for this case study revealed that the technique was a contribution to the project security decision-making. Once the results of TaSPer were presented to the software architects, they gave their feedback about it. For NFR1, the most attractive security tactics were C6 and C7. From the point of view of data protection, the architects had a notion of authenticating and authorizing users. However, seeing that the subjects of the studies also coincided with their vision, they are now more certain that the final design of the project must involve these resistance mechanisms of attacks. For NFR2, the same situation occurs. The security tactics selected by the subjects also coincided with the decisions taken by the architects. From the beginning, they had decided that implementing audit-related decisions could help them recover faster from potential attacks on the project. However, there was much discrepancy for NFR3. The security tactics selected by the subjects did not satisfy the decisions that the architects thought. Because TaSPer was applied in a pilot test, the moderators were unaware of aspects of the business that the architects already knew. Therefore, the architects agreed that TaSPer should consider business guidelines within the TaSPer process.

4.5 Summary

In order to verify our proposal, we prepared a case study called "Innovation Management Project", that allowed us to verify the following expected objectives: rapid familiarization of subjects with the TaSPer process; good interaction and integration of the different participants; quick understanding of architectural tactics; proper selection of tactics by different subjects and collection of valuable data to be used by software architects.

To do this, we present the case of study with three security NFRs too a group of professionals dedicated to IT security, who are not familiar with security tactics. The activity include three phases: the prior phase to the exercise, the phase in which the exercise is performed and the final one of collecting data called Preparation and training, Experimentation and Post-experimentation respectively.

The results obtained revealed that TaSPer allows to fulfill the imposed objectives, showing the subjects' quick familiarization, good process of interaction and integration as well as the possibility to quickly learn the security tactics, the ones with which they were able to select the ones that support an actual architect. Besides that, TaSPer allows the subjects to establish trade-offs between security tactics,

and was observed that the subjects with work experience showed a more accurate selection of security tactics than subjects with less experience.

Chapter 5

Experimental study: "LockInfo" a messaging system for secure communication

THE purpose of this chapter is to validate the TaSPer technique through the experimental process [2][42] considering the stages: definition, planning, evaluating, execution and analysis. Thus, we carried out an experimental study considering the development of the application "LockInfo" (a messaging system for secure communication), considering the participation of a group of 20 practitioners and IT experts in a professional IT Master Program[1] of the UTFSM in Santiago.

The chapter considers the following sections: Section 5.1 describe the introduction of the experimental process and establish their stages; then Section 5.2 describe the experimental process, setting the research objective and questions; Section 5.3 show the planning stage, considering the Context Selection, Hypothesis, Selection of Subjects, Study object, Choice of design type, Instrumentation; Section 5.4 describe the evaluation, where we made two pre experiments; Section 5.5 present the execution of the experiment, and Section 5.5 the analysis of the experiment describing the results obtained and the analysis made; Section 5.7 describe the post-experiment survey; and finally, Section 5.8 summarize the results of the experiment.

5.1 Introduction

In "Experimentation in software engineering," Basili et al. [2] presented a framework that establishes the fundamental bases of the experimental process. In it, they consider the stages of definition, planning, and orchestration, which allows us to observe and generate a correct evaluation and execution of the experiment and the compilation, analysis, and interpretation of the information obtained. Following the

[1] https://www.mti.cl/

same line of argument, it is possible to find complementary studies and research [46] [42] that adapt and develop the work done in [2]. The works named above were used to evaluate TaSPer's fitness to support decision-making, considering the development in five stages [42] (see fig: 5.1): definition, planning, evaluating, execution and analysis.

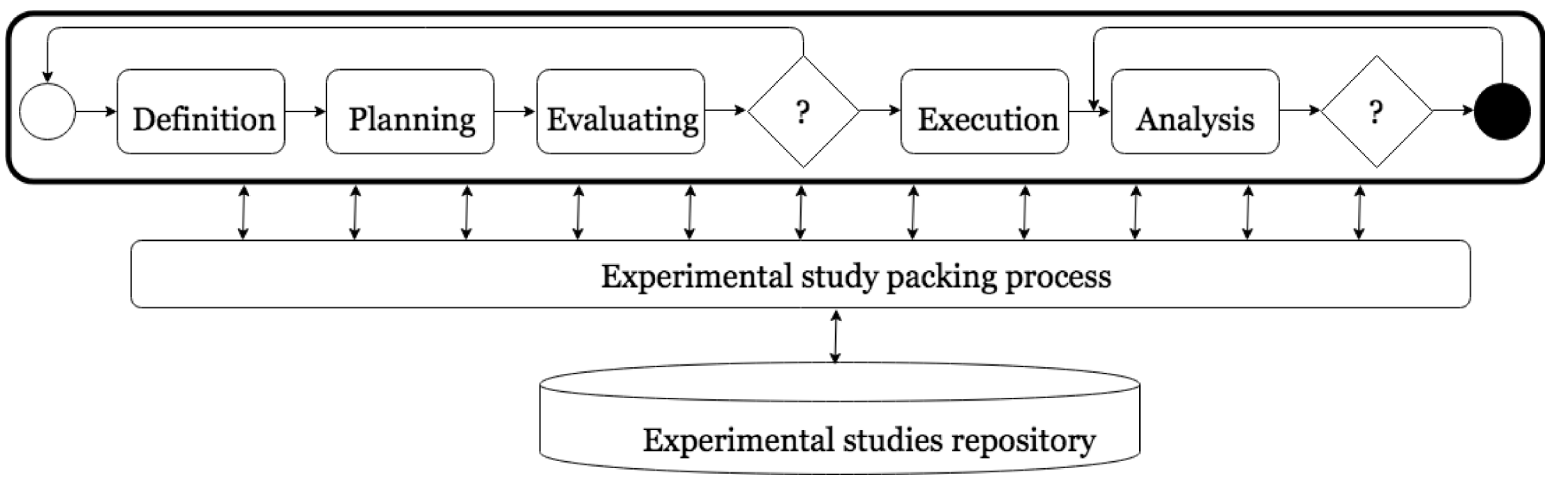


Figure 5.1: Travassos's Experimental Process.

5.2 Definition

Research objective

The research objective is to validate TaSPer to establish a technique that allows the selection of security tactics in a consensual way and contributes to developing secure software related to agile methods. To do this, we evaluate TaSPer in two conditions: (a) the decisions made to select security tactics, carried out individually versus the one made by consensus and (b) the selection using TaSPer versus those that do not, both comparing with the ground truth proposed by a group of experts.

Research question Experimental Study

Based on the research objectives we established the following research questions (RQ):

- **RQ5:** Does TaSPer improve individual design decisions? This question is highly relevant. It seeks to determine that the TaSPer process, through consensual decisions, generates a selection closer to the ground truth defined by the experts than the decisions made individually; if possible, compare the actual decisions taken by consensus versus made in isolation.

- **RQ6:** Having an established technique allows us to obtain results closer to the ground truth than not having it? Regarding this question, we expected

that using a technique will allow better results when selecting tactics than not possessing them. For this, we will compare the selection of tactics using TaSPer versus groups without it.

5.3 Planning

The planning process establishes the parameters related to how the experiment will be carried out, from selecting the context to achieving the experimental design as results. This study was planned following the Wohlin at al.'s proposal [46] (see figure 5.2) and is divided into seven steps

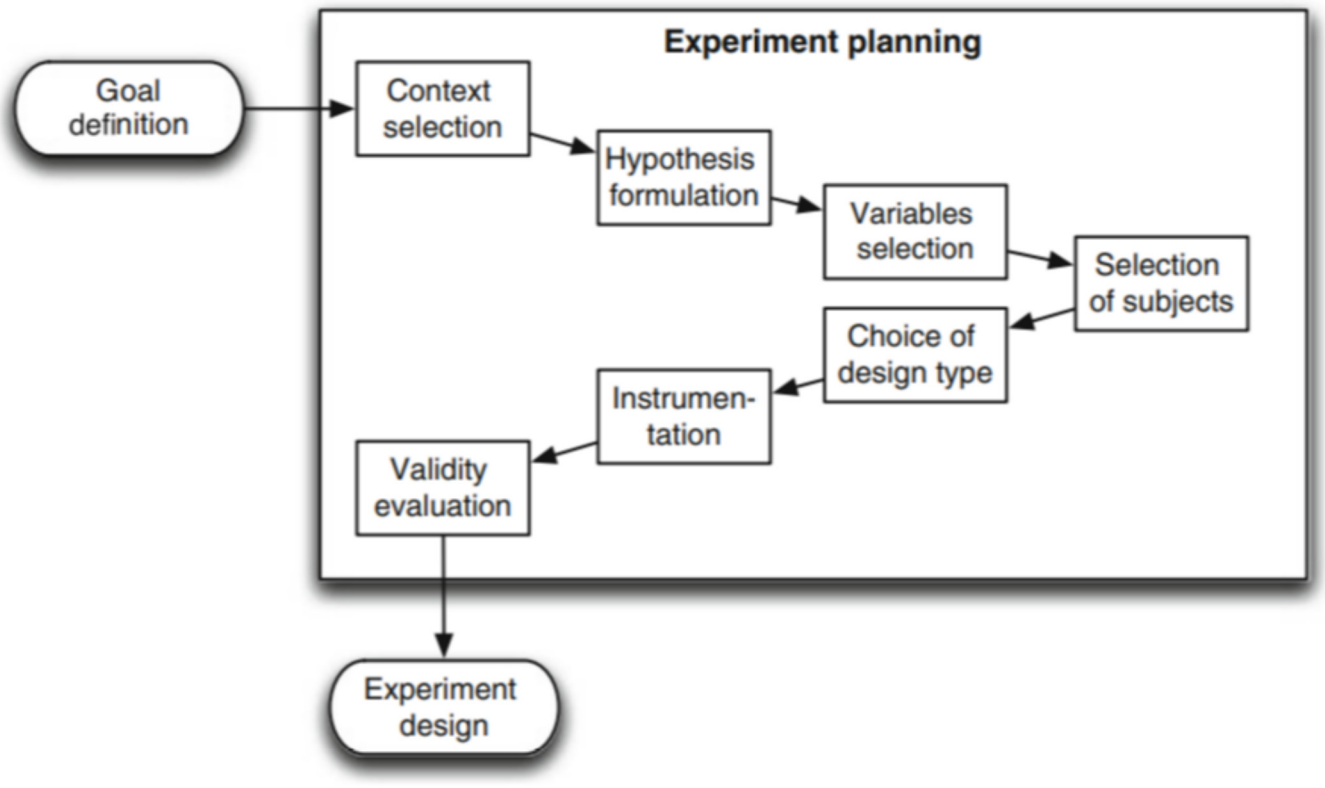


Figure 5.2: Planning phase overview [46]

5.3.1 Context Selection

The technique's execution considers establishing a controlled environment, which would allow obtaining valuable information related to the selection of architectural tactics. For this, we experimented with a group of 20 practitioners and IT experts in a professional IT master program (MTI)[2] of the UTFSM in Santiago. Professionals received an induction related to design decisions and security tactics, later separated into groups, to finally develop the activity, presenting a case study with four scenarios taking fifteen minutes to select tactics per stage.

[2]https://www.mti.cl/

5.3.2 Hypothesis

The hypothesis allows us to specify what we want to achieve in the investigation and determine the intermediate results required to direct the experiment's conclusions:

- $\mathbf{H_{0.1}}$ *Consensual decision-making through the interaction between stakeholders does not produce better results, on average, concerning the precision, recall, and accuracy compared with a ground truth in the context of architectural tactics selection in a specific quality attribute scenario than the decision made by a stakeholder individually.*

- $\mathbf{H_{1.1}}$ *Consensual decision-making through the interaction between stakeholders produce better results, on average, concerning the precision, recall, and accuracy compared with a ground truth in the context of architectural tactics selection in a specific quality attribute scenario than the decision made by a stakeholder individually.*

- $\mathbf{H_{0.2}}$ *A Decision-making technique through interaction between stakeholders does not produce a difference between the ranks of precision, recall, and accuracy regarding the ground truth about architectural tactics selection in a specific quality attributes scenario.*

- $\mathbf{H_{1.2}}$ *A Decision-making technique through interaction between stakeholders produces better ranks of precision, recall, and accuracy regarding the ground truth about architectural tactics selection in a specific quality attributes scenario.*

Variables Selection

Regarding this point, the study was carried out considering keeping the scope concentrated on the following variables:

- **Dependent:** The dependent variable is the selected Tactic.
- **Independent:** The independent variable is the Decision-making Technique.

Measurement

To measure these variables and evaluate our hypothesis, we used as measurement instruments the *loss function* (to evaluate the error rate) [18], and the *precision, recall, and accuracy ratios* (to evaluate the accuracy and coverage) [30]. These instruments seek to measure the difference between decision making of security tactics when interaction among participants is allowed and not.

- **Loss function:** The loss function outlines values of one or more variables over a real number, intuitively representing some *cost* associated with the event [18], achieving the measuring of accuracy. The error rate function used in our experiment was:

$$l(y, \hat{y}) = \begin{cases} 0 & if\ y = \hat{y} \\ \\ 1 & if\ y \neq \hat{y} \end{cases}$$

  Although the loss function has many potentials for our purposes, the main reasons why we have used this function are described below:

  - It allows comparing the quality of the results
  - The loss function penalizes (punishes) the wrong choice of security tactics, as compared to the ground truth
  - The loss function average is obtained from the number of security tactics, getting the error
  - The expected result is a decrease in error rate, i.e. comparing the mean of the individual decisions versus the consensual decisions

- **Precision, recall, and accuracy:** To measure accuracy and coverage, we used ratios widely used in Information Retrieval [30]:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

  Where:

  - True Positive TP= Tactics correctly selected.
  - True Negative TN= Tactics correctly NOT selected.
  - False Negative FN= Tactics that had to be selected but were not selected.
  - False Positive FP= Tactics that should not be selected, but were selected.

- **Ground Truth:** To enable results evaluation, it was necessary to work with a group of expert architects from the Toeska laboratory who made the selection of tactics and determined the ground truth for the experimental study (See Table 5.1).

Table 5.1: Ground truth experts

Id	Experience (Years)	Industry	Security Domains
1	7	Financial and Shipping	Secure Financial Transactions
2	14	IT consulting, Environmental Consulting, Networking and Telecommunications	Ethical Hacking, System Hardening, Information Security Management
3	8	R&D in Measurement Equipments, R&D in Networking and Telecommunications, R&D in Robotics, R&D in Automation Systems, IT consulting	Information Security Management
4	25	Financial, Transportations, R&D, Consulting	Financial, Distributed DB, OODB, White Hat, Stress Testing
5	7	Networking and Telecommunications	Information Security Management

5.3.3 Selection of Subjects

The experimentation phase considered a group of 20 practitioners with varying levels of experience in software development, IT, IT architectures, or related fields, and ideally, have experience in security issues. The subject's diversity allows us to evaluate the technique usability and its impact on people with different experience levels.

Four groups of five people each, randomly separated, were considered for experimentation. Before experimentation, we studied the participants' experience to determine that there were no person with specific software architecture experience, design decisions, or tactics.

5.3.4 Study object

The study object was about a freeware cross-platform messaging application called "LockInfo" with four specific scenarios. In each session, subjects received a document that described the case study context and scenarios. Each scenario was related to a particular security issue.

Application Context: It is required to create a messaging system that allows

the company members a protected communication between them to protect the information exchanged. Thus, it is necessary to develop a software/application that guarantees the transfer of data. Simultaneously, it must be a multiplatform system; this means that it has to work on mobile phones, tablets, and desktop computers.

The aim is to create a system similar to WhatsApp, Telegram, Line, or Messenger; however, it must be specially adapted to the company's needs, considering the scenarios described below.

- **Scenario 1:** LockInfo must ensure that platform users can only perform the allowed actions for their profile.
- **Scenario 2:** Both audio and video calls made through LockInfo must be protected at all times. Both emitter-receiver and receiver-emitter (end-to-end). It is necessary to guarantee the protection of the information from the very beginning until the communication ends.
- **Scenario 3:** LockInfo must ensure that data is stored securely on the different installed devices.
- **Scenario 4:** LockInfo must allow historical records to be consulted and the ability to rebuild and monitor system resources. It is necessary to be able to control the security study on the platform in case of attacks.

5.3.5 Choice of design type

This section describes how the tests are organized and run, considering the general design principles of blocking and balancing:

- **Randomization:** The selection of the persons (subjects) will represent the stakeholders by a random sample of the available subjects. The assignment to each treatment (consensus decision over the individual) is selected randomly.
- **Blocking:** The persons (subjects) used, for this experiment, have a different experience. Some of them are students, and others are professional IT. To minimize the effect of the participants' experience, the persons were grouped in a randomization way using random selection.
- **Balancing:** The experiment uses a balanced design, which means that there is the same number of persons in each group (block).

5.3.6 Instrumentation

The instrumentation is the phase that determines the actions that must be done for experimentation before execution. In this section, we present the processes, documents, procedures, questionnaires, control, and monitoring used during the experiment.

Experimental sequence

The experimental sequence has three stages which are complementary to each other. We want to give the context and guide to the participants and a scheme that allows the experiment's control and monitoring.

- **Preparation and Training**:

  The first stage considers the experiment's preparation, called "Preparation and Training," which follows the following sequence:

  - Presentation of the people in charge of the experiment's execution and supervision (Leader and monitors of the experiment).
  - Presentation of the study, where the introduction to the activity to be carried out is made, emphasizing the design decision process in software architecture.
  - Induction to the participants: Process that considers the introduction to software architecture. Additionally, examples of decision making are shown with a model that contains requirements and related decisions.
  - Induction regarding security tactics: Introduces and explains security tactics.
  - Groups separation, participants were separated into two groups of five people, controlling for the experience.
  - The group that will use TaSPer is introducing how the technique works, the cards' use, and explaining the steps to follow.
  - Delivery of documentation: one group received the **catalog of security tactics** and **textual description of the security tactics**, and the other group received **the TaSPer card game** (see table 5.3) with the instructions of use (see Section 3.2.3).

- **Experimentation**:

  This phase considers the execution of the experiment and the interaction of all the participants.

  - Presentation of the general scenario conducted by the leader of the experiment
  - Presentation of the stage with 15 minute spaces between them.
  - Decision making for each scenario.

- **Post-experimentation**:
  - Brainstorming to obtain experience of the subjects, allowing drivers to obtain information about improvements, challenges, and benefits when using TaSPer.
  - Data analysis, a process carried out only by the driving group allowing to obtaining data regarding the use of the technique. The data obtained is verified against the ground Truth carried out by the group of experts

Conducting and Monitoring

It is essential to have a leading group to guide the exercise. This group should continuously monitor the experiment, verifying that it was following the established procedures. We determined that the ideal number of monitoring the investigation was five people, which should be considered a minimum of 3. The important positions are:

- **Director:** Responsible for guiding the efforts of the experiment.
- **Registration:** In charge of registering all the data regarding the exercise (time, participants, questions asked, etc.).
- **Monitor:** Person in charge of supporting the director's work and supervising the development of the study.

Data registration forms

We developed some registration forms to collect the information in a unified and reliable way. The template has all the selectable tactics that will be discussed by the participants. In the example, we can see the template designed for the 17 security tactics.

Table 5.2: Sample template for selecting tactics

GROUP N°____			
Scenario N° _____			
NUMBER	TACTICS NAME	NOT SELECTED TACTICS	SELECTED TACTICS
		Mark with an "X"	
Detect Attacks			
1	Verify Message Integrity		
2	Verify Storage Integrity		
3	Maintain Audit Trail		
4	Identify Intrusion by Signature		
5	Identify Intrusion by Behavior		
Stop or Mitigate Attacks			
6	Authenticate subjects		
7	Authorize Subjects		
8	Manage Security Information		
9	Filter Data		
10	Verify Origin of Message		
11	Establish Secure Channel		
12	Hide Data by Encription		
13	Hide Data by Steganography		
React to Attacks			
14	Alerts Subjects		
15	Apply Institutions Policies		
Recover from Attacks			
16	Audit Actions		
17	Apply Institutions Policies		

Tactics for the experiment

Regarding the experiment's tactics, the "Toeska" Software Engineering Group work to refine them for a better understanding. Also, we translate them into Spanish for future work.

Table 5.3: Security Tactics used in the experiment

CATEGORY	TACTIC NAME	OBJECTIVE	DECISION
Detect attacks	Verify message integrity	Define measures to make sure that message have not been modified	Procedure to ensure verification of data integrity (eg. checksum or hash values)
	Verify storage integrity	Ensure data is recorded exactly as intended.	Procedure to ensure verification of data integrity (ex. manage hash values)
	Maintain audit trail	Collect, group and evaluate evidence of attacks	Trace and identify the actions of an attacker
	Identify intrusion by signature	Footprint left behind by perpetrators of a malicious attack on a computer network or system.	Determine by evidence what the intrusion was, how and when it was perpetrated, and even how skilled the intruder is
	Identify intrusion by behavior	Identify malicious behavior stored in systems	Detect malicious behavior in protocols, applications, and others. baseline or learned pattern of normal system activity to identify active intrusion attempts.
Mitigate attacks	Authenticate subjects	Ensure that a subject (a user or a remote computer) is actually who or what it purports to be	Password, one-time password, digital certificates, and bio metric identification provide a means for authentication
	Authorize subjects	Ensure that an authenticated user has rights to access and modify either data or services	User can be define by individual, user groups, roles, or list
	Manage security information	A way to maintain the integrity of data and to make sure that the data is not accessible by unauthorized parties or susceptible to corruption of data	Management of keys for cryptography, the secure storage of authorization rules, and other ways to handle information
	Filter data	Avoid attacks based on abnormal inputs or from untrusted sources	Implement content filters on the data received through user requests to the system
	Verify origin of message	Determining the location of the sender of the message	Implement the sender verification mechanism. ex: use a digital signature to verify that the source is reliable and avoid the issuer's impersonation.
	Establish secure channel	Provide secure communications	Enables additional endpoint authentication, message encryption, and message authentication to be
	Hide data by encryption	Data must be protected from unauthorized access	Confidentiality, data protection, virtual private network (vpn), secure sockets layers (ssl)
	Hide data by stenography	Hiding secret data inside other innocent-looking data, called the container, carrier or cover.	Algorithms or techniques of concealing data.
React to attacks	Alert subjects	Notify a certain actor	Report when the system has detected an attack
	Apply institutions policies	Securing an organization's information	Develop security policy and put into practice throughout the organization
Recovery from attacks	Audit actions	Manual or systematic measurable technical assessment of a system or application	Perform security vulnerability scans, reviewing application and operating system access controls, and analyzing physical access to the systems
	Apply institutions policies	Securing an organization's information	Develop security policy and put into practice throughout the organization

5.3.7 Validity Evaluation

- **Conclusion validity:**
  - Before the experimentation, different activities are carried out, including a pilot study and a pre-experimental process, which allowed us to verify the technique's operation.
  - A guide was developed that would allow the correct development of the experiments.
  - The experimentation was carried out in large spaces, allowing to guarantee the participating groups' isolation, assuring that there was no interference between them. Thus, no external factor affected the experiment.
- **Internal validity:**
  - To not affect the internal validity, we present the scenarios consecutively.
  - The study considers that all subjects face the same experimental situation.
  - The forms used for the study were developed to serve only as a guide and not to induce responses.
  - We allocated the Subjects randomly to the groups.
- **External validity:**
  - Due to the nature of the study carried out and the fundamental basis of decision-making in a consensual manner, it is possible to replicate this experiment in different situations outside the context of software Architecture.
- **Construct validity:**
  - Training activities (fifteen minutes) were performed to avoid confounding threats due to previous lack of knowledge of tactics.
  - The training covered concepts and application of security tactics, including using a security tactics catalog (taxonomy).
  - After the induction process that was carried out on the subjects, the interaction of the experiment's conductors was reduced to a minimum, not to induce the decision-making of each of the groups.
  - The study was not evaluated in any part of the process, in order not to generate higher expectations from the subjects; on the other hand, they were encouraged to be part of the study by establishing an incentive to appear in the experiment.

5.4 Evaluation

Continuing with the evaluation, two pre-experiments were carried out with undergraduate and graduate students in different universities. The undergraduate students were teaching assistants in Software Engineering[3] of the UTFSM university. The postgraduate students were Master of Science students at Bío-Bío University[4], Chile.

Once we made the pre-experiments, the students revealed that they found original how to assess security aspects in architecture using TaSPer. Many of them did not know security terms used to evaluate architectures, so this activity allowed them to enlarge their knowledge regarding security in software systems. Independent of the experience lived in these pre-experiments, an essential point of these activities was to refine our experimental design. Therefore, we describe the key findings from these pre-experiments:

- Students (both undergraduate and graduate) who use TaSPer required less time to complete the study than those students who only had the security tactics descriptions.
- TaSPer allowed selecting options within the range of available security tactics but can generate mechanization of solutions found by the subjects. This procedure can accelerate the selection of security tactics but limits the possible solutions to a particular problem in a specific area since cards prevent the decision-makers from innovating in any way before an attack event.

We notice three main lessons learned from the experimental process:

- Training before the exercise is essential; it must explain what security tactics are and how they work.
- The timing of each evaluation phase must be controlled, especially for decision making.
- Well-defined scenarios are critical to avoid unfeasible conclusions when subjects choose tactics.

[3]https://www.inf.utfsm.cl/english
[4]http://www.mcc.ubiobio.cl/

5.5 Execution

On December 03, 2017, the experiment execution was from 15:30 to 17:50, driven by five people considering one director, one registration, and three monitors (see figure 5.3 . Twenty professionals participated in the experiment, separated into four groups of five people (G1, G2, G3, and G4) applying TaSPer to groups G2 and G4. We improved the technique execution phase considering the lessons learned and key findings from the case study and pre-experiments. The experimental sequence is shown in the table 5.4:

Table 5.4: Experimental sequence

Stage	Initial	Finish
Preparation and training	15:30	16:00
Experimentation	16:30	17:41
Post-experimentation	17:41	17:50

We guide the experimentation with the support of a presentation showed to all participants. We base the presentation on the experimental sequence 5.3.6, considering the formation and training process highlighting the following points of the *preparation and training* stage:

- What are design decisions?: An explanation regarding design decisions is made along with explaining the definition of security tactics, ending with an example of a real use case of requirements and decisions (CGI Energy Data Architecture).

- Group separation: The participants were separated into groups of five people.

- Explanation and development: An induction process was carried out regarding the activity, emphasizing the context of the experiment, what each participant must do, the presentation of the scenarios to evaluate, and the explanation of the TaSPer technique.

Figure 5.3: Monitor explaining the preparation and training stage

5.6 Analysis

The analysis section is the last stage of the experimental process [2][42]; the main objective is to determine whether or not it is possible to reject the hypothesis about research questions RQ5 and RQ6. We divide this process into two phases, as shown in figure 5.4, scenario analysis, and statistical analysis.

1. **Scenario Analysis:** This phase begins on the experiment's execution with particular attention to the participants' behavior, emphasizing their individual and group actions for decision-making (subjects rationale). Subsequently, the participants' decisions must be collected individually and in groups (after the treatment used for the selection of tactics), analyzing each scenario's results correlating the findings with the expert's ground truth and thus obtain precision, recall, and accuracy.

2. **Statistical analysis:** The results obtained of the first phases must be grouped by metric, letting them compare with each other and, therefore, verify, through a statistical test, if it is possible to determine that the results are statistically significant and accept or reject the established hypotheses. To achieve this,

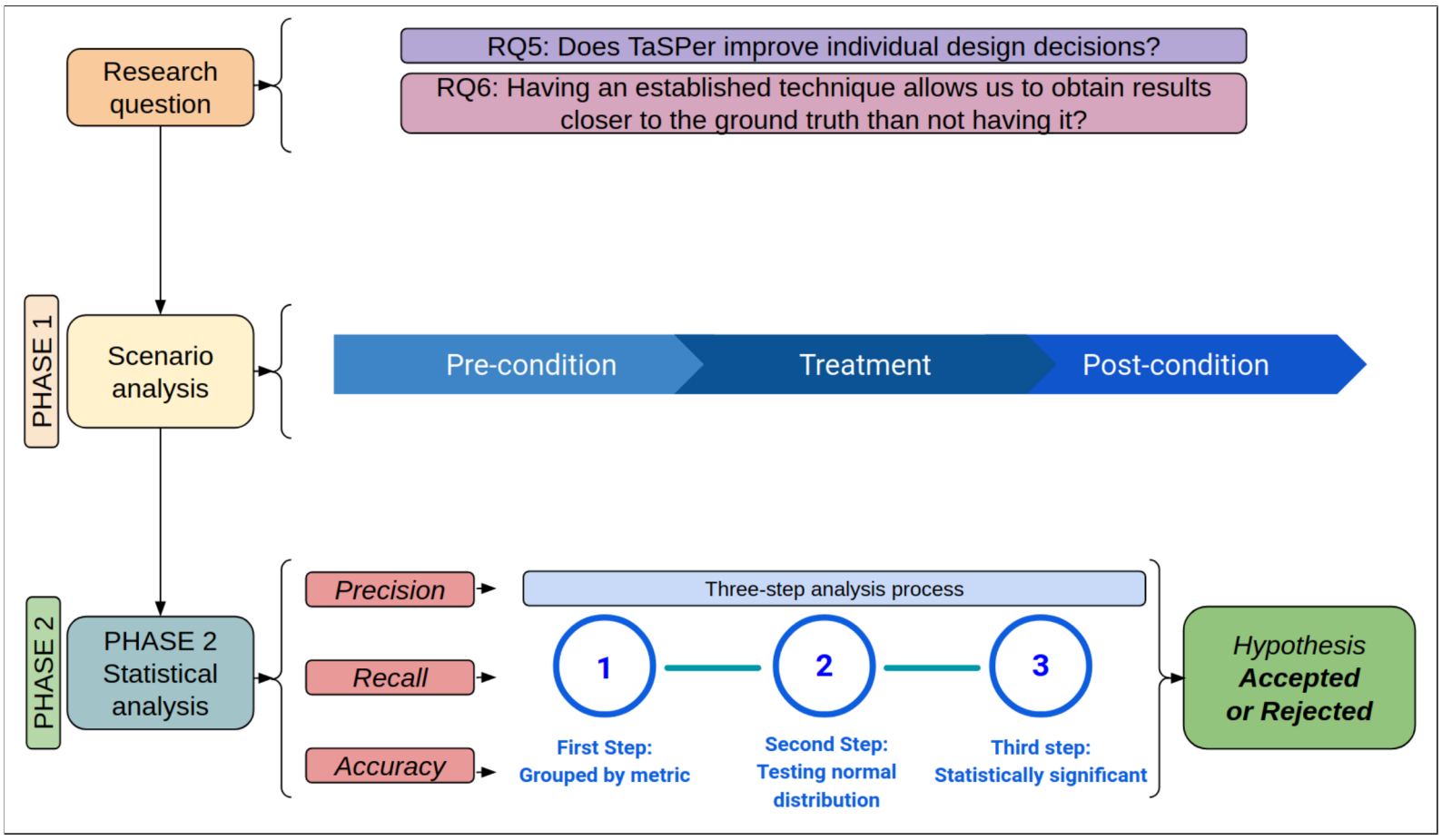


Figure 5.4: Analysis phases

we set a three-step analysis process to carry out this phase, showing a specific order to carry out the analysis.

- **First Step:** ***Grouped by metric***
  The first step considers sorting and grouping the results by metric obtained in the scenarios (from S1 to S4).
- **Second step:** ***Testing normal distribution***
  With the grouped results obtained from step one, we must test the normal distribution of the results. For this we apply the Shapiro-Wilk test, a normality distribution test to verify the data's distribution and establish if they are parametric or non-parametric.
- **Third Step:** ***Statistically significant***
  To determine the statistical test to be applied, it is crucial to recognize that the experimental study considers comparing a precondition and post-condition relative to the treatment (technique) that we used. The treatment with the result obtained from the second step (Shapiro Wilk test) empowers us to chose the T-test if the results were parametric, and if not, we choose the Wilcoxon or Mann – Whitney tests [15][24].
  Finally, knowing the test, it is necessary to apply it over the grouped decision making, letting us calculate the p-value, which, if it is less than 0.5, confirms that the results are *statistically significant*, representing a powerful indicator to reject the null hypothesis.

5.6.1 Research question 5 (RQ5)*: Does TaSPer improve individual design decisions?*

Scenario analysis

To answer this question, it is necessary to apply the treatment (proposed technique) to the same subjects, comparing them in two different conditions, considering the results before treatment and after treatment (see table 5.5). For this, we applied the TaSPer technique to groups 2 and 4 in each scenario, obtaining the selection of security tactics individually by each participant, then applying the treatment and subsequently obtaining the results in a consensual way. Finally, we analyzed the subjects' decisions based on the precision, recall, and accuracy results obtained.

Figure 5.5: Treatment applied for subjects of groups G2 and G4.

To collect the individual decisions, we gave them a template to establish their records (see 5.2). On the other hand, each group's moderator had a particular template that would allow recording individual selection and consensus.

The results obtained can be seen from tables 5.5 to 5.8. These tables present 17 security tactics that identify the different scenarios from 1 to 4, separating each scenario in two groups (G2 and G4), presenting each subject's decisions (S1, S2, S3, S4, and S5). Finally, it is possible to appreciate the consensual results (CR) obtained by each group and the ground truth (GT) defined by the experts. It is essential to highlight that, on these tables, the 1s represent the selected tactics, and the 0s represent not selected tactics.

Once we obtained the tables described above, it is possible to determine and evaluate the results considering the measurements set in 5.3.2 and with this establishing True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) of each stakeholder, as well as obtain their consensual result (CR), which finally allows us to calculate **precision, recall, and accuracy** for each group and scenario. The results obtained can be seen from table 5.5 until table 5.8.

Table 5.5: Decisions scenario 1

		Scenario 1												
		GT	***GROUP G2***						***GROUP G4***					
N°	TACTICS		**CR**	S1	S2	S3	S4	S5	**CR**	S1	S2	S3	S4	S5
Detect Attacks														
1	Verify Message Integrity	1	**1**	1	0	1	1	1	**1**	1	1	1	1	1
2	Verify Storage Integrity	0	**0**	1	0	1	0	0	**0**	0	0	1	1	0
3	Maintain Audit Trail	0	**1**	0	1	1	0	1	**1**	0	1	1	0	0
4	Identify Intrusion by Signature	0	**0**	0	1	0	0	0	**0**	0	1	0	0	1
5	Identify Intrusion by Behavior	0	**1**	0	1	1	0	1	**0**	1	0	0	1	1
Stop or Mitigate Attacks														
6	Authenticate subjects	1	**1**	0	0	1	1	0	**1**	1	1	1	1	1
7	Authorize Subjects	1	**1**	0	1	1	1	0	**1**	1	1	1	1	1
8	Manage Security Information	1	**1**	1	0	0	0	0	**0**	0	1	0	1	1
9	Filter Data	0	**1**	0	1	1	0	1	**0**	0	0	0	0	1
10	Verify Origin of Message	1	**1**	0	0	1	0	1	**0**	0	0	1	0	0
11	Establish Secure Channel	1	**1**	1	1	1	1	1	**1**	1	0	1	1	1
12	Hide Data by Encription	1	**1**	1	0	1	1	1	**1**	1	1	1	1	1
13	Hide Data by Steganography	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0
React to Attacks														
14	Alerts Subjects	0	**0**	0	0	1	0	0	**0**	0	1	1	1	0
15	Apply Institutions Policies	0	**0**	0	0	0	1	0	**1**	1	1	0	1	0
Recover from Attacks														
16	Audit Actions	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0
17	Apply Institutions Policies	0	**0**	0	0	1	0	0	**1**	1	1	0	1	0

Table 5.6: Decisions scenario 2

		Scenario 2												
		GT	***GROUP G2***						***GROUP G4***					
N°	TACTICS		**CR**	S1	S2	S3	S4	S5	**CR**	S1	S2	S3	S4	S5
Detect Attacks														
1	Verify Message Integrity	1	**1**	1	1	1	1	1	**0**	1	0	1	0	1
2	Verify Storage Integrity	0	**0**	0	0	0	0	0	**0**	1	0	0	1	0
3	Maintain Audit Trail	0	**1**	0	1	1	1	1	**0**	1	0	0	1	1
4	Identify Intrusion by Signature	0	**0**	0	0	0	0	1	**0**	0	0	1	0	1
5	Identify Intrusion by Behavior	0	**0**	0	0	0	0	1	**0**	0	0	0	0	0
Stop or Mitigate Attacks														
6	Authenticate subjects	0	**0**	1	0	0	1	0	**0**	0	0	1	1	1
7	Authorize Subjects	0	**1**	1	1	0	1	0	**0**	0	0	0	1	1
8	Manage Security Information	0	**1**	0	0	0	0	1	**0**	0	1	1	1	0
9	Filter Data	0	**0**	1	1	0	0	0	**1**	0	0	0	1	1
10	Verify Origin of Message	1	**1**	1	1	1	1	0	**1**	1		1	1	1
11	Establish Secure Channel	1	**1**	1	1	1	1	1	**1**	1	1	1	1	1
12	Hide Data by Encription	1	**1**	1	1	1	0	1	**1**	1	1	1	1	1
13	Hide Data by Steganography	0	**0**	0	0	0	0	1	**0**	1	1	0	1	0
React to Attacks														
14	Alerts Subjects	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0
15	Apply Institutions Policies	0	**0**	0	0	0	0	0	**1**	1	0	1	1	1
Recover from Attacks														
16	Audit Actions	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0
17	Apply Institutions Policies	0	**0**	0	0	0	0	0	**1**	1	0	1	1	1

Table 5.7: Decisions scenario 3

		Scenario 3												
		GT	***GROUP G2***						***GROUP G4***					
N°	TACTICS		**CR**	S1	S2	S3	S4	S5	**CR**	S1	S2	S3	S4	S5
Detect Attacks														
1	Verify Message Integrity	0	**0**	0	0	1	1	1	**0**	0	1	0	0	0
2	Verify Storage Integrity	1	**1**	1	0	1	0	1	**1**	1	1	0	1	1
3	Maintain Audit Trail	0	**0**	0	0	0	1	1	**1**	1	0	1	0	0
4	Identify Intrusion by Signature	0	**0**	0	1	0	0	1	**0**	0	0	0	0	0
5	Identify Intrusion by Behavior	0	**0**	0	1	0	0	1	**0**	0	0	0	0	0
Stop or Mitigate Attacks														
6	Authenticate subjects	0	**0**	0	0	0	1	0	**0**	0	0	1	0	0
7	Authorize Subjects	0	**1**	0	0	0	1	0	**0**	0	0	1	0	0
8	Manage Security Information	1	**1**	1	0	1	1	1	**1**	0	1	1	1	0
9	Filter Data	0	**1**	1	0	0	0	0	**0**	0	0	0	0	0
10	Verify Origin of Message	0	**0**	0	0	0	1	0	**0**	0	0	1	0	0
11	Establish Secure Channel	0	**0**	0	0	1	1	0	**0**	1	0	0	0	0
12	Hide Data by Encription	1	**1**	1	0	0	1	1	**1**	1	1	0	0	1
13	Hide Data by Steganography	0	**0**	1	0	0	0	0	**0**	0	0	0	1	0
React to Attacks														
14	Alerts Subjects	0	**1**	0	0	0	0	0	**1**	0	1	1	0	1
15	Apply Institutions Policies	0	**0**	0	0	0	0	0	**0**	1	1	0	0	1
Recover from Attacks														
16	Audit Actions	0	**1**	1	1	0	0	0	**1**	0	1	1	0	1
17	Apply Institutions Policies	0	**0**	0	1	0	0	0	**0**	1	1	0	0	1

Table 5.8: Decisions scenario 4

		Scenario 4												
		GT	***GROUP G2***						***GROUP G4***					
N°	TACTICS		**CR**	S1	S2	S3	S4	S5	**CR**	S1	S2	S3	S4	S5
Detect Attacks														
1	Verify Message Integrity	0	**0**	0	0	1	1	0	**0**	0	0	1	0	0
2	Verify Storage Integrity	1	**1**	0	1	1	1	0	**1**	0	1	1	1	0
3	Maintain Audit Trail	1	**1**	1	1	1	0	1	**1**	1	1	1	1	0
4	Identify Intrusion by Signature	1	**1**	1	1	1	0	1	**1**	0	1	1	1	0
5	Identify Intrusion by Behavior	1	**0**	0	0	0	0	0	**1**	0	1	1	1	0
Stop or Mitigate Attacks														
6	Authenticate subjects	0	**0**	0	0	1	0	0	**0**	0	1	1	1	0
7	Authorize Subjects	0	**0**	0	1	1	0	0	**0**	0	1	0	0	0
8	Manage Security Information	0	**1**	0	0	1	1	1	**0**	1	0	0	0	0
9	Filter Data	0	**0**	0	1	0	0	0	**0**	0	0	0	0	0
10	Verify Origin of Message	0	**1**	1	0	0	0	1	**1**	0	1	1	1	0
11	Establish Secure Channel	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0
12	Hide Data by Encription	0	**0**	0	0	1	1	0	**0**	0	0	0	1	0
13	Hide Data by Steganography	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0
React to Attacks														
14	Alerts Subjects	0	**0**	0	1	0	1	0	**0**	0	0	0	1	0
15	Apply Institutions Policies	0	**0**	0	1	0	1	0	**1**	1	1	1	1	1
Recover from Attacks														
16	Audit Actions	1	**1**	1	0	0	1	1	**1**	1	0	0	1	1
17	Apply Institutions Policies	1	**1**	0	0	1	1	0	**1**	0	1	1	1	1

- **Scenario 1:**

  With respect the table 5.9, it is possible to note that the consensual result (CR) of group 2 managed to select seven tactics correctly and three tactics incorrectly, according to the Ground Truth. Concerning group 4, its CR presents two not selected tactics, and three tactics selected incorrectly.

  It is also possible to compare the results obtained for Precision, Recall, and Accuracy between the consensual results and the average obtained from the individual results, making it possible to appreciate that five of the six results obtained were higher through the use of consensual decisions.

Table 5.9: Scenario results 1

Scenario 1									
GROUP N°2									
		GT	CR-2	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	7	7		4	2	6	5	4
TN	Tactics correctly NOT selected	10	7		9	6	4	9	7
FN	Tactics that had to be selected but were not selected		0		3	5	1	2	3
FP	Tactics that should not be selected, but were selected		3		1	4	6	1	3
Precision			0.700	0.608	0.800	0.333	0.500	0.833	0.571
Recall			1.000	0.600	0.571	0.286	0.857	0.714	0.571
Accuracy			0.824	0.659	0.765	0.471	0.588	0.824	0.647
GROUP N°4									
		GT	CR-4	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	7	5		5	4	6	6	6
TN	Tactics correctly NOT selected	10	7		7	5	7	5	7
FN	Tactics that had to be selected but were not selected		2		2	3	1	1	1
FP	Tactics that should not be selected, but were selected		3		3	5	3	5	3
Precision			0.625	0.590	0.625	0.444	0.667	0.545	0.667
Recall			0.714	0.771	0.714	0.571	0.857	0.857	0.857
Accuracy			0.706	0.682	0.706	0.529	0.765	0.647	0.765

- **Scenario 2:**

  Table 5.10 reveals that the consensual result (CR) of group 2 managed to select four tactics correctly and three tactics incorrectly. Group 4 CR presents one not select not selected tactic and three tactics selected incorrectly. With this, the results obtained for Precision, Recall, and Accuracy between the consensual results and the average obtained from the individual results made it possible to appreciate that five of the six results obtained were higher through the use of consensual decisions, same as in the previous scenario.

Table 5.10: Scenario results 2

Scenario 2									
GROUP N°2									
		GT	CR-2	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	4	4		4	4	4	3	3
TN	Tactics correctly NOT selected	13	10		10	10	12	10	8
FN	Tactics that had to be selected but were not selected		0		0	0	0	1	1
FP	Tactics that should not be selected, but were selected		3		3	3	1	3	5
Precision			0.571	0.564	0.571	0.571	0.800	0.500	0.375
Recall			1.000	0.900	1.000	1.000	1.000	0.750	0.750
Accuracy			0.824	0.800	0.824	0.824	0.941	0.765	0.647
GROUP N°4									
		GT	CR-4	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	4	3		4	2	4	3	4
TN	Tactics correctly NOT selected	13	10		8	11	8	4	6
FN	Tactics that had to be selected but were not selected		1		0	2	0	1	0
FP	Tactics that should not be selected, but were selected		3		5	2	5	9	7
Precision			0.500	0.401	0.444	0.500	0.444	0.250	0.364
Recall			0.750	0.850	1.000	0.500	1.000	0.750	1.000
Accuracy			0.765	0.635	0.706	0.765	0.706	0.412	0.588

- **Scenario 3:**

  Concerning the table 5.11, it is possible to note that the consensual result (CR) of group 2 managed to select all the True Positive tactics; nevertheless, they have four False Positives. Concerning group 4, its CR shows that they managed to select all the True Positive Tactics; however, they have 3 False Positives. The previous results allow us to compare the Precision, Recall, and Accuracy results obtained between the consensual results and the average obtained from the individual results, allowing us to appreciate that all the consensual results were superior, being the first scenario to present this.

Table 5.11: Scenario results 3

Scenario 3									
GROUP N°2									
		GT	CR-2	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	3	3		3	0	2	2	3
TN	Tactics correctly NOT selected	14	10		11	10	12	8	10
FN	Tactics that had to be selected but were not selected		0		0	3	1	1	0
FP	Tactics that should not be selected, but were selected		4		3	4	2	6	4
Precision			0.429	0.336	0.500	0.000	0.500	0.250	0.429
Recall			1.000	0.667	1.000	0.000	0.667	0.667	1.000
Accuracy			0.765	0.718	0.824	0.588	0.824	0.588	0.765
GROUP N°4									
		GT	CR-4	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	3	3		2	3	1	2	2
TN	Tactics correctly NOT selected	14	11		10	9	8	13	10
FN	Tactics that had to be selected but were not selected		0		1	0	2	1	1
FP	Tactics that should not be selected, but were selected		3		4	5	6	1	4
Precision			0.500	0.370	0.333	0.375	0.143	0.667	0.333
Recall			1.000	0.667	0.667	1.000	0.333	0.667	0.667
Accuracy			0.824	0.706	0.706	0.706	0.529	0.882	0.706

- **Scenario 4:**

  About the table 5.6.1, it is possible to note that the consensual result (CR) of group 2 had one False Negative and two False Positive. At the same time, group 4 selects all the True Positive tactics; however, they have 2 False Positives. The previous results allow us to compare the Precision, Recall, and Accuracy results obtained between the consensual results and the average obtained from the individual results, allowing us to appreciate that like in the scenario three, all the consensual results were superior.

Table 5.12: Scenario results 4

Scenario 4									
GROUP N°2									
		GT	CR-2	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	6	5		3	3	4	3	3
TN	Tactics correctly NOT selected	11	9		10	7	6	6	9
FN	Tactics that had to be selected but were not selected		1		3	3	2	3	3
FP	Tactics that should not be selected, but were selected		2		1	4	5	5	2
Precision			0.714	0.520	0.750	0.429	0.444	0.375	0.600
Recall			0.833	0.533	0.500	0.500	0.667	0.500	0.500
Accuracy			0.824	0.635	0.765	0.588	0.588	0.529	0.706
GROUP N°4									
		GT	CR-4	$\overline{x}$	S1	S2	S3	S4	S5
TP	Tactics correctly selected	6	6		2	5	5	6	2
TN	Tactics correctly NOT selected	11	9		9	7	7	6	10
FN	Tactics that had to be selected but were not selected		0		4	1	1	0	4
FP	Tactics that should not be selected, but were selected		2		2	4	4	5	1
Precision			0.750	0.565	0.500	0.556	0.556	0.545	0.667
Recall			1.000	0.667	0.333	0.833	0.833	1.000	0.333
Accuracy			0.882	0.694	0.647	0.706	0.706	0.706	0.706

- **Subjects rationale:**

  During the experimental process execution, one of the most outstanding observations made by the conducting and monitoring group was the interaction developed between the different participating subjects, who shared their impressions, ideas, and decisions at all times. Regarding this, it was possible to appreciate collaborative work in all the groups that used the technique, complying with the established rules, and carrying out decisions in a consensual manner. This point allows us to determine that having an established technique for consensual decision-making allows the active participation of the different actors, generating ideas, increasing trust between them, supporting the transfer of information, and leveling the participants' knowledge.

Statistical analysis

To analyze the results obtained and answer this research question we establish the need to determine whether it is possible to reject the null hypothesis $\mathbf{H_{0.1}}$.

For this, it is essential to consider that in RQ5, we apply the treatment to the same subjects, comparing them in two different conditions, before and after treatment. Thus, if the results are parametric, then it is necessary to apply the Paired T-test; if not, we need to apply the Wilcoxon Single Rank Test [24] (see 5.6).

The above means that we compare the subjects decision making obtained from the average individual decisions and the consensual decision-making according to precision, recall, and accuracy metrics and then analyze the results to accept or reject the null hypothesis. Below are the analysis carried out concerning the metrics precision, recall, and accuracy in function to the analysis steps (see figure 5.4).

Figure 5.6: Shapiro-Wilk RQ5

Precision

- ***First Step*: Grouped by metric**

  To complete the first step, we present table 5.13 and figure 5.7. According to them, it is possible to realize that all consensual decisions of group G2 and G4 improved compared to the average individual decisions in each of the four scenarios. The above means that both the G2 and G4 groups' consensual decisions obtained decisions with better precision on the ground truth than individually.

Table 5.13: Precision.

		Individual decisions	Consensual decisions
S1	G2	0.608	0.7
	G4	0.59	0.625
S2	G2	0.564	0.571
	G4	0.401	0.5
S3	G2	0.336	0.429
	G4	0.37	0.5
S4	G2	0.52	0.714
	G4	0.565	0.75

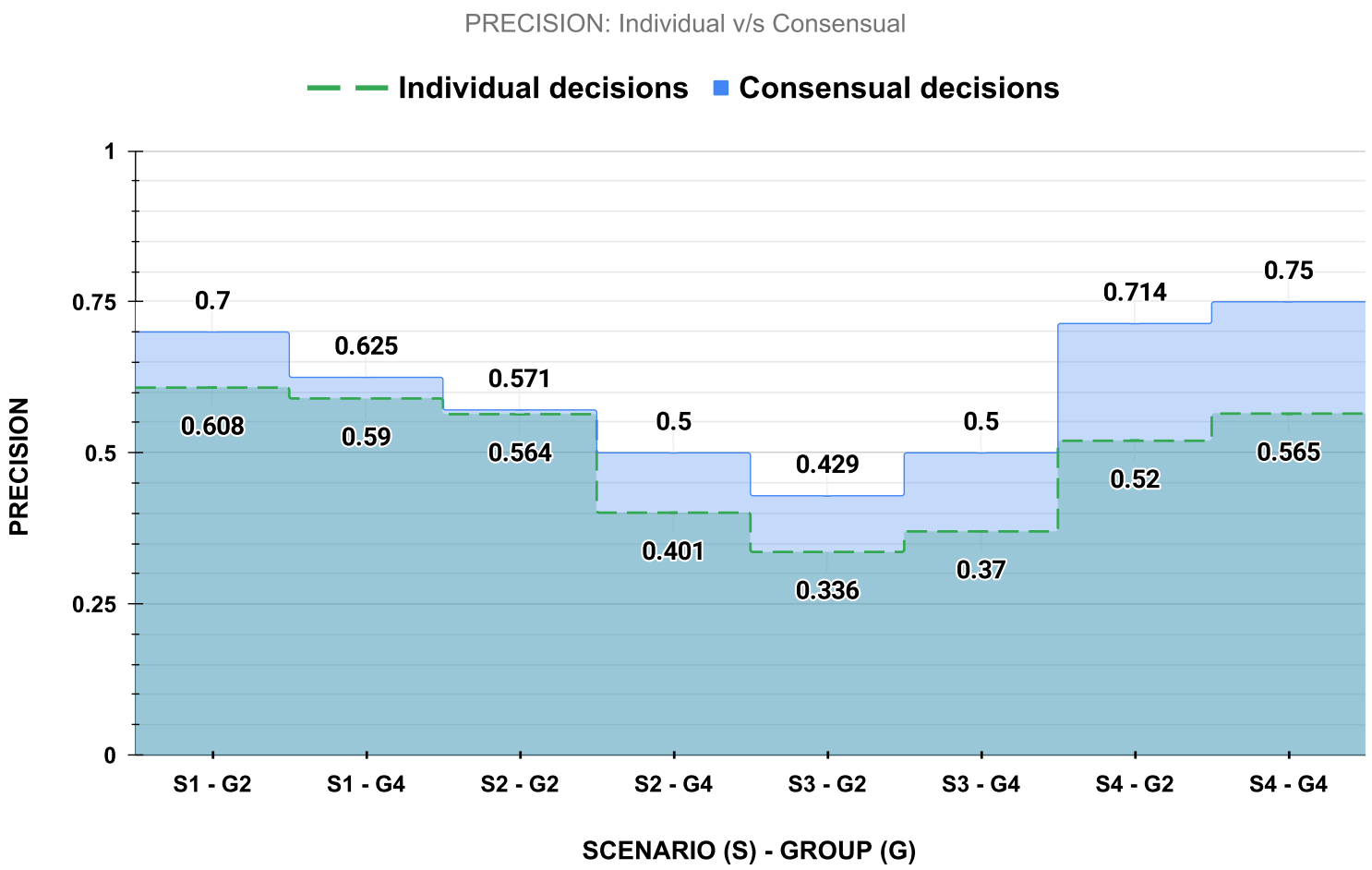


Figure 5.7: Precision G2 and G4, individual v/s consensual

- ***Second Step*: Testing normal distribution**

  The results obtained when applying the Shapiro-Wilk test allowed to determine that both the individual decisions and the consensual ones correspond to a parametric distribution.

Figure 5.8: Shapiro-Wilk test over precision

- ***Third step*: Statistically significant**

  Considering the result obtained in phase two, we can determine that we must apply the paired T-Test to get *p-value*. The result of the p-value was 0.00136, being less than 0.05, which allows us to indicate that there is a significant difference between individual decisions and consensual decisions. Therefore, it can be concluded that precision in $H_{0.1}$ is rejected.

Recall

- ***First Step*: Grouped by metric**

  About the recall results, particularly all consensual decisions from G2 improved compared to the average of individual decisions. Regarding G4, there were two instances, in S1 and S2, that the average obtained from individual decision-making was closer to the ground truth than the consensus.

Table 5.14: Recall

		Individual decisions	Consensual decisions
S1	G2	0.6	1
	G4	0.771	0.714
S2	G2	0.9	1
	G4	0.85	0.75
S3	G2	0.667	1
	G4	0.667	1
S4	G2	0.533	0.833
	G4	0.667	1

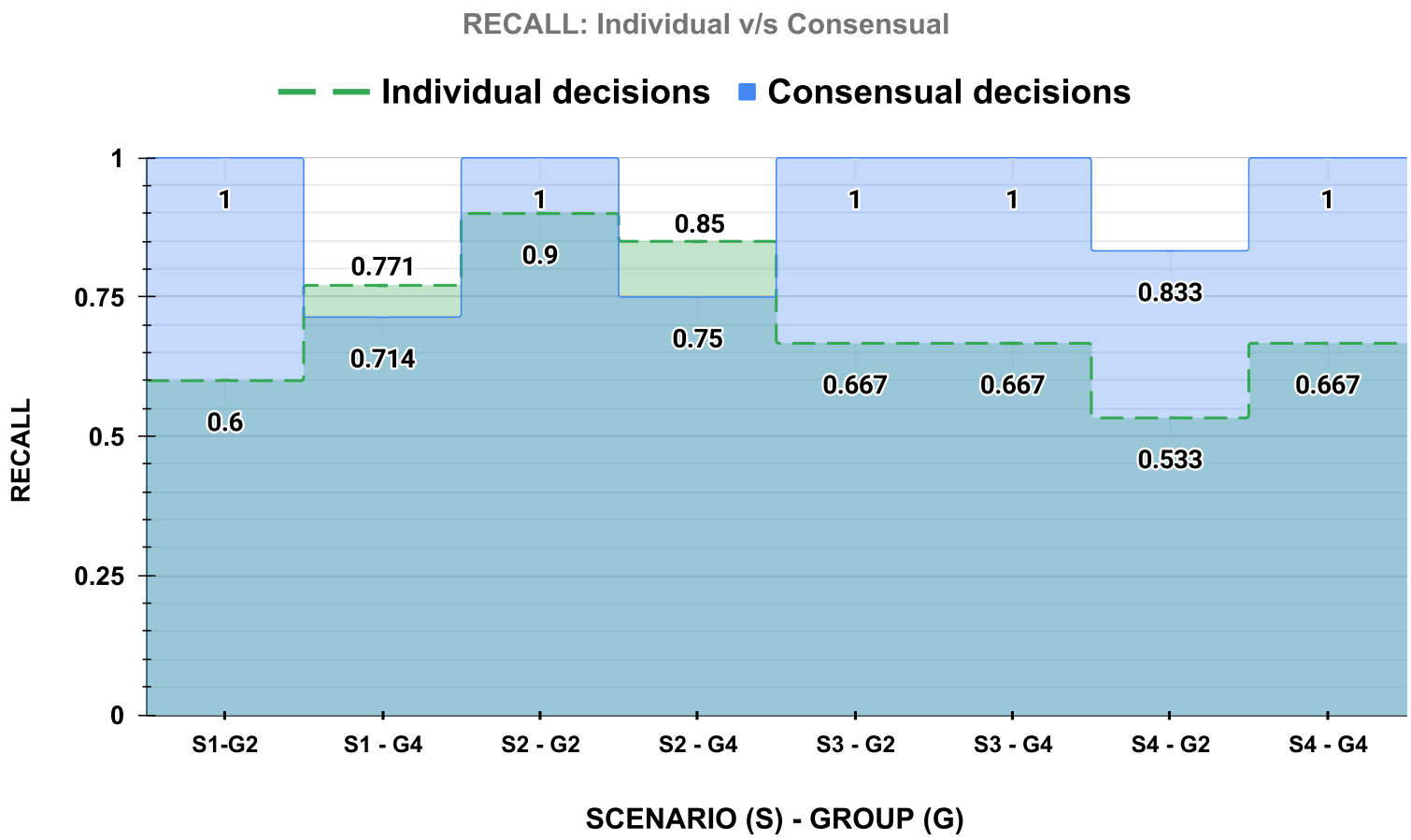


Figure 5.9: Recall G2 and G4, individual v/s consensual

- ***Second Step*: Testing normal distribution**

  The results obtained when applying the Shapiro-Wilk test allowed to determine that both the individual decisions and the consensual ones correspond to a non parametric distribution.

Figure 5.10: Shapiro-Wilk test over recall

- ***Third step*: Statistically significant**

  Considering the result obtained in phase two, we can determine that we must apply the paired Wilcoxon Single Rank Test to get *p-value*. The result of the p-value was 0.0243, being less than 0.05, which allows us to indicate that there is a significant difference between individual decisions and consensual decisions. Therefore, it can be concluded that precision in $H_{0.1}$ is rejected.

Accuracy

- ***First Step*: Grouped by metric**

  In this case, it is possible to realize from figures 5.11 that all consensual decisions were closer to the ground truth than the individual decisions.

Table 5.15: Accuracy

		Individual decisions	Consensual decisions
S1	G2	0.659	0.824
	G4	0.682	0.706
S2	G2	0.8	0.824
	G4	0.635	0.765
S3	G2	0.718	0.765
	G4	0.706	0.824
S4	G2	0.635	0.824
	G4	0.694	0.882

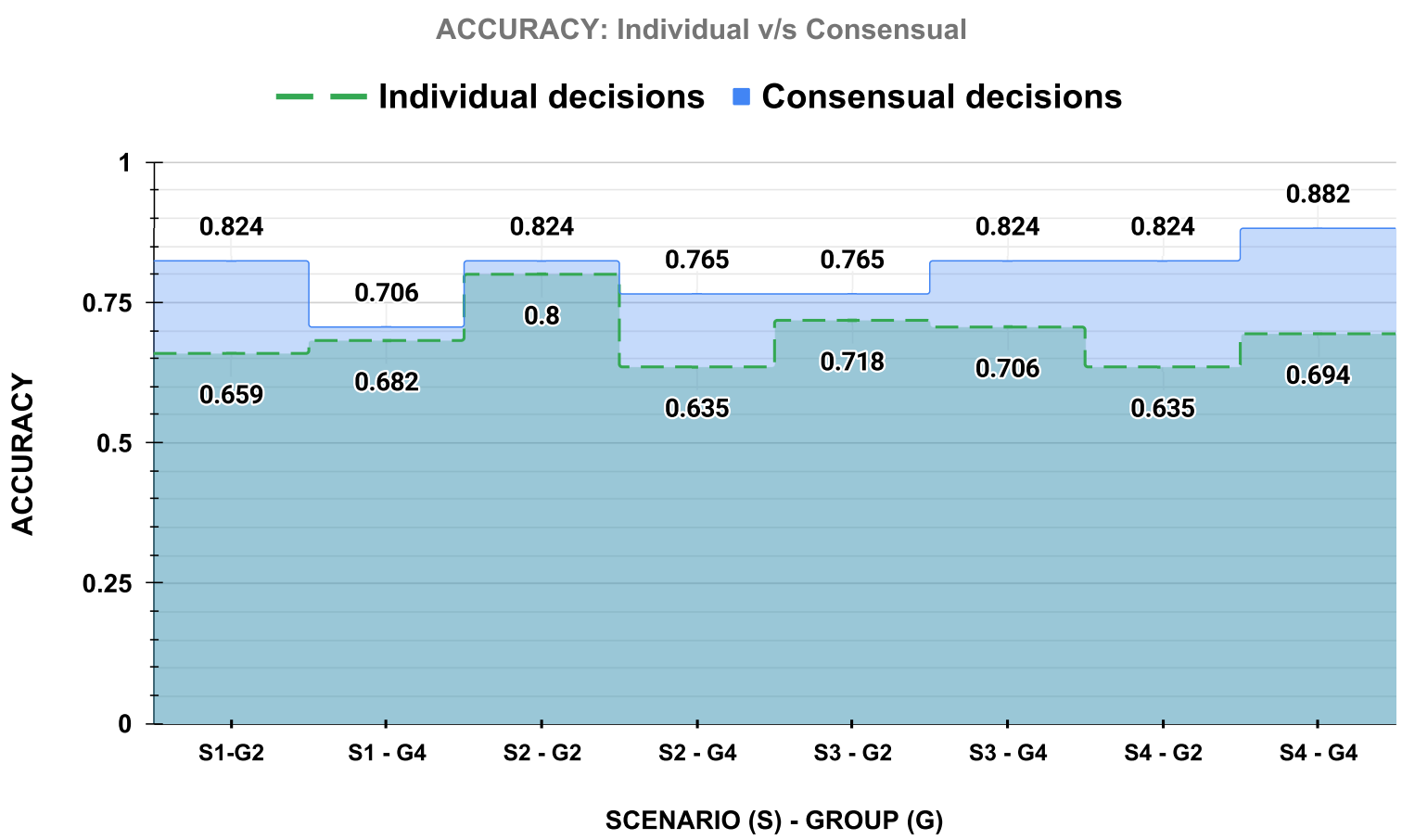


Figure 5.11: Accuracy G2 and G4, individual v/s consensual

- ***Second Step*: Testing normal distribution**

  The results obtained when applying the Shapiro-Wilk test allowed to determine that both the individual decisions and the consensual ones correspond to a parametric distribution.

Figure 5.12: Shapiro-Wilk test over accuracy

- ***Third step*: Statistically significant**

  Considering the result obtained in phase two, we can determine that we must apply the paired T-Test to get *p-value*. The result of the p-value was 0.0015,, being less than 0.05, which allows us to indicate that there is a significant difference between individual decisions and consensual decisions. Therefore, it can be concluded that precision in $H_{0.1}$ is rejected.

Analysis summary

Finally, according to the results obtained for ***precision, recall and accuracy***, it can be concluded that $H_{0.1}$ is rejected. Table 5.16 summarize the analysis executed.

Table 5.16: Hypothesis $H_{0.1}$

	RQ5		
	Precision	Recall	Accuracy
$H_{0.1}$	Rejected	Rejected	Rejected
Summary	*Rejected*		

Overall RQ5 analysis

Considering what we described regarding the scenarios, the rational subjects, and finally the analysis of the results, it is possible to establish that the TaSPer technique improves individual decision-making through consensual results closer to experts' ground truth.

5.6.2 Research question 6 (RQ6): *Having an established technique allows us to obtain results closer to the ground truth than not having it?*

Scenario analysis

Regarding the RQ6 question, we determined to answer it that we need to compare the results of selecting security tactics using the TaSPer technique versus not using it. We established two different treatments to select security tactics called "No TaSPer" and "TaSPer" due to the above (see figure 5.13). It is important to emphasize that we establish four groups, groups G1 and G3 did not use TaSPer, and groups G2 and G4 did it collecting the decision made from all the subjects.

Figure 5.13: RQ6

To obtain the results, we followed the defined experimental process, allowing determine and evaluate the results considering the measurements determined in 5.3.2 and with this establishing True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) of each group and scenario, which finally allows us to calculate **precision, recall, and accuracy**.

- **Scenario 1:** Table 5.17 presents scenario 1, where it is possible to observe the results obtained from the groups that used and not the technique. Regarding this, it is possible to see that the groups that used TaSPer obtained on average results closer to True Positive; however, it was the opposite in True Negative. Due to the above, it is impossible to determine a clear trend regarding precision, recall, and accuracy results.

Table 5.17: RQ6 Scenario 1

		Scenario 1			
		No TaSPer		TaSPer	
	GT	**GR-1**	**GR-3**	**CR-2**	**CR-4**
TP	7	6	3	7	5
TN	10	9	10	7	7
FN		1	4	0	2
FP		1	0	3	3
Total cards	17	17	17	17	17
Precision		0.857	1.000	0.700	0.625
Recall		0.857	0.429	1.000	0.714
Accuracy		0.882	0.765	0.824	0.706

- **Scenario 2:** Regarding table 5.18, it is possible to see that, on average, the results of the groups that used TaSPer, G2, and G4 obtained better results in terms of True Positive and True Negative. The above was triggered due to the lower results obtained by the G3 group, which did not use the technique, compared to Ground Truth. These results are also possible to appreciate through the precision, recall, and accuracy results obtained.

Table 5.18: RQ6 Scenario 2

		Scenario 2			
		No TaSPer		TaSPer	
	GT	**GR-1**	**GR-3**	**CR-2**	**CR-4**
TP	4	3	2	4	3
TN	13	10	6	10	10
FN		1	2	0	1
FP		3	7	3	3
Total cards	17	17	17	17	17
Precision		0.500	0.222	0.571	0.500
Recall		0.750	0.500	1.000	0.750
Accuracy		0.765	0.471	0.824	0.765

- **Scenario 3:** The results obtained in scenario 3 for groups with and without TaSPer are similar, reflecting this in table 5.19 and appreciating it in the precision, recall, and accuracy results.

Table 5.19: RQ6 Scenario 3

		Scenario 3			
		No TaSPer		TaSPer	
	GT	**GR-1**	**GR-3**	**CR-2**	**CR-4**
TP	3	3	2	3	3
TN	14	10	12	10	11
FN		0	1	0	0
FP		4	2	4	3
Total cards	17	17	17	17	17
Precision		0.429	0.500	0.429	0.500
Recall		1.000	0.667	1.000	1.000
Accuracy		0.765	0.824	0.765	0.824

- **Scenario 4:** Scenario 4 shows (see table 5.20) better results obtained for groups G2 and G4 concerning Ground Truth. The data allow us to observe that the GR3 group, without technique, obtained results less close to the GT than the other groups, reflected in precision, recall, and accuracy results.

Table 5.20: RQ6 Scenario 4

		Scenario 4			
		No TaSPer		TaSPer	
	GT	**GR-1**	**GR-3**	**CR-2**	**CR-4**
TP	6	5	3	5	6
TN	11	7	8	9	9
FN		1	3	3	0
FP		4	3	2	2
Total cards	17	17	17	19	17
Precision		0.556	0.500	0.714	0.750
Recall		0.833	0.500	0.833	1.000
Accuracy		0.706	0.647	0.824	0.882

- **Subjects rationale:**

  The most impressive thing was to observe that in the groups that did not use TaSPer, there was no process for selecting tactics in a consensual way. It was common during the experimental process for one person to control the group and generate decision-making without producing an iteration process or a verification that everyone agreed, imposing their vision instead of guiding decision-making and allowing participation of all.

Statistical analysis

This section aims to analyze the results obtained regarding the groups' decisions, using or not TaSPer, about precision, recall, and accuracy, the above seeking to determine whether it is possible to reject the null hypothesis $\mathbf{H_{0.2}}$.

Therefore, we applied two treatments, which allowed us to compare the results in two different conditions. Both treatments are independent of each other. These are of great relevance because they imply applying , if the results are parametric, apply a t-test; if not, we need to apply Mann-Whitney U test [24] (see 5.14).

The above means that we compare the subjects decision making obtained from the groups according to precision, recall, and accuracy metrics and then analyze the results to accept or reject the null hypothesis. Below are the analysis carried out concerning the metrics precision, recall, and accuracy in function to the analysis steps:

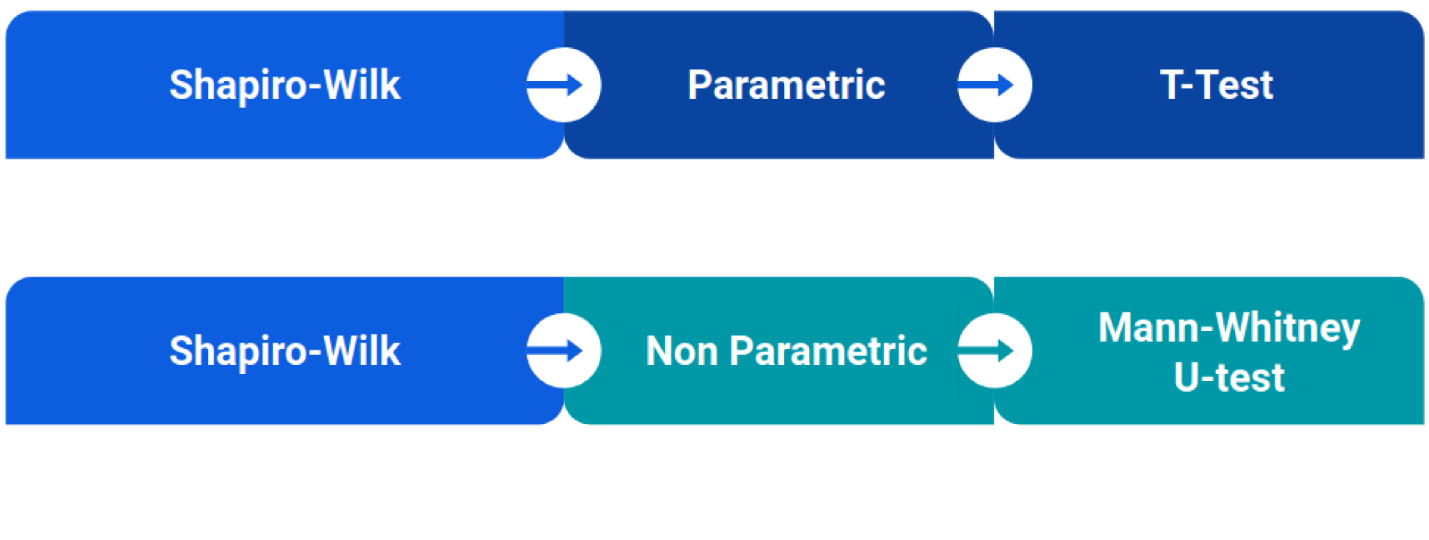


Figure 5.14: Shapiro-Wilk RQ6

Precision.

- ***First Step*: Grouped by metric**

  Regarding Precision, the mere fact of observing the graphs does not allow us to determine if there is any positive difference with the use of the TaSPer technique. It is even possible to appreciate that both groups' results without TaSPer are notably better in the first scenario, being an unexpected result. On the other hand, when analyzing the results obtained from scenarios 2 and 3, these do not allow us to determine that when using TaSPer, a critical difference is generated when not using it. Finally, in scenario 4, if it is possible to see an improvement in the results obtained by groups G2 and G4 when using TaSPer.

Table 5.21: Precision.

	No TaSPer	TaSPer
S1	0.857	0.700
	1.000	0.625
S2	0.500	0.571
	0.222	0.500
S3	0.429	0.429
	0.500	0.500
S4	0.556	0.714
	0.500	0.750

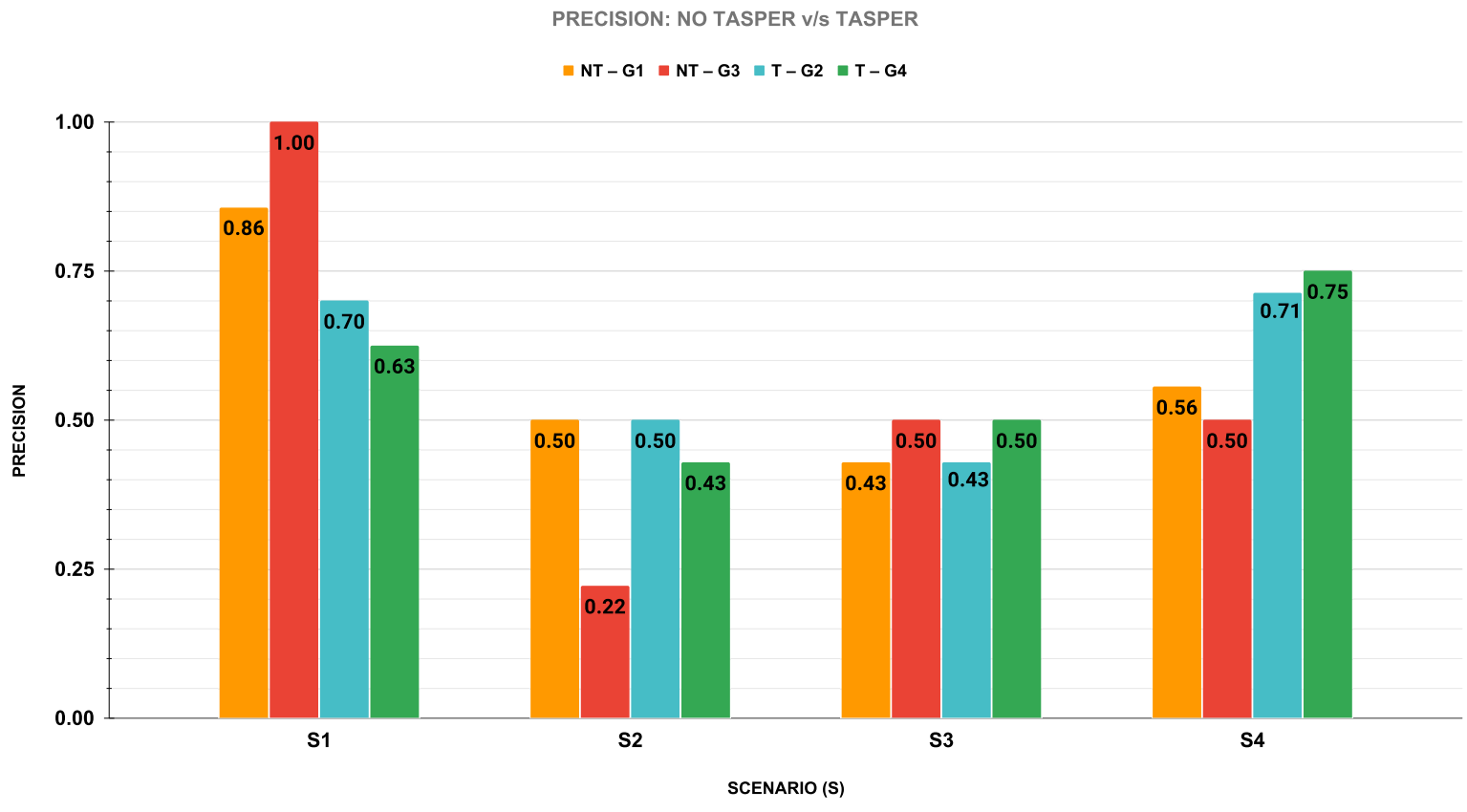


Figure 5.15: Precision, No TaSPer v/s TaSPer

- ***Second Step*: Testing normal distribution**

  The results obtained when applying the Shapiro-Wilk test allowed to determine that both the group decisions and the consensual ones correspond to a parametric distribution.

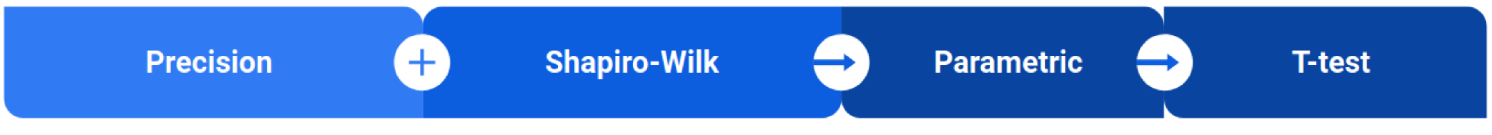


Figure 5.16: Shapiro-Wilk test over precision

- ***Third step*: Statistically significant**

  According to the results for precision, we can determine that we must apply the T-Test to get the *p-value*. The result of the p-value was 0.240652, being more than 0.05, which allows us to indicate that there is not a significant difference between individual decisions and consensual decisions. Therefore, it can be concluded that precision hypothesis in $H_{0.2}$ is accepted.

Recall

- ***First Step*: Grouped by metric**

  Regarding the results obtained for Recall, when observing the graphs, it is possible to determine that TaSPer does generate a positive result, being able to appreciate that all the results of G2 and G4 are equal to or greater than those obtained by groups G1 and G3.

Table 5.22: Recall.

	No TaSPer	TaSPer
S1	0.857	1.000
	0.429	0.714
S2	0.750	1.000
	0.500	0.750
S3	1.000	1.000
	0.667	1.000
S4	0.833	0.833
	0.500	1.000

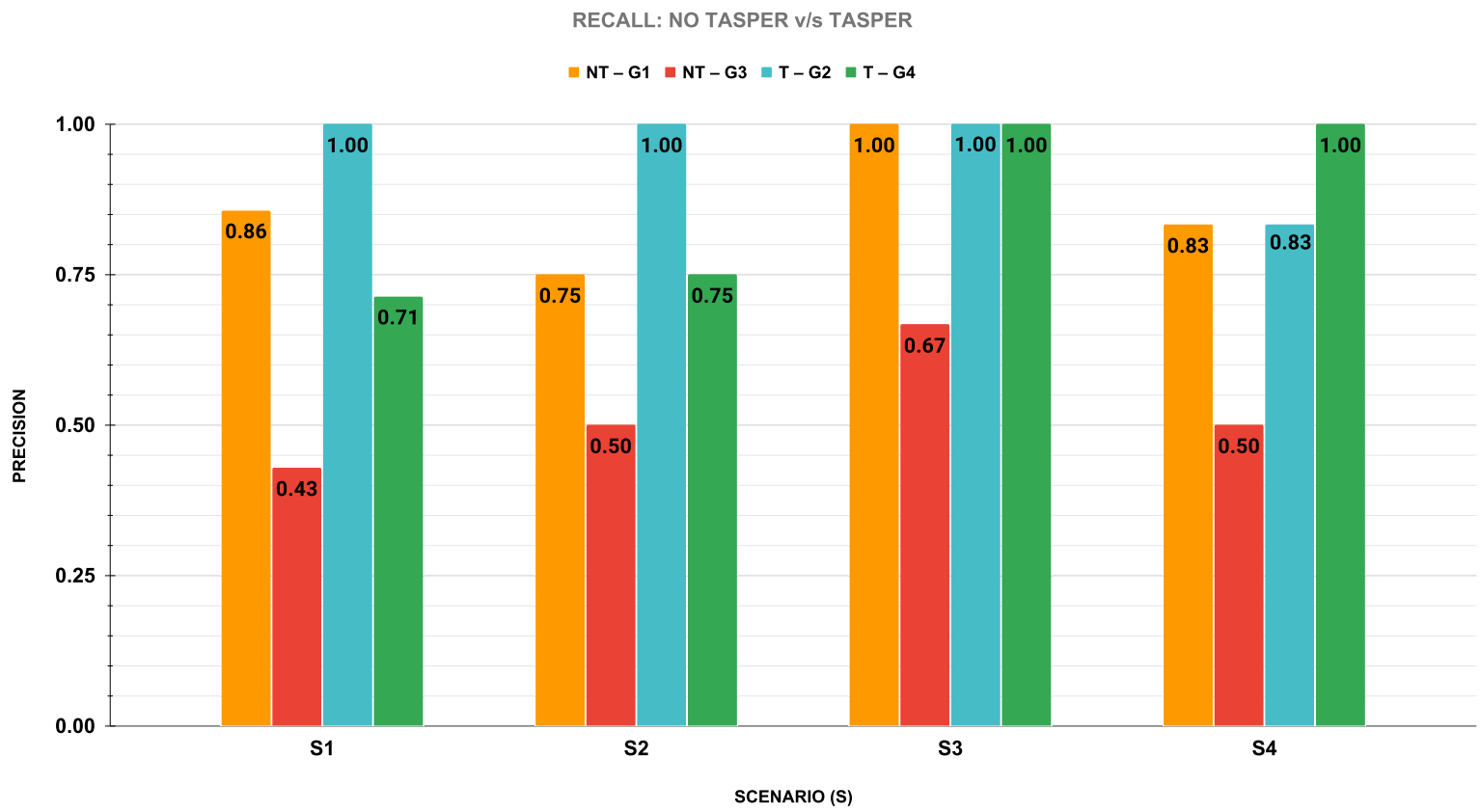


Figure 5.17: Recall, No TaSPer v/s TaSPer

- ***Second Step*: Testing normal distribution**

  The results obtained when applying the Shapiro-Wilk test allowed to determine that both the group decisions and the consensual ones correspond to a non parametric distribution.

Figure 5.18: Shapiro-Wilk test over recall

- ***Third step*: Statistically significant**

  According to the results for recall, we can determine that we must apply the Mann-Whitney U test to get the *p-value*. The result of the p-value was 0.0039, being less than 0.05, which allows us to indicate that there is a significant difference between individual decisions and consensual decisions. Therefore, it can be concluded that recall hypothesis in $H_{0.2}$ is rejected.

Accuracy

- ***First Step*: Grouped by metric**

  Regarding the results obtained for Accuracy, it is possible to see in the graph that both group and consensual decision-making achieve a high percentage of correct decisions than Ground Truth. Notwithstanding the preceding, it is possible to observe that the results obtained both with the technique and those that did not use it generate similar results between them.

Table 5.23: Accuracy.

	No TaSPer	TaSPer
S1	0.882	0.824
	0.765	0.706
S2	0.765	0.824
	0.471	0.765
S3	0.765	0.765
	0.824	0.824
S4	0.706	0.824
	0.647	0.882

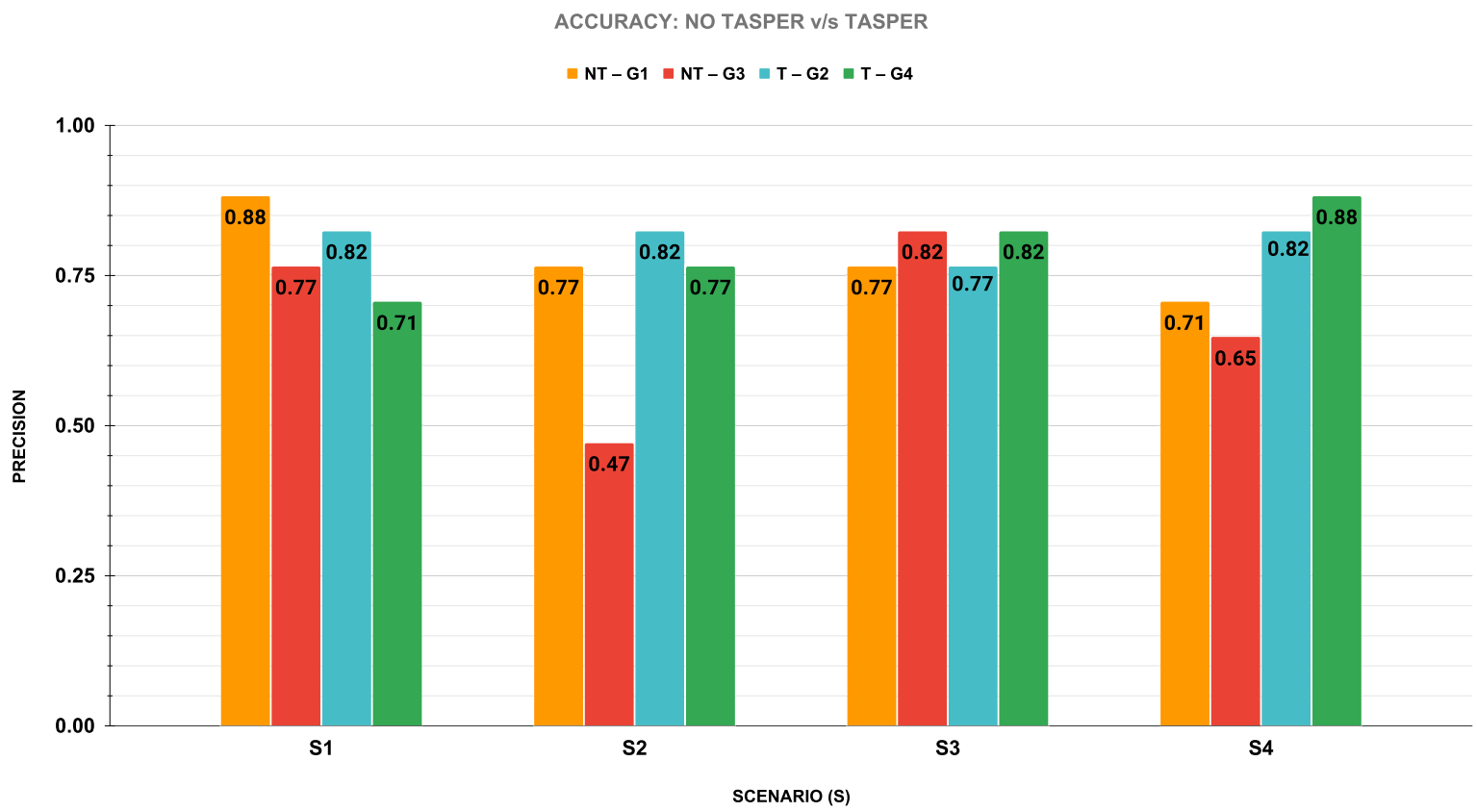


Figure 5.19: Accuracy, No TaSPer v/s TaSPer

- ***Second Step*: Testing normal distribution**

  The results obtained when applying the Shapiro-Wilk test allowed to determine that both the group decisions and the consensual ones correspond to a parametric distribution; therefore, to obtain the *p-value*, it is necessary to apply the T-test.

Figure 5.20: Shapiro-Wilk test over accuracy

- ***Third step*: Statistically significant**

  According to the results for accuracy, we can determine that we must apply the T-Test to get the *p-value*. The result of the p-value was 0.219, being more than 0.05, which allows us to indicate that there is not a significant difference between individual decisions and consensual decisions. Therefore, it can be concluded that accuracy hypothesis in $H_{0.2}$ is accepted.

Analysis summary.

To summarized all, first, we applied the Shapiro-Wilk test to determine which test should be performed over the precision, recall, and accuracy results. This to establish if the data correspond or not to a parametric distribution. Table 5.24 shows the results obtained.

Table 5.24: Shapiro-Wilk test

	RQ6		
	Precision	Recall	Accuracy
Parametric	x	-	x
Non Parametric	-	x	-

We could then determine for precision, recall, and accuracy the following tests shown in table 5.25.

Table 5.25: Testing *p*-values

	RQ6		
	Precision	Recall	Accuracy
Parametric	T-test	-	T-test
Non Parametric	-	Mann-Whitney U test	-

Finally, according to the tests made, the p-value for ***precision and accuracy*** is more than 0.05, and the ***recall*** is less than 0.05. These three results together indicate that there is not a significant difference between group decisions and consensual decisions. Therefore, we can conclude that we can not reject the hypothesis $H_{0.2}$; therefore, the hypothesis $H_{0.2}$ is accepted (see 5.26).

Table 5.26: Hypothesis $H_{0.2}$

	RQ6		
	Precision	Recall	Accuracy
$H_{0.2}$	Accepted	Rejected	Accepted
Summary	*Accepted*		

Overall RQ6 analysis

Regarding question 2, we could not establish that having a technique to select tactics allows obtaining results closer to the ground truth than not having it. Notwithstanding the above, we note that the groups without the method did not generate a collaborative process for decision-making without exchanging ideas, thereby not achieving a consensual selection. The results obtained and the hypothesis's rejection could be related to what we observed in the subject rationale and the decision-making done individually by the person who took control of the decisions.

5.7 Post experiment survey

Regarding the experimentation carried out, one of the critical factors for us was to obtain the participants' impression. For this, a brainstorming process was carried out, which was guided by the leading group. Regarding the results obtained through the feedback received by the participants, we can say that:

- Using a guide for decision making allows a better flow of information.
- The experiment allowed them to understand security tactics better.
- Another idea was to uses TaSPer to teach in the academic field.
- The interaction was significant for decision-making, and it was quite pleasant to share experiences among the participants.

- Most the subjects would recommend the usage of security tactics.
- Most of the subjects would use the security tactics in their projects.
- Most of the people knew the *security tactics* term.
- Most of the suggestions were based on the elimination of some tactics not related to software architecture eg. apply institutions policies, the enhancement of the times between activities and give better examples and details about the tactics.

5.8 Summary

The experimental sequence was vital to achieving the correct execution of the experimental study, allowing the correct preparation and training of the participants and generating the necessary spaces for the participants to be able to make decisions and analyze the results later.

Regarding the results obtained, it is essential to emphasize that the data obtained about question RQ5, allowed us to reject the null hypothesis and, at the same time, be able to establish that TaSPer improves individual decision making through consensual decision making.

On the other hand, we could not appreciate a statistically significant difference that would allow us to reject hypothesis two and thus be able to determine a difference in the results when using TaSPer. One interesting observation was that subjects who did not use the technique had great difficulty in making decisions.

Finally, the global results reveal that TaSPer had a positive effect on the subjects, because it supported architectural group decision-making and required the subjects' active participation. Additionally, since the technique is based on Planning Poker, it takes advantage of its key characteristics to achieve active participation of those involved during the development and to enable information exchange among the participants.

Chapter 6

Conclusions and future work

THIS chapter describe in the Section 6.1 the conclusions of this Master Thesis and in Section 6.2 the future work related.

6.1 Conclusion

In this thesis, we propose the Tactics Selection Poker "TaSPer," a technique for selecting software architecture tactics by consensus based on Planning Poker philosophy whose objective is to choose security tactics in a collaborative and integrated manner.

TaSPer design presents 17 cards related to the 17 security tactics described by Bass et al. [5]. Each of the cards contains five fields: number, quality attribute, name, objective, decision, and category. The procedure considers the approach of the NFR to the stakeholders. Each stakeholder will have the possibility of privately select the tactics that they deem appropriate to satisfy the security requirement. Subsequently, all stakeholders will reveal the chosen cards simultaneously, and then each subject will have the possibility to argue the reason for the selection of that card. Finally, when they have completed their participation, they can modify their decisions based on the discussion and then record the final results, performing this for each existing NFR.

To evaluate our proposal, first, we prepared a case study that allowed us to verify the following expected objectives:

- Rapid familiarization of subjects with the TaSPer process.
- Good interaction and integration of the different participants.
- Quick understanding of architectural tactics.
- Proper selection of tactics by other subjects and collecting valuable data to be used by software architects.

After the case study and considering the lessons learned, we developed an experimental study regarding the stages definition, planning, evaluating, execution, and analysis based on [2] and [42] experimental process. The main objective was to validate TaSPer to establish a technique that allows the consensual selection of security tactics and develops secure software related to agile methods. To do this, we define two research questions, the first to identify if the technique improves the individual decisions and the second to analyze the use of the TaSPer versus not using it. Those questions were:

- Does TaSPer improve individual design decisions?: We first established two groups for this question, and we applied the same treatment using the proposed technique TaSPer. Then we analyzed the average results of the individual decisions obtained for precision, recall, and accuracy, and we compare it against the consensual results. The analysis carried out on the results obtained allowed us to reject the null hypothesis, permitting us to answer the question and establish that TaSPer improves individual design decisions. Simultaneously, among the most outstanding observations is the high level of interaction between the different participants, maintaining collaborative work, and making decisions in a consensual manner.

- Having an established technique allows us to obtain results closer to the ground truth than not having it?: We established four groups for this question, two using TaSPer to select security tactics and two groups without an established technique. The results did not allow rejecting the hypothesis because they were not statistically significant. However, we observed that the groups that did not use TaSPer did not generate clear spaces for collaborative action; therefore, the decisions were not consensual.

On the other hand, the global results reveal that TaSPer positively affected the subjects supporting architectural group decision-making and required active participation. Additionally, since we base the technique on Planning Poker, it takes advantage of its essential characteristics to achieve the participants' active involvement during the development and enable information exchange among the participants.

Finally, considering the results obtained during this master's thesis, it is possible to establish that the TaSPer technique can be used as a collaborative technique for consensual decision-making to select software architecture security tactics.

6.2 Future Work

Future works are related to the "Toeska" 3.1 Software Engineering Group, under the Department of Computer Science of the Technical University Federico Santa María, especially to be able to verify the use of TaSPer with students of related courses to software development, especially focused on learning the use of security

tactics. At the same time, support the thesis in development related to the work and thesis "Evaluating Impact of Experience in Architectural Design Decision-Making Techniques" by Juan Pablo Brito.

Bibliography

[1] A. M. Alashqar, H. M. El-Bakry, and A. A. Elfetouh. A framework for selecting architectural tactics using fuzzy measures. *International Journal of Software Engineering and Knowledge Engineering*, 27(03):475–498, 2017.

[2] V. R. Basili, R. W. Selby, and D. H. Hutchens. Experimentation in software engineering. *IEEE Transactions on software engineering*, 1(7):733–743, 1986.

[3] M. Baslyman and S. Chiasson. "Smells phishy?": An educational game about online phishing scams. *APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–11, 2016.

[4] L. Bass, J. Bergey, P. Clements, P. Merson, I. Ozkaya, and R. Sangwan. A Comparison of Requirements Specification Methods from a Software Architecture Perspective. *Software Engineering Institute*, 1(August), 2006.

[5] L. Bass, P. Clements, and R. Kazman. *Software Architecture in Practice (3rd Edition)*. SEI Series in Software Engineering, 2013.

[6] K. Beckers and S. Pape. A serious game for eliciting social engineering security requirements. *IEEE 24th International Requirements Engineering Conference (RE)*, pages 16–25, 2016.

[7] F. Benavides, S. Segura, P. Trinidad, and A. Ruiz-Cortés. Fama: Tooling a framework for the automated analysis of feature models. *Proceeding of the First International Workshop on Variability Modelling of Software-intensive Systems (VAMOS*, 2007.

[8] H. Cervantes, S. Haziyev, O. Hrytsay, and R. Kazman. Smart decisions: an architectural design game. *International Conference on Software Engineering Companion (ICSE-C)*, pages 327–335, 2016.

[9] A. Chavarriaga, C. Noguera, R. Casallas, and V. Viviane Jonckers. Architectural tactics support in cloud computing providers: the jelastic case. *Proceedings of the 10th international ACM Sigsoft conference on Quality of software architectures (QoSA '14). ACM, New York, NY, USA*, pages 13–22, 2014.

[10] J. Chavarriaga, C. Noguera, R. Casallas, and V. Jonckers. Managing trade-offs among architectural tactics using feature models and feature-solution graphs. In *2015 10th Computing Colombian Conference (10CCC)*, pages 124–132. IEEE, 2015.

[11] J. Chavarriaga, C. Noguera, R. Casallas, and V. Jonckers. Managing trade-offs among architectural tactics using feature models and feature-solution graphs. *10th Computing Colombian Conference (10CCC), Bogota*, pages 124–132, 2015.

[12] M. Cohn. *Agile estimating and planning*. Pearson Education, 2005.

[13] J. C. da Silva Santos. Toward establishing a catalog of security architecture weaknesses. *Thesis. Rochester Institute of Technology. Accessed from http://scholarworks.rit.edu/theses/9004*, 2016.

[14] T. Denning, A. Lerner, A. Shostack, and T. Kohno. Control-alt-hack: the design and evaluation of a card game for computer security awareness and education. *Proceedings of the 2013 ACM SIGSAC conference on Computer and communications security*, pages 915–928, 2013.

[15] T. Dybå, V. B. Kampenes, and D. I. Sjøberg. A systematic review of statistical power in software engineering experiments. *Information and Software Technology*, 48(8):745–755, 2006.

[16] E. B. Fernandez and H. Astudillo. Should we use tactics or patterns to build secure systems. In *First International Symposium on Software Architecture and Patterns, in conjunction with the 10th Latin American and Caribbean Conference for Engineering and Technology, Panama City, Panama*, 2012.

[17] E. B. Fernandez, H. Astudillo, and G. Pedraza-García. Revisting architectural tactics for security. *Software Architecture. Springer International Publishing*, pages 55–69, 2015.

[18] J. H. Friedman. On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, 1(1):55–77, 1997. doi:10.1023/A:1009778005914.

[19] D. Gatica, G. Márquez, and H. Astudillo. Systematic selection of software components through architectural tactics. Is a relationship between tactics and NFRs possible? *XX Ibero-American Conference on Software Engineering (CIbSE)*, 2017.

[20] M. Gondree and Z. N. Peterson. Valuing security by getting [d0x3d!]: Experiences with a network security board game. In *6th Workshop on Cyber Security Experimentation and Test ({CSET} 13)*, 2013.

[21] M. Gondree, Z. N. Peterson, and T. Denning. Security through play. *IEEE Security and Privacy*, 11(3):64–67, 2013.

[22] J. Grenning. Planning poker or how to avoid analysis paralysis while release planning. *Hawthorn Woods: Renaissance Software Consulting*, 3:22–23, 2002.

[23] N. B. Harrison and P. Avgeriou. How do architecture patterns and tactics interact? a model and annotation. *Journal of Systems and Software*, 83(10):1735–1758, 2010.

[24] M.-B. Ibanez, A. Di-Serio, and C. Delgado-Kloos. Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on learning technologies*, 7(3):291–301, 2014.

[25] A. Jansen and J. Bosch. Software architecture as a set of architectural design decisions. *5th Working IEEE/IFIP Conference on Software Architecture (WICSA'05)*, pages 109–120, 2005.

[26] S. Kim. A quantitative and knowledge-based approach to choosing security architectural tactics. *Ad Hoc and Ubiquitous Computing*, 18(1/2):45–53, 2015.

[27] S. Kim, D.-K. Kim, L. Lu, and S. Park. Quality-driven architecture development using architectural tactics. *Journal of Systems and Software*, 82(8):1211–1231, 2009.

[28] S. Kim, D.-K. Kim, L. Lu, and S.-Y. Park. A tactic-based approach to embodying non-functional requirements into software architectures. In *2008 12th International IEEE Enterprise Distributed Object Computing Conference*, pages 139–148. IEEE, 2008.

[29] E. Koziolek, K. A., and R. Reussner. Peropteryx: automated application of tactics in multi-objective software architecture optimization. *Proceedings of the joint ACM SIGSOFT conference–QoSA and ACM SIGSOFT symposium–ISARCS on Quality of software architectures–QoSA and architecting critical systems–ISARCS*, pages 33–42, 2011.

[30] J. Makhoul, F. Kubala, R. Schwartz, R. Weischedel, et al. Performance measures for information extraction. In *Proceedings of DARPA broadcast news workshop*, pages 249–252. Herndon, VA, 1999.

[31] G. Márquez and H. Astudillo. Selecting components assemblies from non-functional requirements through tactics and scenarios. *35th International Conference of the Chilean Computer Science Society, SCCC*, 2016.

[32] G. Márquez and H. Astudillo. Selection of software components from business objectives scenarios through architectural tactics. *Proceedings of the 39th*

International Conference on Software Engineering Companion, pages 441–444, 2017.

[33] M. Mirakhorli. Common architecture weakness enumeration (CAWE). *IEEE Software Blog*, 2016.

[34] V. Nestler. Cyber realm card game. *Available at: http://gencybercards.com/. [Accessed 01 Jul 2020].*, 2020.

[35] R. Noel, G. Pedraza-García, and H. Astudillo. An exploratory comparison of security patterns and tactics to harden systems. *Proceedings of the 11th Workshop on Experimental Software Engineering (ESELAW 2014), ser. CibSE*, 2014.

[36] G. Pedraza-Garcia, H. Astudillo, and D. Correal. A methodological approach to apply security tactics in software architecture design. *2014 IEEE Colombian Conference on Communications and Computing, COLCOM 2014 - Conference Proceedings*, 2014.

[37] G. Pedraza-Garcia, H. Astudillo, and D. Correal. A methodological approach to apply security tactics in software architecture design. In *2014 IEEE Colombian Conference on Communications and Computing (COLCOM)*, pages 1–8. IEEE, 2014.

[38] G. Pedraza-Garcia, H. Astudillo, and D. Correal. A methodological approach to apply security tactics in software architecture design. *IEEE Colombian Conference on Communications and Computing (COLCOM)*, pages 1–8, 2014.

[39] J. Ryoo, B. Malone, P. A. Laplante, and P. Anand. The use of security tactics in open source software projects. *IEEE Transactions on Reliability*, PP(99):1 – 10, 2015.

[40] M. Thompson and H. Takabi. Effectiveness of using card games to teach threat modeling for secure web application developments. *Issues in Information Systems*, 17(3), 2016.

[41] D. Tofan, M. Galster, P. Avgeriou, and W. Schuitema. Past and future of software architectural decisions – a systematic mapping study. *Information and Software Technology*, 56(8):850 – 872, 2014.

[42] S. L. Travassos, G. H. Pfleeger and V. R. Basili. Experimental software engineering: an introduction. *1st Experimental Software Engineering Latin American Workshop-ESELAW*, 2004.

[43] J. Tyree and A. Akerman. Architecture decisions: Demystifying architecture. *IEEE Software*, 22(2):19 – 27, 2005.

[44] J. S. van der Ven, A. G. Jansen, J. A. Nijhuis, and J. Bosch. Design decisions: The bridge between rationale and architecture. *Springer, Berlin, Heidelberg)*, pages 329–348, 2006.

[45] L. Williams, A. Meneely, and G. Shipley. Protection poker: The new software security "game". *IEEE Security and Privacy*, 8(3):14–20, 2010.

[46] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. Experimentation in software engineering. *Springer Science and Business Media*, 2012.