To create a collaborative network amongst different service providers to facilitate the sharing of information and cyber threat intelligence we need to organize, maintain and evaluate trust amongst the autonomous members who have their own desire and goals to achieve that may result in conflicting interests. In this research, we operationalized the trust concept and researched the impact of trustworthiness factors on the success of a collaboration.

The research resulted in a computational trust model and a normative BDI agent based model. The BDI agent based model called N-BDI* is able to reason about the norm and behavior of members. The trust model motivates the selection of the three main trustworthiness factors, benevolence, competence and integrity. We showed that our SCTM helps the members of the alliance to select the right partner to collaborate within the situation at hand, while keeping the interaction risk to a minimum.

Ameneh Deljoo received her MSc degree in 2012 from Shiraz University. After gaining experience as a researcher in the ICT research group in the Delft University of Technology, she decided to pursue a PhD degree in 2015.This thesis is a collection of Ameneh's work between 2015 and 2019 under supervision of prof. dr. ir. C.T.A.M de Laat, prof. dr. T.M. van Engers and prof. dr. ing. L.H.M. gommans.

Computational Trust Models for Collaborative Network Orchestration   Ameneh Deljoo

Computational Trust Models for Collaborative Network Orchestration

Ameneh Deljoo

Computational Trust Models for Collaborative Network Orchestration

Ameneh Deljoo

This work was financially supported by the Dutch Science Foundation project SARNET(grant no: CYBSEC.14.003/618.001.016) and the Dutch project COMMIT(WP20.11).

Cover design by Yas Montazer.

Printed and bound by Ipskamp Printing.

ISBN: 978–94–6421–156–6

Computational Trust Models for Collaborative Network Orchestration

Ameneh Deljoo

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties
ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 20 januari 2021, te 14.00 uur

door

Ameneh Deljoo

geboren te Abadeh

Promotiecommisie

Promotor:	Prof. dr. ir. C.T.A.M. de Laat	Universiteit van Amsterdam
Promotor:	Prof. dr. T.M. van Engers	Universiteit van Amsterdam
Co-promotor:	Prof. dr. ing. Leon Gommans	Universiteit van Amsterdam
Overige leden:	prof. dr. M.V. Dignum	Umeå University
	prof. dr. S. Klous	Universiteit van Amsterdam
	prof. dr. H. Afsarmanesh	Universiteit van Amsterdam
	prof. dr. P.W. Adriaans	Universiteit van Amsterdam
	dr. P. Grosso	Universiteit van Amsterdam
	dr. C.H.M. Nieuwenhuis	Thales
	dr. ir. A. Taal	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

To my Parents

For raising me to believe that anything was possible

And

To Sam and Hassan

For making everything possible

CONTENTS

I APPENDICES

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

SPG	Service Provider Group
SCTM	Social Computational Trust Model

SDN	Software Defined Networking
ABM	Agent Based Model
N-BDI*	Normative Belief-Desire and Intention*
BDI	Belief-Desire and Intention
NFV	Network Function Virtualization
SPs	Service Providers
AI	Artificial Intelligence
MSC	Message Sequences Chart
IdP	Identity Provider
NRENs	National Research and Education Networks
STDMPs	Secure Trustworthy Digital Marketplaces
TEI	Trusted Electronic Institutions agent
GDPR	General Data Protection Regulation
RC	Recognition Context
NC	Normative Context
GDPR	General Data Protection Regulation

1 INTRODUCTION

More than 96,000 different types of attacks were reported between 2000 and 2018 [133]. Many security researchers and affected organizations try to identify and mitigate known attacks by implementing detection systems that can trigger various countermeasures. However, attackers always seem to discover new vulnerabilities and stay ahead of the game, in part by finding and exploiting weaknesses, and by sharing this information with other attackers. To decrease vulnerability to such attacks effectively, organizations could form an alliance that enables cybersecurity information sharing across organizations to develop and distribute new countermeasures. Inter-organizational cybersecurity information sharing is essential to enable organizations to protect and defend their critical network infrastructure from attacks [www.cyberthreatalliance.org/].

Establishing collaborations among different organizations has a direct impact on the defensive capabilities of all members of such an alliance [122]. A barrier to successful collaboration is trust. Collaborators need to trust other parties and find the right partner to collaborate for joint tasks. Finding a reliable partner to collaborate with is challenging because there is an inherent risk in collaborating with competitors or unknown partners. One way to manage that risk is to organize and maintain trust among the members of an alliance, where a trust model allows autonomous members to select the right partners.

The goal of this thesis is to study how trust can be organized, maintained and evaluated. It presents the development of computational trust models that can evaluate and select members that are able to collaborate in orchestrating their defensive network capabilities. We must also recognize that each member has his own desires and goals that may result in conflicting interests.

The need for the research described in this thesis emerged from severe DDOS attacks performed on the services of partners of the Systems and Networking Lab (SNE) at the University of Amsterdam in the spring of 2013. This resulted in the Secure Autonomous Response NETworks (SARNET) project that aimed to create an Information and Communications Technology (ICT) system that can respond to attacks autonomously to maintain a safe security state of the network.

The goal of the SARNET Alliance project is to research how the function and operation of the ICT systems can be orchestrated in a distributed and collaborative way, which is trusted among autonomous organizations. We assume that a group of autonomous organizations will hereto create an alliance. To create an alliance, we need to:

- Recognize and define a common benefit that no single member can achieve on its own, providing a strong incentive to join. The benefits must outweigh the risk of sharing the information.
- Define a trust framework to create and organize trust among the members in order to evaluate and reduce risk.
- Have a federated governance model to create common policies and standards for the alliance members.

As we will further motivate in section 1.2, to support the SARNET Alliance project goal, we define the main research question as: *What dynamic computational trust models enable cyber-intelligence sharing through partner selection for collaborative cyber defense operations?*

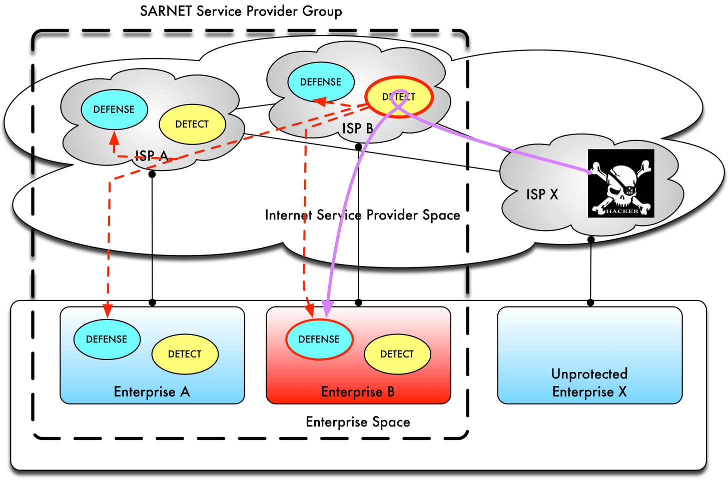


Figure 1.1: A SARNET Service Provider Group organizes a collaborative way to provide cyber security services.

Fig. 1.1 shows the SARNET alliance schema, where Service Provider Group partners organize a collaborative way to provide cybersecurity services [54]. A cybersecurity attack may come from any place within the Internet. Many defensive measures are currently placed in the channel between Internet Service Providers and the Enterprise network uplink (scrubbing, firewalling, intrusion detection, etc.). Detecting and defending against attacks near the source of the attack is potentially more scalable, increases the probability of repressing attackers by collecting forensics, avoid misuse of ISP networks bandwidth channels, etc. Internet Service Providers must however, collaborate to deliver such cybersecurity services by sharing detection intelligence and by providing a set of defensive capabilities that have been agreed within the Service Provider Group. Also, Enterprises can collaborate in the group in the same way as ISPs do. An upstream ISP, which detects an attack (red dashed arrows in fig. 1.1), may enable pro-active defensive measures in both Enterprises and possibly other ISP's. Enterprises may also detect attacks that are signaled to Internet Service Provider defensive measures. Large enterprises and ISPs will benefit from such an approach as it will save expensive first response skills, experience less security incidents because of the pro-activeness of the approach, and have the ability to file charges against criminals considering the potential ability to collect more elaborate forensic data.

1.1 RESEARCH CONTEXT

In the SARNET project, we recognize that cybersecurity problems are more than operational problems that can be addressed by an engineering team. To resolve a cybersecurity problem, a collaboration is required among all levels within and among (an) organization(s). At the start of the SARNET project we recognized that defending against cybersecurity attacks requires decisions to be made at three levels: The strategic-, tactical- and operational level (see Fig. 1.2). This sub-devision is based on a well known management approach, with examples also found in security context [127]. The strategic, tactical, and operational layers are explained briefly in the following sections. In this thesis, we focus on the strategic layer.

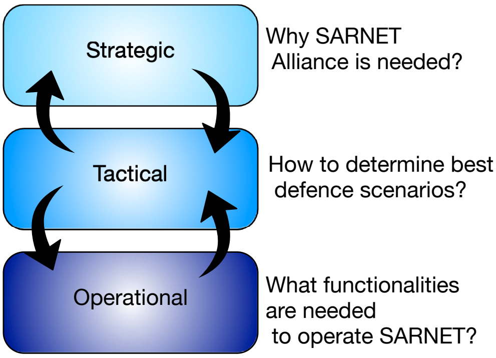

Strategic
Why SARNET Alliance is needed?
Tactical
How to determine best defence scenarios?
Operational
What functionalities are needed to operate SARNET?

Figure 1.2: Three layers of the SARNET project: strategic, tactical, operational.

1.1.1 *Strategic*

The strategic layer aims to define a set of rules and these rules are applied to guide defense strategies against cyber attacks. The defense strategies can be applied to a single organization, or to multiple organizations that collaborate with interconnected SARNETs. We developed a computational trust model that allows us to select the most trusted parties for joint tasks. Such a model can help find the best strategies against attacks and, consequently, help organizations minimize their risks.

1.1.2 *Tactical*

The tactical layer aims to investigate response scenarios that can mitigate the negative impact of an attack on a network. In an ideal situation, a network should autonomously anticipate new attacks by considering response scenarios and efficiently recover from such an attack. Determining the response scenarios depends on different factors such as network topology, cost of countermeasures, and the risk of implementing the countermeasures.

1.1.3 *Operational*

To create an autonomous response network that measures and qualifies states (or situations), by reasoning if and how an attack is developing and calculating the associated risks, controllability and flexibility from a network are expected. In this regard, cloud technology combined with optical networks, Software Defined Networking (SDN), and Network Function Virtualization (NFV) provide this flexibility to the network. The combination of SDN, NFV and cloud resources allow the SARNET to scale up or adopt resources when it is required.

Nevertheless, another responsibility of the operational layer is to provide accurate information about the network status to the tactical layer, in which this information can be used to improve the decision-making process. The operational layer will provide information such as physical inventories and virtual systems, the topology of the network, and monitoring information. The operational and partly tactical issues are researched in the PhD thesis of Ralph Koning [77].

1.2 RESEARCH QUESTIONS

During an attack, a set of functions will be selected and executed by the operational layer to mitigate and defend against the attack. At the tactical layer, the members must take decisions on how to perform the selected functions. At the strategic layer, the members must collaborate to perform the selected functions. Therefore, we must study how members can trust each other before the collaboration starts. In this research, we are specifically interested in constructing a trust model that can be applied to support the members' decision-making in selecting the right (i.e., capable and effective) partner for such joint tasks. Since the choice for the right partners to collaborate with are based on past experiences as well as estimates of future behavior, the model should include such dynamics.

This leads us to the main research question:
What dynamic computational trust models enable cyber-intelligence sharing through partner selection for collaborative cyber defense operations?

To create a computational trust model that can facilitate the decision-making process in selecting the right partner, we need to define sub-research questions that help us to answer the main research question. Therefore, the main research question is split into four sub-research questions. To create a computational trust model, we have to understand what trust means, and this leads to the first sub-research question:

- ***RQ 1*** *What does trust mean, and how can the defined concept of trust be applied in collaborative networks?*

  The research described in this thesis was done as a part of the SARNET project, in which the Service Provider Group (SPG) framework was used as a starting point. Leon Gommans [54] presented this Service Provider Group (SPG) framework as a way to create and administer the common policies. Gommans states that to have a successful collaboration, common policies and standards are needed. In this thesis, we employ the SPG framework as the governance framework for creating such common policies and rules in the context of a collaborative network. To form a collaborative network, a set of questions arises, such as *what is trust? How can trust be created, maintained, and evaluated?* In Chapter 2, we present the definition of trust and trust factors that we will use to evaluate members' trustworthiness. The common policies form the guiding mechanism for the interactions among the members. This leads to the following sub-research question:

- ***RQ 2*** *How can we model the interactions of a collaborative network?*
  This research question aims to identify the representational aspects of collaborative networks' structure and the techniques suited for modeling the interactions of collaborative networks. An introduction to the modeling techniques and representational aspects are described in Chapter 3, and we will elaborate on this when presenting the case studies in Chapter 4.
  Modeling, the interactions of a collaborative network, helps us to understand the behavior of the network.
  After modeling the interactions of a collaborative network, trust emerges from the interactions, and we need to express trust in a computational model. The members use the computational trust model to evaluate the given member's trustworthiness and up-

date their trustworthiness over time. Therefore, we define the third sub-research question as follows:

- ***RQ 3*** *How can we express trust among members in a collaborative network in a dynamic computational model?* For the reason explained before, this computational model should be dynamic in order to grasp the dynamics of the interacting members. To define a computational trust model and evaluate members' trustworthiness, we look into the different trustworthiness factors and define the following sub-research questions:

  - ***RQ 3.1*** *What are the trustworthiness factors? Do these factors have a unique impact on the trust value?*

    To construct a computational trust model, it is necessary to know which factors play a role when evaluating a member's trust. We identify three independent trustworthiness factors that potentially meet all the requirements to evaluate members' trust, namely, competence, benevolence, and integrity.
    Trust and risk are different but related concepts. In literature however, the relation between the risk and trust remains underexposed (see e.g., [31]) and most of the trust models are based upon intuitive observations. In this research, we distinguish between the risks of interactions for the collaborative network members, and trust that results from these perceived risks. This brings us to the next research question RQ 3.2:

  - ***RQ 3.2*** *What generic risk factors can be identified, and can they be evaluated by an automated process?*

    Many researchers take risk as an inherent factor of trust (see e.g., [31]). Various computational models have been proposed in literature, but only a few models explicitly take risk factors into account. In Chapter 5, we address this sub-research question and present the risk concept and risk factors, namely relational and performance risk and show how these factors can be used to select the most capable and effective partner for joint tasks.

Answering the previous sub-research questions enable us to create our Social Computational Trust Model (SCTM). We will validate this SCTM with a series of experiments, including a real networking emulation, the SARNET emulation. This evaluation in the real networking emulation, set up with Ralph Koning [77], was conducted to answer our last sub-research question:

- ***RQ 4*** *How can the computational trust model practically facilitate the selection of partners in the SARNET emulation?*
  Answering this question was central to the SARNET project wherein practical situations, partners have to be selected for taking countermeasures against attacks where these partners can only base their decisions on past and current observations of the actions of the potential allies in similar situations, weighing cost and benefits.

With the research described in this thesis, we aim to have the following scientific contributions:

- A normative Belief-Desire and Intention* (N-BDI*) agent model for reasoning about the behavior of collaborative network members.
- A simulation environment using this N-BDI* to model such collaborative networks.
- The Social Computational Trust Model (SCTM) to evaluate trust among alliances members. The validation of the SCTM model in practice is demonstrated in the operational layer of SARNET for decision-making in multi-domain defense orchestration.
- A risk estimation framework to decide on the most appropriate action.

1.3 THESIS OUTLINE

This thesis consists of seven chapters. Fig. 1.3 represents an outline of this dissertation and the relationship between the chapters and the research questions.

- A detailed description of the SARNET alliance is given in **Chapter 2**. It presents the need and requirements to create an alliance. Afterwards, the SGP framework and its role as a governance model to define common policies for the members of an alliance is given. Next, the trust definition and its components are presented.
- The objective of **Chapter 3** is to investigate an environment to model the multi-domain systems. In this chapter we present a Belief-Desire and Intention (BDI) agent model and the extension of a BDI agent model to meet the requirement of multi-domain systems.
- **Chapter 4** places the concepts of **Chapter 3** into a digital market place. The digital market place can be considered as one type of alliance, aimed at sharing cyber intelligence. We apply a normative agent based model to create a secure trustworthy digital market place. In this chapter, we outline the extended BDI (called N-BDI*) model that allows for the secure data sharing model presented.

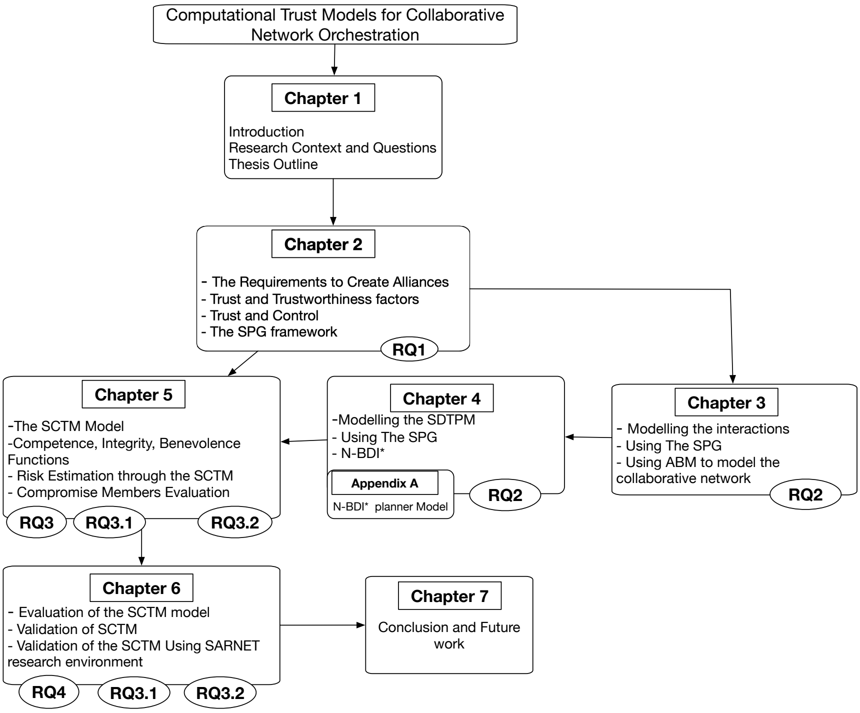

Computational Trust Models for Collaborative Network Orchestration
Chapter 1
Introduction
Research Context and Questions
Thesis Outline
Chapter 2
- The Requirements to Create Alliances
- Trust and Trustworthiness factors
- Trust and Control
- The SPG framework
RQ1
Chapter 5
-The SCTM Model
-Competence, Integrity, Benevolence Functions
- Risk Estimation through the SCTM
- Compromise Members Evaluation
RQ3
RQ3.1
RQ3.2
Chapter 4
-Modelling the SDTPM
- Using The SPG
- N-BDI*
Appendix A
N-BDI* planner Model
RQ2
Chapter 3
- Modelling the interactions
- Using The SPG
- Using ABM to model the collaborative network
RQ2
Chapter 6
- Evaluation of the SCTM model
- Validation of SCTM
- Validation of the SCTM Using SARNET research environment
RQ4
RQ3.1
RQ3.2
Chapter 7
Conclusion and Future work

Figure 1.3: Thesis outline.

- **Chapter 5** presents the proposed Social Computational Trust Model (SCTM) and its components, namely competence, benevolence, and integrity. This model is used to evaluate trust among the members of an alliance will be discussed in this chapter. Indeed, a risk estimation framework to estimate the interaction risk among the members is discussed in this chapter.
- **Chapter 6** describes the proof of concept of experiments performed with different scenarios. These scenarios are modeled as Agent Based Models that use as an internal model the SCTM model to decide on the agents to collaborate with. The experiments show how these can be modeled as real case studies in practice. The experiments were performed to evaluate trustworthiness among the members of the alliance in the situation at hand, involving different stages of relations and evidence on a given member under evaluation. This chapter demonstrates the use of two types of evidence on the given member and their combination to evaluate the member's competence, benevolence and integrity. The SCTM components were used to evaluate the trustworthiness of a given member individually. Additionally, each component's outcome allows to estimate the related interaction risks for the members of the alliance and recommend a "right" member to

collaborate. Chapter 6 also presents the implementation of the SCTM model on the SARNET emulation and the comparison between the result of the SARNET emulation and the simulation result. The work of Chapter 5 and 6 provide answers to the research questions 3, 3.1, 3.2, and 4.

- **Chapter** 7 gives a summary of the answers to the research questions, discusses the conclusions, and describes possible future directions.

1.4 PUBLICATIONS

A complete list of the author's publications is provided in page 131. The following publications are used in the body of this thesis:

- Koning, R., **Deljoo, A.**, Trajanovski, S., De Graaff, B., Grosso, P., Gommans, L., van Engers, T., Fransen, F., Meijer, R., Wilson, R. and de Laat, C., 2017, March. Enabling e-science applications with dynamic optical networks: Secure autonomous response networks. In Optical Fiber Communication Conference (pp. Tu3E-1). Optical Society of America.
  Deljoo, A. wrote the strategic and alliances sections. R. Koning. wrote the operational section, and S. Trajanovski wrote the tactical section. The remaining co-authors supervised and proofread the written work.

- **Deljoo, A.**, Taal, A., van Engers,T., Gommans, L. & de Laat C., Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners, New Generation Computing, [under review] © Springer.
  Deljoo, A. formalized, developed and performed the analytic calculations and performed the numerical simulations of the SCTM model. Arie Taal, provided guidance on the notation and formalization of the SCTM model. The remaining co-authors supervised the written work.

- **Deljoo, A.**, Koning, R., van Engers,T., Gommans, L. & de Laat C., Managing Effective Collaboration in Cybersecurity Alliances Using Social Computational Trust, Journal of Annals of Telecommunications [Under Review], © Springer.
  Deljoo, A. developed the framework and performed the simulations, derived and analyzed the results. Koning, R. assisted with the SARNET emulation measurements. Koning, R. wrote the SARNET environment that **Deljoo, A.** used to evaluate the SCTM model. The remaining co-authors supervised the written work.

- **Deljoo, A.**, van Engers,T., Gommans, L. & de Laat C., Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners, IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS), © IEEE/ACM.
  Deljoo, A. formalized, developed and performed the analytic calculations and performed the numerical simulations of the SCTM model. The remaining co-authors supervised the written work.

- **Deljoo, A.**, van Engers,T., Koning, R., Gommans, L. & de Laat C., Towards trustworthy information sharing by creating cyber security alliances, 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 1506-1510, (2018) © IEEE/ACM.
  Deljoo, A. formalized and presented the SCTM model and its components. Koning. R., provided the SARNET case study.

- **Deljoo, A.**, van Engers,T., Gommans, L. & de Laat C., The impact of competence and benevolence in a computational model of trust, IFIP International Conference on Trust Management, 45-57(2018) © Springer.
  Deljoo, A. designed the model and the computational trust framework and analyzed the result. The remaining co-authors supervised the written work.

- **Deljoo, A.**, Koning, R., van Engers,T., Gommans, L. & de Laat C., Managing Effective Collaboration in Cybersecurity Alliances Using Social Computational Trust, 3rd Cyber Security in Networking Conference (CSNet) (2019) (CSNet'19), © IEEE.
  Deljoo, A. developed the framework and performed the simulations, derived and analyzed the results. Koning, R. assisted with SARNET testbed measurements. Koning, R. wrote the SARNET environment that **Deljoo, A.** used to evaluate the SCTM model. The remaining co-authors supervised the written work.

- Koning, R., **Deljoo, A.**, Meijer, L., de Laat C., & Grosso, P., Trust-based Collaborative Defences in Multi Network Alliances, 3rd Cyber Security in Networking Conference (CSNet) (2019) (CSNet'19), © IEEE.
  Deljoo, A. wrote the SCTM model and formalized the trustworthiness factors. Koning, R. implemented and developed the SARNET testbed. Koning, R. wrote the SARNET environment and provided the defense scenarios. The remaining co-authors supervised the written work.

- **Deljoo, A.**, van Engers,T., Gommans, L. & de Laat C., What is going on: Utility-based plan selection in bdi agents, workshops at the Thirty-First AAAI Conference on Artificial Intelligence (AAAI) (2017) (AAAI 2017) © AAAI.
  Deljoo, A. designed and performed the simulations, derived the results and analyzed the results. The remaining co-authors supervised the written work.

- **Deljoo, A.**, van Engers,T., Gommans, L. & de Laat C., The Service Provider Group Framework, Looking Beyond the Internet: Workshop on Software-defined Infrastructure and Software-defined Exchanges, 2016.
  Deljoo, A. presented the SPG framework and the SPG applications. The co-authors supervised the written work.

- **Deljoo, A.**, van Engers,T., Gommans, L. & de Laat C., An agent-based framework for multi-domain service networks, In Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART'16), 290-296 (2016) © SCITEPRESS
  Deljoo, A. designed, developed, and presented the agent based model. The remaining co-authors supervised the written work.

- **Deljoo, A.**, van Engers, T. M., van Doesburg, R., Gommans, L., & de Laat, C. (2018). A Normative Agent-based Model for Sharing Data in Secure Trustworthy Digital Market Places. In ICAART (1) (pp. 290-296)© SCITEPRESS.
  Deljoo, A. developed the conceptual framework, implemented the N-BDI* simulations. Doesburg, R. provided the GDPR section that Deljoo, A. used as input for evaluating the N-BDI* model. Both Deljoo, A and Doesburg, R. authors contributed to the final version of the article. The remaining co-authors supervised the project.

2 PRINCIPLES TO CREATE AN ALLIANCE

In this chapter, we introduce the concept of cybersecurity alliances shaped by different organizations that facilitate collaboration. We explain how we define trust, control and risk that drives the social computational trust model that we will introduce in Chapter 5. We consider that the purpose of a collaborative network is to share cyber intelligence and defense capabilities. Members need to be able to select most optimal and trusted members to organise a defense in an attack with. This brings questions as:

1. Alliance members must trust each other before interacting. How can trust be created and maintained?
2. Alliance members must thereto create common policies and standards in a federated model. How can such policies be created, administered, enforced and judged upon?
3. Alliance members must understand common benefits and recognize that no single member could create them on their own. How can members maximize their benefit whilst avoiding instability of the alliance?

This chapter is based on:

- **Deljoo, A.**, van Engers,T., Koning, R., Gommans, L. & deLaat C., Towards trustworthy information sharing by creating cybersecurity alliances, 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 1506-1510, (2018) © IEEE/ACM
- **Deljoo, A.**, van Engers,T., Gommans, L. & deLaat C., The Service Provider Group Framework, Looking Beyond the Internet: Workshop on Software-defined Infrastructure and Software-defined Exchanges, 2016.

2.1 INTRODUCTION

In this chapter, we research cybersecurity alliances, where organizations can form strong partnerships to collaboratively notify each other about novel threats and protect against corresponding attacks. Many researchers have focused on sophisticated technical means to set up effective countermeasurements (e.g., event logging, correlation over new data and reasoning algorithms and anomaly detection approaches) [80]. We focus on the need and requirements to create an alliance to share cyber information. In particular, a cybersecurity alliance requires:

- a common benefit no single member can achieve on its own, providing a strong incentive to join an alliance, and encouraging members to actually share information as sharing outweighing the risk,
- a trust framework to create and organize trust among the members,
- a federated governance model to create common policies, standards for the members of the alliances.

Besides others, tackling these aspects is of paramount importance when it comes to sharing potentially sensitive and company-internal information. A well-defined trust model as a means to reduce risk helps to dispel reservations. However, since such trust relations can be complex to enforce digitally, we employ a social model to evaluate trust of different parties in the network. The network of organizations evolves over time, therefore, we need to define a more sophisticated method to select the trusted peer for sharing the information.

The common goals for the partners can be categorized from: reduce the cost and utilize the economies of scale of their organizations, to learn from the partners, or developing the new skills, sharing or minimizing of risks of collaboration, accessing to the new markets and the technologies, overcoming the political challenges, etc. These are some of the reasons that motivate parties in an alliance to join such a collaboration [7]. In this chapter, we discuss the following contributions:

1. First, we will explain the challenges in creating the cybersecurity alliance that need to be addressed.

2. Secondly, we will adopt the Service Provider Group SPG Framework [37, 54]. We use the SPG framework as a common framework to arrange trust by defining a set of common rules for the members. And, we review some of the SPG examples.

3. Thirdly, we will explain what trust means within the alliance. We will discuss the relationships between trust, risk and control in the alliance and will present our proposed management package. The detailed proposed trust model will be presented in Chapter 5.

2.2 CHALLENGES IN CREATING ALLIANCES

At first glance, as sharing of cyber threat information is beneficial to organizations, collaborations are easy to setup. However, there are several risk factors that discourage organizations from sharing information about cybersecurity incidents that they experience. These factors include:

- Competition, in particular, about their capabilities, an organization is often hesitant to share information with its competitors.
- Trust. Organizations have to rely on their partners' performance and remain vulnerable to partners' actions.
- Reputation. Public disclosure of security information often damages an organization's reputation, especially commercial organizations such as financial institutes.
- Legal context. Alliances consist of different companies with different legal frameworks as they may operate in different countries.

The ultimate goal is to design a framework under which the organizations are willing to share their cyber intelligence and cyber defense capabilities where the extent of incident information sharing among alliances' members is maximized–while the concerns as mentioned earlier and discouraging factors are sufficiently respected and taken into consideration. The mentioned requirements are addressed through this thesis.

2.3 THE CONCEPT OF TRUST

The Merriam-Webster dictionary defines trust as an " assured reliance on the character, ability, strength, or truth of someone or something" [134]. Based on this definition, we can conclude that trust appears in both personal and impersonal forms. Trust has been studied in different areas from sociology to psychology [93]. The concept of trust has been used by different theoretical frameworks [29, 110] such as:

1. Transaction cost theory, where economists studied the concept of trust from the calculative [129] or institutional [97] view,

2. Social exchange theory [15] where sociologists investigated trust in the social relations among people or institutions [59, 132],

3. Agency theory introduced by Eisenhardt *et al.* [46] where trust reduces the complexity of relations by excluding opportunistic behavior in advance [102],

4. The resource-based theory [11] that studied the impact of resources on trust in the organizations where resources provide a competitive advantage [39],

5. System theory presented by Luhmann [91] where trust in the systems (such as a collaborative network) has been studied, and

6. Attribution theory by Kelley [72], where psychologists study trust in terms of a trustor and a trustee attributes and focus upon the cognition of these attributes such as personal attributes [109, 124].

This thesis focuses exclusively on trust amongst organizations i.e. as an intra-organizational phenomenon (see e.g., Dietz *et al.* [40]). We will have a look at how both forms (the interpersonal and institutional form) are distinguished to form a collaborative network or an alliance. Trust in the alliance is always presented in every form of relations from internal relations to the relationships between the organizations. Trust among the members of alliances has been empirically demonstrated to be important for alliance formation [108]. Trust has some benefits for alliances; being a substitute for formal control mechanisms, reducing transaction costs, facilitating dispute resolution, and allowing for more flexibility [31].

2.3.1 *Personal Trust and System Trust*

In order to manage and organize trust within collaborative networks that are formed by different organizations, two different forms of trust, i.e., personal and impersonal forms of trust, have been extensively studied. Luhmann [91] defined personal trust as face–to–face interactions between two parties which help the parties to become familiar with the interests and preferences of each other. In this type of trust, the institutional arrangements are excluded from their interactions. Personal trust grows by the number of interactions between a trustor and a trustee. Undoubtedly, personal trust once established, has an important role in any business transaction [8, 91].

Luhmann distinguishes between personal trust and system trust. The system trust or institutional trust is a cognitive or rational process discriminating among individuals or institutions that are untrustworthy, distrusted, or unknown [87]. In this type of trust, a trustor cognitively chooses whom he will trust, in what context and under which conditions. The trustor makes a decision not based on his knowledge about the intentions of the trustee or his behavior, instead, the trustor will make a decision based on stable social institutions and legal systems (Luhmann, 1979; p. 34). As Luhmann and Bachmann state that these institutions reduce the risk in conferring trust [8, 91].

2.3.2 *Trust and its Antecedents*

Our aim in this thesis is to model and evaluate trust in the concept of collaborative networks. Therefore, to model trust, we need to define trust and identify trust antecedents.

In this thesis, we consider the following description given by Mayer [93] "*Trust is the willingness* of a trustor to be vulnerable to the actions of a trustee based on the expectation that the trustee will perform a particu-

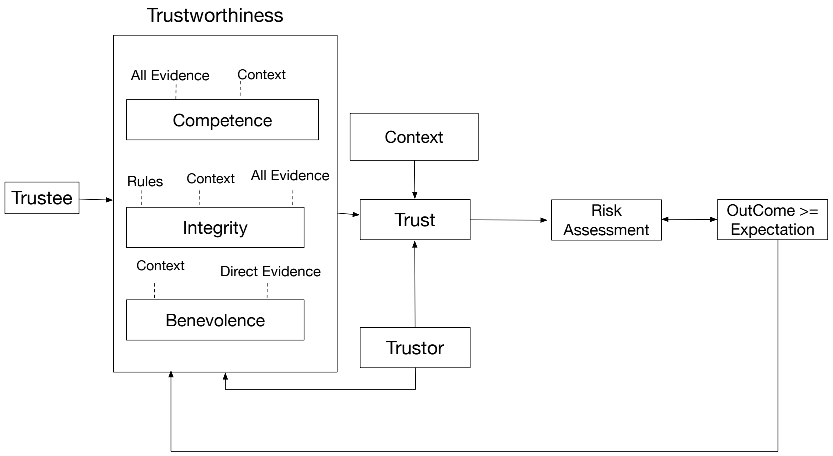


Figure 2.1: Three trustworthiness components namely benevolence, integrity and competence, are presented in this framework. We evaluate trust, by combining the outcome of these three components.

lar action important to the trustor, irrespective of the ability to monitor or control the other parties" [10, 36, 83, 93].

This expectation is realized when the given member:

- Has the potential ability to perform a given task , for example, a task could be defending against a type of attack. This is called competence.
- Adheres to a set of rules agreed upon and acts accordingly to fulfill the commitments, a commitment could be: providing cyber intelligence and adequate detail for others to act upon. This in known as integrity.
- Acts and does good even if unexpected contingencies arise.[1] This is called benevolence.

The trust framework is depicted in Fig. 2.1. Essentially, the framework says that a member is trustworthy if he has an ability to perform a task in a given situation, his integrity, and has a positive relationship with the trustor. Once trust is established, the trustor is willing to take the risk, and the outcome of the risk estimation block will serve as feedback to update the perception about the trustee's factors (i.e., competence, integrity and benevolence).

2.3.3 *Trustworthiness*

Trustworthiness and trust are two distinct terms. According to [23, 62, 74, 93] trust is a property of a trustor in respect of his relationships with a trustee, while trustworthiness is an attribute of the latter, i.e., a multi-

[1] Acts toward the interest of the alliances.

faceted approach that captures the competence of trustees and other characters of the trustees [26]. A trustworthy trustee of a collaborative network is the one that would have the highest values of competence, integrity and benevolence in a given situation, and the trustor can predict the trustee's behavior based on the trustee's competence, integrity and benevolence values in a given situation. In this respect, Hardin [62] says, "if, on your own knowledge, I seem to be trustworthy to some degree with respect to some matter, then you trust me with respect to that matter." Castelfranchi and Falcone [23] state that the trustee's trustworthiness in a given situation is objective; however, the trustor may make a mistake in evaluating the trustee in some cases, where the trustor conducted an insufficient or deficient evidence gathering on the given trustee, which may lead to misjudging of the trustor's trust. This means that the trustors deal with the perceived or evaluated value of trustworthiness, which is subjective [23].

2.4 THE FACTORS OF TRUSTWORTHINESS

As we introduced three factors of trustworthiness in section 2.3.2, we explain each of them in details in this section.

2.4.1 *Competence*

Competence, also referred to as ability, refers to the potential ability of the evaluated member to perform a given task, and is one of the trustworthiness factors. The domain of competence is varied for each domain, for example, the trustee may be highly competent in some technical domains, which shows that the trustee can perform the given task related to that area. On the other hand, the trustee may have little experience and knowledge in another area, such as interpersonal communication, which may not be able to perform the task in that area. This factor studied and applied in different areas by scholars (see e.g., Mayer, Castelfranchi *et al.*[5, 23, 86, 93]). Competence relates to a set of qualities that makes the trustee able to perform the given task. These qualities are mentioned as expertise to perform, knowledge of the domain, skills, know how to do the given tasks, self-esteem and self-confidence, interpersonal skills, and leadership [93]. A trustee that has and can prove to have all or some of the mentioned qualities is taken into account from the trustor's perception that he has the required ability to do the given task [62].

2.4.2 *Integrity*

Integrity can be defined as a trustee's commitment to the rules that a trustor finds acceptable, more generally, to a set of sound rules (see e.g., [5, 48, 93]). When evaluating the trustee's integrity considering a given commitment and situation, the trustor seeks evidence about the trustee's integrity. The related evidence must show the trustee's capacity to fulfill promises, keep consistency in his actions, and compliance with the agreed norms or rules [93]. Therefore, if the trustor finds any evidence that shows the expediency, the shallowness, and the artificiality, of the given trustee, this shall lead to lack of the trustee's integrity [27]. For example, in the collaborative network, members need to act according the agreed rules (i.e., the rules of the SPG), therefore, any deviation from the agreed rules can have an impact on the trustee's integrity.

2.4.3 *Benevolence*

Benevolence has been considered by different scholars as a key element in trusting scenarios [85, 86, 93]. Mayer *et al.* define benevolence as "a trustee is believed to want to do good to the trustor, aside from an egocentric profit motive" [93]. A number of scholars have defined benevolence as a good–will feeling towards the trustor [47], excluding any harmful intention even the trustee has an opportunity to do so [86].

This can also be defined as positive intentions towards a trustor, which is presented in the trust model by Adali *et al.* [5]. In this thesis, we adopt Mayer's benevolence definition by considering that in a collaborative network, partners are motivated to behave benevolently when they expect joint gains from their collaborations[2] [22, 85]. Harding [62] presents that in this type of trusting scenario, each member has the incentive to collaborate with the other party, if both parties take into account the interests of the interacting peer [62]. Therefore, in the collaborative networks, the members of the network will put the network's goals ahead of their individual goals.

2.5 SERVICE PROVIDER GROUP

In this section, we adopt the SPG as a framework to consider how trust and power can be organized to govern an alliance. The SPG represents a group of organizations that act together as one single business delivering a service. The SPG framework provides one or more services that none of its members could provide on their own. To

[2] In this type of benevolence, voluntary help still exists.

a user, the SPG appears as a single autonomous provider. To members, the SPG appears as a collaborative group with standards and rules that each member translates into its own conforming policies. The policies regulate the provisioning of the services and the user terms and conditions that are enforced by the group. A user signs a service agreement with a member representing the SPG. The SPG recognizes the directorated role that oversees the interactions and inter-operation of its members. Fig. 2.2 shows the schema of the SPG framework.

As we mentioned, we use the SPG as a governance framework to manage an alliance by creating and maintaining group policies and standards. In our research, the cybersecurity alliance consists of different SPG members that collaborate to share incident information and collaboratively take actions [37]. A priori identification of benefits and risks for each members' alliances is essential. This is a challenging task that needs coordination and oversight to ensure quality, and manage risk and liability. Leon Gommans *et al.* [54] describe the SPG as a way to coordinate the collaborative network activities by defining a set of rules that promotes trust among the members of the alliance. Trust inherently introduces risks as trust may be disappointed. The risks associated with information sharing and safeguarding sensitive information are reduced through the adoption of sound policies and standards. Building trust in sharing and safeguarding requires the ability to manage risk [8]. Risk decreases with sound policies and standards, increased awareness and comprehensive training, effective governance, and enhanced accountability.

Instituting the SPG is a way to establish and maintain a common set of intra-organizational policies that are translated into inter-organizational policies such that each entity knows that the policy it is authorizing is correct. In the case of the rule violation, the SGP has an enforcement component that is used to enforce the agreed SPG rules according to the objective of the group. The authors [54] made the assumptions that protocols, exchanging authorization transactions between organizations will provide enough message confidentiality, authenticity and integrity such that the security of an exchange is never disputed. In this thesis, we adopt the SPG framework as a way to define a set of common rules and establish the alliance model. The set of the SPG rules are used to monitor the members' behavior and evaluate an integrity of a member. The SPG rules are defined under the assumption and expectation that each member knows that he should behave according to these rules. In general, the behavioral variance in a society subjected to rules is smaller compared to a non-regulated one, i.e., the behavior is more predictable, therefore the risks for each society member is reduced [63]. The SPG provides a way to justify trust among members by observing the members' behavior, nevertheless, members may not act according to the rules (non-compliant behavior of a member). Within the original

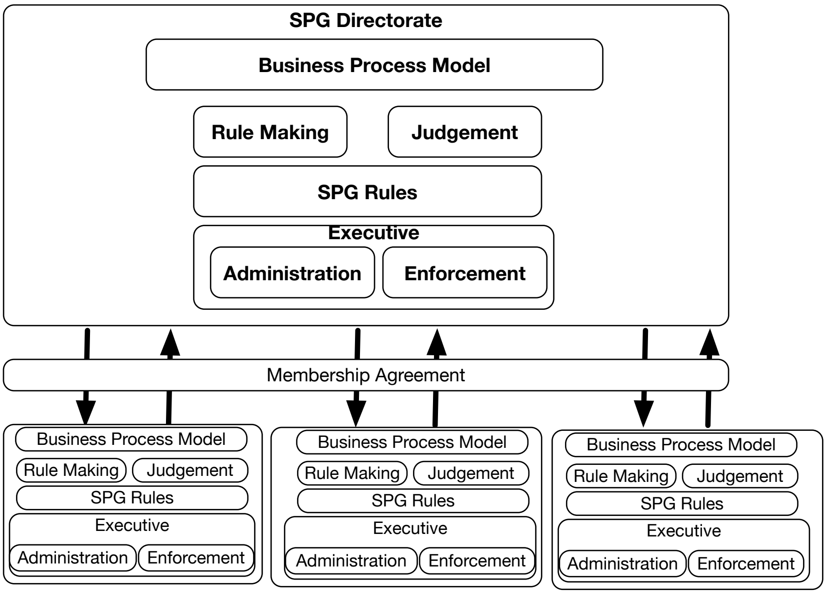


Figure 2.2: The SPG framework, which contains the rule making, rule administration, and judgment components. The members have the same component to translate and implement the common policies into their internal policies.

SPG research, the "user" is external to the SPG. In our Alliance research, the "user" is a alliance member that relies on the information provided by other alliance members. The SPG rules are used as an input for one of the components of our trust, which is presented in the following section.

2.5.1 *Application of Service Provider Group*

Two well known examples of an SPG are MasterCard and eduroam, where none of the members of these networks can deliver the services on their own. In the MasterCard example, competitive banks collaborate to provide payment card services to their customers (i.e., merchants and cardholders). An Eduroam network provides WiFi service to the members of participating research & education institutions across the globe.

Leon Gommans [55] studied these two examples in detail in his thesis. Also, competitive airlines collaborate in alliances such as Skyteam to expand their networks of destinations that can be offered to their passengers and agree on a common standard to offer services such as allowing lounge access, priority lane access, etc. Another application of the SPG model is managing and organizing the services delivered by multiple autonomous Service Providers (SPs) of the Geni testbed [13]. The Geni testbed provides the slivers delivered by different SPs and which will need coordination in order to comply with the particular agreed rules namely the service level that is required by an SP. The willingness of an SP to collaborate within a group is dependent on the benefits each provider can achieve through this collaboration. Investigating how the SPG concept can be applied to a collaboration of service provider organizations, encompasses:

- Providers can offer services that can become part of service stretching across multiple autonomous service providers,
- Such services have a common notion of service quality requirements. This is the key subject this research contributes to.

2.6 CONTROL AND RISK

Control and risk have been studied and described as one of the steps towards establishing trust [91, 93]. The theoretical bases for the control concept of the alliance governance model are derived from two sources: transaction cost economics and a key mechanism to control opportunistic behavior [108].

Risk has been defined as one of the elements of trust by different scholars [31, 93, 115]. Das *et al.*[31] and Mayer *et al.* [93] presented two types of risks (i.e., rational risk and performance risk) that are involved in the process of creating and managing trust among the members of the alliances. We will explain these types of risks in Chapter 5.

The alliance governance model has been proposed to prevent partners from abusing the alliance by taking advantage of opportunist possibilities. Adequate legal and ownership safeguards, detailed contracts, equity investments, and strict rules agreed between the partners are examples of governance models. We use the SPG as a governance model to present the adequate standards and policies for the members. The SPG framework is also responsible to coordinate the activities of autonomous members of the alliances. This model also describes the need to monitor members to detect unwanted behavior such as opportunism, abuse or fraud. The SPG framework has been introduced in previous publications [36, 37], where we investigated the role of the SPG in defining the set of common rules for the alliances by detecting and applying the consequences of undesirable behavior in an alliance (e.g., violation of the agreements[3]).

We use the basic concepts of the SPG and combine them with the trust and risk literature to explain how these factors impact the structure of

[3] We use the SPG as the input to evaluate the integrity of the given member (or trustee)

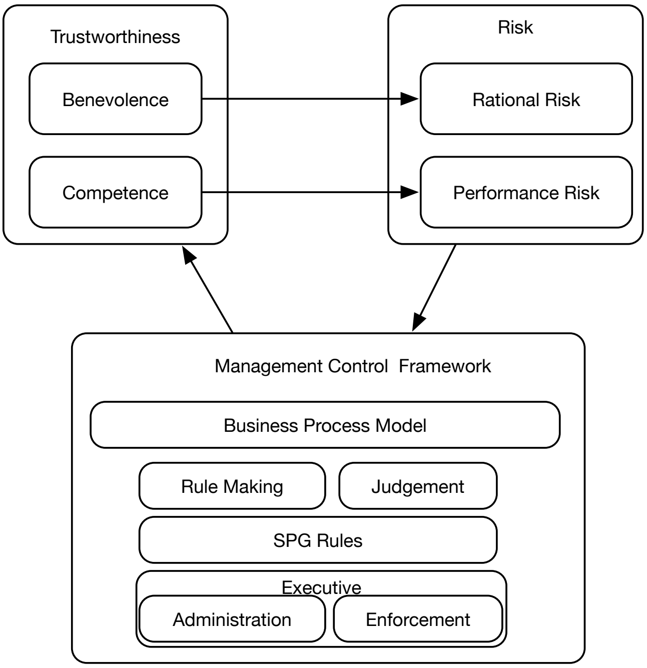


Figure 2.3: Management control package. At the bottom of this Figure, we have the SPG framework as the governance model responsible for defining the SPG rules and policies of the alliance. In the risk module, we estimate the relational and performance risks based on the benevolence and competence components. The SPG rule is used to evaluate the given member's integrity and monitor the given member's behavior.

the governance model and the control mechanisms in alliances. The proposed model of relations that we will use is presented in Fig. 2.3.

CONCLUSION

In this chapter, we described the concept of collaborative networks as a way to share cyber intelligence, cyber defense capabilities, and effective collaboration decisions. The overall aim of this approach is to help organizations to share critical information on security incidents among trusted parties and increase the efficiency of defending against attacks. Information sharing is crucial for the members to give insights into ongoing attacks, new malware and detected vulnerabilities. We presented the requirements to create a cybersecurity alliance.

The inter–organizational trust definition and trust components consist of three different components competence, integrity, and benevolence, where we identified them beside the trustworthiness of the trustee to evaluate the trustee's trustworthiness. Therefore, it is important to estimate the trustee's trustworthiness by considering each of these three dimensions individually and dynamically combining them by considering different situations and stages of relations described in Chapter 5. However, most of the computational trust approaches evaluate the trustee's trustworthiness as a black box and do not consider different trustworthiness dimensions, we will review some of the recent computational trust models in Chapter 5.

Our computational trust model is based on the multidisciplinary literature on trust [8, 23, 93], describing the estimation of competence, integrity, and benevolence of the trustee under evaluation. We will present the computational model of the presented trust framework in Chapter 5. We used the SPG framework as a governance framework to establish and maintain a common set of inter-organizational rules. Therefore, we present the governance model of the alliance by combining the SPG concept with a qualification of trust and risk.

3 MODELING A COLLABORATIVE NETWORK

In this chapter, we introduce an Agent Based Model (ABM) to study how an alliance, as an open system[1] can be modeled as a collaborative network. We identify the requirements and tools to model the collaborative network by describing the interactions in a collaborative network. Using the SPG concept (see Chapter 2), we introduce a methodology for the acquisition of the computational model of the SPG and its transformation into an ABM. Our research methodology to answer *RQ*2: (i.e., "How can we model the interactions of a collaborative network?") is as follows: First, we analyze interactions in the network at the signal layer, i.e., the message exchange between actors, and model them with the components of the BDI agent architecture. In the next step, we identify actions, intentions, and conditions necessary for the interactions to occur. These steps are required to model the Eduroam case study, that is covered in this chapter, and the Secure Trustworthy Digital Marketplaces (STDMPs), which will be presented in Chapter 4.

This chapter is based on:

- **Deljoo, A.**, van Engers,T., Gommans, L. & deLaat C., An agent-based framework for multi-domain service networks, In Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART'16), 290-296 (2016), © SCITEPRESS.

3.1 INTRODUCTION

A priori identification of benefits and risks to stakeholders that collaborate to provide a service across multiple service domains is a problem that depends on the goals, benefits and capabilities of multiple service provider networks involved in producing the services. In this chapter, we describe and demonstrate the modeling of the benefits and risks identification mechanism in such an open system using an ABM. Agent based modeling is an intuitive way to model a collaborative network where members are self-governed autonomous entities [1].

Open systems have an intuitive mapping onto an ABM. An open system consists of rational, cooperative and autonomous agents where each of them has its own goal to achieve. Agents present specific roles

[1] open systems in which countably many members may leave and join the network at run-time.

in this system and interact with others as a means to accomplish their goals. The ABM may, therefore, offer a way to investigate the benefits and risks for collaborating autonomous agents. Modeling such systems receives considerable attention from both the Artificial Intelligence (AI) and the communications network communities [1, 43]. A service provider network is an example of an open system. Service provider networks are composed of competitive service providers that see benefits in collaborating. It is important to note that in such networks, each member cannot provide the requested services on its own and collaboration provides benefits such as reduced cost or increased revenue.

The SPG [54] is a way to describe multi-domain collaborations. The SPG framework presents a way to model multi-domain service provider networks and can be used to describe the structure of such a collaboration. Eduroam [128] is a good example of such collaborations. We take the eduroam confederation as an example of an open system, which consists of multiple autonomous agents, where each of them has their own goals and intentions to collaborate. For instance, providing authenticated eduroam WiFi access to visiting students is an example of a campus IT service that a single university is unable to provide on its own without collaboration with other universities. In this chapter, Eduroam is used as a simplified example of the SPG framework.

ABMs can be used to represent the collaborative behavior in SPGs. An ABM provides mechanisms to allow organizations represented as agents to advertise their goals, negotiate their terms, exchange rich information, and synchronize processes at a high-level of abstraction [106]. Considering the autonomous behavior of the SPG, a comprehensive model for an ABM must be able to express the global goal and the requirements of the domain in a distributed way. In this chapter we will present an ABM for a multi-domain service provider network. We demonstrate the transformation of a sequence of inter-agent interactions into intra-agent characterizations.

3.2 METHODOLOGY

The method we used to create an ABM model of an SPG is as follows: First, the case study was analyzed at the signal layer, i.e., we identified all the events and their messages; then we visualized them by using a Message Sequences Chart (MSC), see Fig. 3.2. Next, we integrated the signal layer with an internal behavioral characterization and a specification. Consecutively, we constructed an agent layer where we addressed the intentions of agents. For this purpose, we defined the mental objects and events from a BDI perspective, maintaining the relationship between components. Finally, we embedded an institutional layer to consider the normative aspects. It should be noted that our conceptual

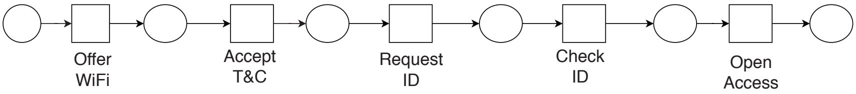


Figure 3.1: Example of occurrence: a WiFi connection instance happens in four steps: university **offers** WiFi access to students and staff members who are registered in the eduroam identity database (check ID). For brevity we use the example of a student using the eduroam system to gain internet acces, but the model applies to anybody affiliated with an educational or research institution participating in the eduroam system. The student **accepts** the terms and conditions of this free service. The university **requests** an identity from the student. The student **provides** a valid identity. Finally, the university will **deliver** the service (free WiFi).

framework covers several realities: physical (message exchange), mental (social network), and institutional (normative relations).

3.3 EDUROAM AS A SERVICE PROVIDER GROUP

Eduroam [128] allows students, researchers and staff from participating institutions to obtain wireless Internet connectivity (WiFi) across the campus and when visiting other participating institutions. Eduroam allows participating research and education institutions, known as an eduroam SPG, to provide internet access for students and staff from any other participating institute. The participating institutes act as an Identity Provider (IdP). In the following section, we describe the scenario of eduroam[2]. Fig. 3.1 shows a successful eduroam connection between the service provider and the student (users).

Eduroam works as follows: An university offers WiFi access to students and staff members who are registered in the eduroam identity database. The student accepts the terms and conditions of this free service. The university requests an identity from the student. The student provides a valid identity. Finally, the university will deliver the service (free WiFi). A successful WiFi connection is a fundamental cross domain transaction. Consequently, what the case study describes is a collaboration among the SPG (university) and their users (the student), which is just one of many other possible scenarios.

[2] eduroam is a federated roaming service that provides such secure network access by authenticating a user with their own credentials issued by their IdP. To reduce the complexity of the case study, we only consider one Service Provider (a university) and students who are willing to use this free service. A group of National Research and Education Networks (NRENs) are providing this service for the institutes organized by TERENA [54, 116, 128].

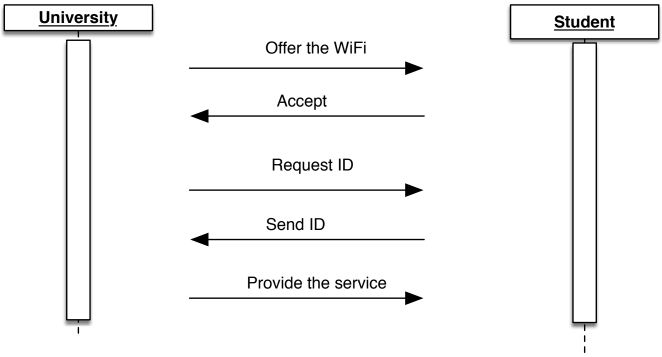


Figure 3.2: Message sequence chart of eduroam WiFi service.

3.3.1 *Signal Layer*

In order to initiate the modeling, first we look at the speech acts (message sequence) of agents and all the events to illustrate the first layer of our framework (called the signal layer). As a first definition, we may consider a scenario as a chain of events which represents the flow of the scenario from a beginning to a conclusion (from this perspective, we can consider all of the occurrence events as communications acts, such as messages going from a sender to a receiving party). In addition, when the sender performs an action, this action is coupled with an acknowledgment by the receiving party. The eduroam service delivery process is basically characterized by the actions: offer, accept, request, provide and open access. These actions are performed by one of the parties as a represented in Fig. 3.1. This process is protected by an eduroam confederation agreement. To provide the access, the student or the university needs to perform the required action such as providing the credential, therefore, when the required action is not performed from the student side, the university can enforce on a failed action. Fig. 3.2 illustrates the eduroam scenario.

Accepting the terms and conditions allow students to connect to the eduroam WiFi. To be registered as a student is one of the preconditions for a successful authorization step. Both being in the WiFi range and holding a valid ID are examples of necessary conditions for completing the process by this scenario. Such necessary conditions are in general associated with the ability or, more generally, to the power of the agent, in a specific context ,i.e.,"agent + environment". These kind of necessary conditions must be satisfied in any framework where agents need to perform an associated action.

3.3.2 *Message Topology*

The message topology of the eduroam scenario, as presented in 3.3 and 3.4 is based on the collection of messages between parties, namely students and universities. In Fig. 3.3 on the right side, the small boxes are responsible for message queues and the lines show the direct communication among the parties. In Fig. 3.3 on the left side, the dashed lines refer to actions that have important results besides the direct communication. For simplicity, we only consider two possible representations of the message topology (direct and indirect communication).

A message topology depicts the interactions in the network, and shows the distribution of signals over the agent-roles (see Boer *et al.* [18]). In our approach, the message topology facilitates the identification of a certain agent-role for an agent, which is shown in the MSC (see Fig. 3.2). This part of the research has been inspired by an "actor model" defined by Hewitt *et al.* [64].

In order to take all the events and their side-effects into account, Silno [121] introduced an explicit world actor, "disjointing the sender from the receiver", which we also adopt in this work and is presented in the right side of Fig. 3.4. The world would play as an intermediary component among the agents. In the eduroam case study, the world plays the role of an IdP.

3.4 AGENT PERSPECTIVE

In the previous section we presented the message exchange in eduroam WiFi access. We started from a representation of the eduroam WiFi access scenario on the MSC chart and we present it with Petri nets patterns [107].

Next we added the identified agent-role descriptions to this representation. Such roles are associated with certain beliefs, plans (resulting in actual actions) and goals. From the agent point of view, the precondition and *ex-post* intentional explanation of the scenario contributes to plans that agents want to perform. A possible result of this explanation is presented in Fig. 3.5. Therefore, externalized intents have been considered as the events that trigger the processes of offer open access between a university and a student. Then results of the actions are presented at the end of the MSC chart. In this case study, we know that the university usually accepts a student's requests for using the WiFi service, once their identities has been approved by the IdP. And at the third step, we apply the critical grouping method to show the conditions, which are necessary for that action to be happened.

To summarize, we assume that: (a) the student performs an evaluation of the offer (evaluation action), (b) the student accepts the offer if it is acceptable for him (terms and conditions), and (c) the student

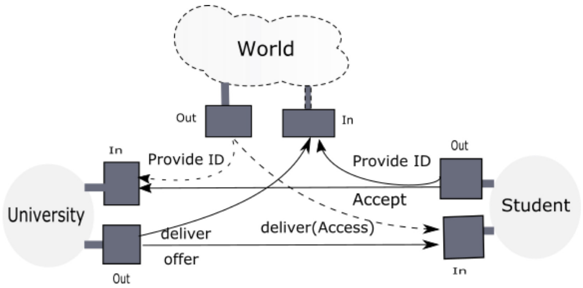


Figure 3.3: Indirect communications in the eduroam scenario. This case happens when the student requests the service from a guest university, therefore, different parties and transactions between the student and the university is required to provide the service to the student.

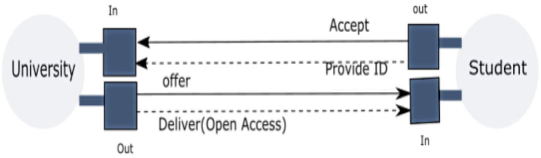


Figure 3.4: Direct communications in the eduroam scenario. In this scenario, the student is in the home university and the home university will check the ID and provide the service.

provides the ID (the university provides access) if the student owns the requested information (a valid identity or ownership condition). The MSC diagram in Fig. 3.5 depicts the eduroam access scenario. This MSC diagram shows the current activities in the eduroam scenario, wherein the vertical lines present the messages exchanged between parties, resulting in a successful WiFI access if all required conditions are satisfied. In the following sections, we introduce some patterns to be attached to the flow of the scenario. Instead of using just one visualization with the MSC diagram as shown in Fig. 3.5, we provide alternative representational models with the Petri nets (see Fig. 3.6).

In our ABM model, we refer to four layers, each of which addresses specific components:

- the *signal layer* represents acts, side-effects and failures (e.g., technical failure, user abuse) such as outcomes of actions,
- the *action layer* represents actions (or activities) such as performances intended to bring about a certain result,
- the *intentional layer* represents intentions such as commitments to actions, or to build up intentions, and

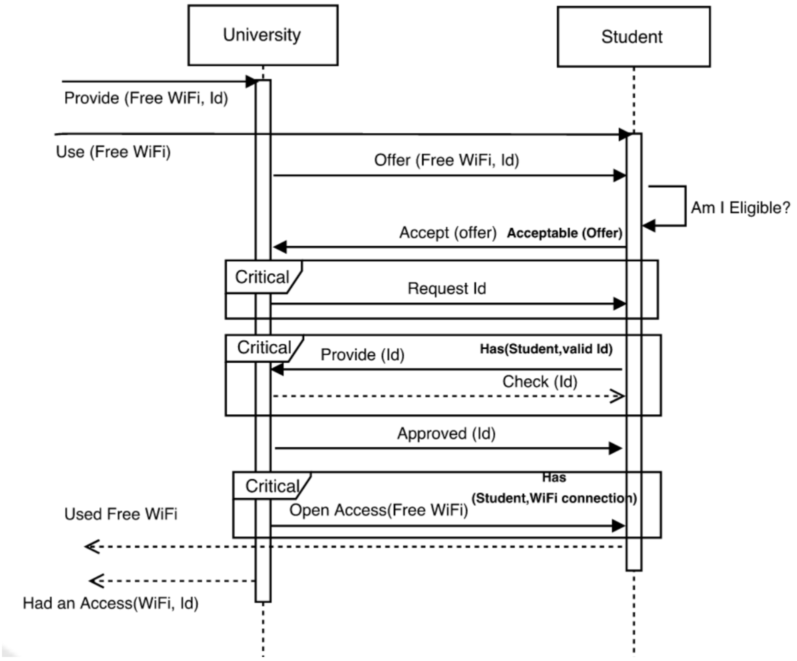


Figure 3.5: MSC of the eduroam scenario with intentions and critical conditions.

- the *motivational layer* represents motives such as events triggering the creation of intentions.

The last three layers compose the agent layer. The closure of the sensing-acting cycle of the agent is guaranteed by the fact that certain signals, when perceived by agents, becomes the motivation for action. In our framework, motivation refers to conditions that make the agent sensitive to a certain fact, which becomes the motive for starting an action. Motivations however often remain implicit in the scenario (see Sileno *et al.* [121]).

3.4.1 *Institution*

In general, we can say that an institution is an intentional social collective entity (see [17]), defined by certain rules and some institutional facts. It is collective and intentional, simply because a group of people recognize its existence. A complementary view on institutions has been presented by Searle *et al.* [118] and in Searle [119]. In this complementary view the concept of an institution unifies games, (social) informal norms and legal norms. Terms like “university" and “student" denote agent roles acting within the free WiFi institution. However, there is a difference between the actual participants and the role that they play.

Agent-role models were first introduced by Boer *et al.* [18] with the purpose of representing scenarios of compliance and non-compliance elicited from legal experts. In this work, an agent-role links the concepts of an institutional role, and an intentional agent. In practice, we add characteristics to the role that are important factors according to the constructed normative theory [121], and we describe its behavior by using an intentional approach.

We start considering only the core functions (events and acts) related to the agent's role (a student or a university). This process proceeds by using a common knowledge interpretation to define the intention of an agent. Then, the analysis of intentions allows us to reconstruct the goal process, setup plans and actions to reach the goal.

Following the description given by the eduroam WiFi agreement, a student with a valid ID assigned to him by the IdP, may expect access to the WiFi at any participating institute (university) that acts as an IdP when he accepts the offer of that IdP. Therefore, institutional roles are defined by actions in order to achieve certain goals, in the eduroam case the goal is providing the WiFi to the Staff and students. Furthermore, we observe that the possibility of a WiFi connection exists because there is a university who has offered and received acceptance and finally delivers the service. Both roles are strictly necessary; there cannot be a student without a university in the case of free WiFi. In this line of thought, a WiFi connection does not concern only one university with its own students. Eduroam connection is a free WiFi service for all students all over the world and composed by several competing parties (e.g., campuses).

Although, it is not explicitly present in the formal description of the internet connection for different institutions, the presence of the IdP and technical partners is obviously not negligible for the institutional role. These relations are involved in the evaluation of the offer (having a power[3] to offer as a university) and the action, which is meant to judge the acceptability of the offer (being in the WiFi range and holds the valid ID). Evaluation however, is not made explicit in the definition of the free WiFi process. A complete scheme about the process can be drawn by unifying procedural and institutional descriptions, which is shown in Figs. 3.7 and 3.6. The university acts as an offeror and a student acts as an offeree. The gray circle in Fig. 3.7 shows that the performance of the action is not sufficient to proceed, but it has to return a positive result.

[3] This power (see [44]) is a concept derived from Hohfeld [65] that is separated from the social reality concept ability.

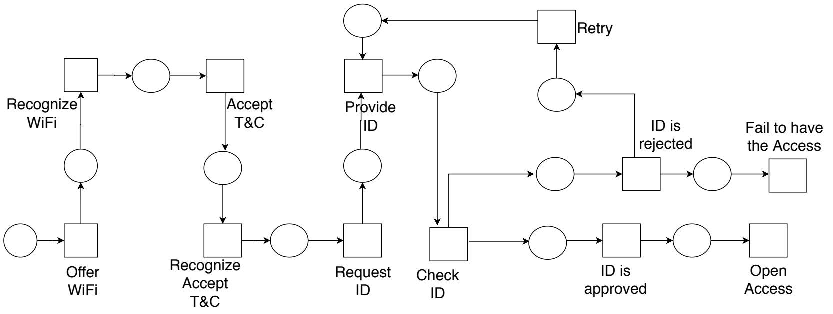


Figure 3.6: Full flow of events associated to an Eduroam WiFi connection between a student as the offeree and the university as the offeror. The flow represents the detailed events between two parties. Also, the flow contains the failure events and the actions to respond to the unsuccessful events.

3.4.2 *Visualization*

We use Petri nets to visualize the flow of the scenario [51]. Petri nets is one of the languages for mathematical modeling to describe a distributed system. The Petri nets is known as a directed graph, in which the nodes in the graph represent transitions and edges show the event that may occur. Therefore, we chose the Petri nets to model the eduroam case study.

3.4.3 *Implementation*

For the discussed model in Fig. 3.6, we have implemented the eduroam case study in the Jadex platform [20]. Jadex is the BDI extension of the Java Agent DEvelopment Framework (JADE). The Jadex environment is a general-purpose development environment for creating and implementing multi-agent system applications, allowing the implementation of the agents with reactive (event-based) and deliberative (goal-driven) behavior. As a proof of concept, we implemented the eduroam scenario. Each student has a unique ID that needs to be validated through the registered university and if the students try to login with a wrong ID, the access will be denied and the university is informed about the wrong attempt. The aim of implementation is to identify the required interaction between different parties in the eduroam network. Fig. B.2 in Appendix B shows screen shots of the implemented eduroam case study in the Jadex framework.

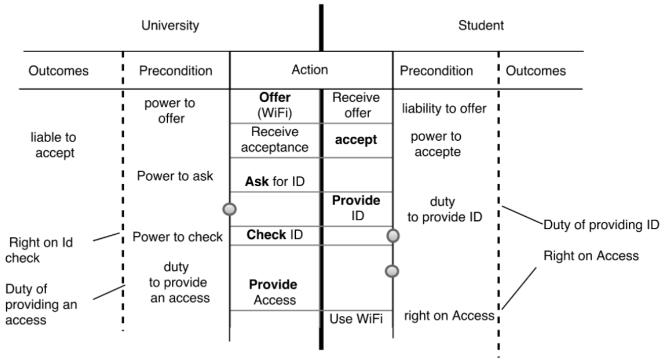


Figure 3.7: Full-action pattern associated to an eduroam WiFi connection. In this figure, we define the actions, preconditions and the outcome that are associated with the actions for the student and the university. The Perti nets of this table is provided in Fig. B.1 in Appendix B.

3.5 CONCLUSION

In this chapter, we presented four layers (i.e., signal, intentional, motivational and action) which are used to model a specific usecase. These four layers allows for reasoning about policies, beliefs, intentions, and are an essential step to provide an ABM framework to describe the behavior of collaborating partners in an SPG.

Our research is intended to model a collaborative network and associated normative reasoning in a completely distributed environment. In particular, we are interested in how to model an SPG from the normative perspective to observe the agent's (member) behavior and identify the benefits and risks. In the current approach, typical strategy decision problems for a given scenario do not take explicitly into account the possibility that the members avoid a rule,or forcing the interpretation of the rule toward their interests if the regulator (consciously or not) left some ambiguity. Using our framework, agent models or roles involved in a social scenario, outlined from a scenario can be described. As an operative result, such a simulation can help to understand the social (institutional) dynamics: validating the domain of conceptualization of the experts, making predictions, suggesting improvements to regulations for the SPG framework and spotting normative weaknesses and vulnerabilities.

4

A NORMATIVE AGENT-BASED MODEL FOR SHARING DATA

In this chapter, we focus on how norms can be used to create so-called Secure Trustworthy Digital Marketplaces (STDMPs). A secure and trustworthy data-sharing infrastructure build according to the STDMP-architecture can be used in various application domains. A good example of such application domains is the health sector, where hospitals, third parties and data analysts try to find the most effective interventions based on patient data. The STDMP will help the stakeholders to protect their interests and to prevent data protection infringements.

Norms guide the behavior in social systems and govern many aspects of individual and group decision-making. Various scholars use agent-based models for modeling such social systems; however, the normative component of these models is often neglected or relies on oversimplified probabilistic models. Within the multi-agent research community, the study of norm emergence, compliance and adoption has resulted in new architectures and standards for normative agents. As the problem we work on is on collaboration between network partners, we must include normative reasoning within the agent architecture to be able to simulate such behavior using agent-based models (ABMs). For this reason, we have developed the Normative Belief-Desire and Intention* (N-BDI*) framework, an extension of the well-known BDI agents. As we will show in this chapter, the framework enables us to construct normative agents that allow us to model social systems and use them as a basis for studying the effects of norms on a society of agents. The N-BDI* provides us more the insights to answer *RQ*2 (“How can we model the interactions of a collaborative network?").

This chapter is based on:

- **Deljoo, A.**, van Engers,T., Gommans, L. & deLaat C., A Normative Agent-based Model for Sharing Data in Secure Trustworthy Digital Market Places, In Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART’18), 290-296 (2018), © SCITEPRESS.

4.1 INTRODUCTION

Norms[1] play a key role in the functioning of societies of agents, such as human groups, teams, and communities, as they are ubiquitous but invisible forces governing many societies. Bicchieri [14] describes human norms as: “the language a society speaks, the embodiments of its values and collective desires, the secure guide in the uncertain lands we all traverse, the common practices that hold human groups together".

A normative agent refers to an autonomous agent who understands and demonstrates normative behavior. Autonomous agents are able to reason about the norms that they are associated with, and may occasionally violate them when they conflict with each other or if these norms conflict with the agent’s interests [90]. For individual agents, reasoning about social norms can easily be supported by many agent architectures. Dignum [41] defines three layers of norms (private, contract, and convention) that can be used to model norms within the BDI framework. Agent-based models (ABMs) are frequently used for analyzing and reasoning about the actions and interactions of members forming such societies of agents that are bound by norms. However, creating realistic large-scale models of social systems is impaired by the lack of good general-purpose computational models that can be used to study an example of a norm-governed behavior of social systems.

A real-world social scenario where these concerns apply is in business relationships. In our research, we are focusing on environments in which agents may agree on collaboration efforts, involving specific interactions during a certain time frame. This environment is regulated by the applicable generic social and legal norms as well as specific norms agreed upon by the collaborative agents. Agents may represent different business units or enterprises, that for instance come together to address new market opportunities by combining skills, resources, risks, and finances that no partner can achieve on its own [42].

Any collaboration activity requires trust between the involved partners. When considering open environments, previous performance records of potential partners may not be accessible. In this chapter, we introduce a specific agent-role that is called the Trusted Electronic Institutions agent (TEI). This agent’s primary role is to regulate and control the interactions between agents. The TEI agent implements a coordination framework to facilitate the establishment of contracts and provide a level of trust by offering an enforceable normative environment. The TEI agent encompasses a set of norms regulating the environment. An

[1] Norms play an important role in open artificial agent systems; they have been said to improve collaboration. As in real-world societies, norms provide a way to achieve social order and raise expectations thus controlling the environment and making it more stable and predictable.

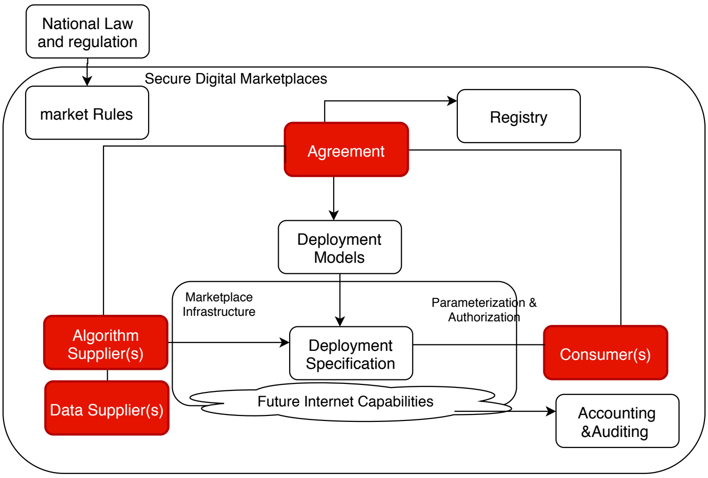


Figure 4.1: The STDMP architecture. The members of the STDMP (data suppliers and data consumers) are collaborating to achieve common goals, such as increase value, improve the accuracy of AI algorithms. The members of the STDMP create and maintain data exchange contracts to exchange data for a particular purpose under specific conditions such as access rights, using AI algorithms to analyze a set of a dataset for a specific purpose ((see [136]).

important role of the TEI agent is to monitor and enforce, through appropriate services, both predefined, institutional norms and the contractual norms that result from a negotiation process. Agents rely on the TEI agent to monitor the contractual commitments of the parties involved in the transactions of the collaborative network.

Previously, we presented the elements of a normative architecture for modeling a collaborative network (see Chapter 3 [32, 38]). This chapter describes an extended BDI architecture for constructing and simulating normative effects on a social system such as STDMPs. Fig. 4.1 shows the STDMP architecture; the color boxes are in the scope of this chapter. Therefore, our research aims to create a general-purpose ABM and simulation system for studying *how norms can be used to create STDMPs and how we can monitor the effects of norms on such system where members of the society are self-governed autonomous entities and pursue their individual goals based only on their beliefs and capabilities [56]*. We have called this ABM framework N-BDI*, which stands for Normative Belief-Desire and Intention.

This chapter presents a study showing the relative contribution of social norms on creating STDMPs-supported collaborations. We used N-BDI* simulation to predict the impact of norms on the members of

STDMPs. This N-BDI* model and the simulations based upon it that you can run to study an STDMP members' behaviors, is described in detail in Section 4.2. Section 4.3 presents this STDMP scenario and the N-BDI* implementation. We illustrate the mechanisms of N-BDI* based simulation using a case study that is about the acceptance of partners' requests to share data, using the STDMP. Such sharing is only allowed if the request meets the requirements of the General Data Protection Regulation (GDPR). How compliance with GDPR is checked will be presented in Section 4.4.

We conclude the chapter with a comparison of our approach with related work on normative agents and alternative normative architectures.

4.2 N-BDI*

In this section, we present an extension of a BDI agent, called a normative BDI* (N-BDI*) framework. Our N-BDI* framework is inspired by the nBDI framework presented by Criado *et al.* [28]. Their framework, like ours, is an extension of the basic BDI agents. The nBDI framework consists of two functional contexts: the Recognition Context (RC), which is responsible for the norm identification process, and the Normative Context (NC), which allows agents to consider the norms applicable in the context of their decision-making processes.

One of the differences of our N-BDI* framework compared to the nBDI framework by Criado *et al.* [28] is the way an agent selects the most appropriate plan that fulfills the expectations of the agent. We accomplish this by integrating probabilities and utility into the BDI agent's planner component. In the N-BDI* framework, an agent has the ability to select the most appropriate plans based on the highest expected utility that fulfills the expectations of the agent (see Algorithm 1. In the nBDI framework, the authors did not consider the utility in the agent's planner.

The second difference between the nBDI framework and the N-BDI* framework is that in the nBDI framework an agent's intention is equal to an agent's action. Whereas in our N-BDI* framework, the intention is separated from an agent's action. The reason to separate intention from the action component is that intention, i.e., the commitment to execute a certain plan, should include compliance checking. Whether an agent first prioritizes the plans before checking if these plans are compliant with the norms the agent is bound by or checks compliance before prioritizing plans is irrelevant. Agents may decide to only commit to compliant plans, but also norm-violation, i.e., non-compliant behavior, is also possible. In the latter case, the agent would weigh the potential sanctions of non-compliant behavior against the gains. This separation is also used when we reason about the integrity of agents; a topic will be addressed in Chapter 5. The third distinction between these two frameworks is that we provide a function that can retrieve norms directly from the agent belief sets.

In the N-BDI* framework, after selecting a plan, the agent intends to execute that plan. In our architecture, before executing, an agent checks its (institutional) *power* to execute the selected plan. This power (see Doesburg *et al.* [44]) is a concept derived from Hohfeld [65] that is separated from social reality concept *ability*. The latter is referring to social power, i.e., the ability to achieve something in social reality. A simple example illustrates the difference between these two subjects. While in a university building, smoking is not allowed, i.e., nobody has the institutional power to smoke, whereas one has social power, i.e., the ability to do so.

The *ability* to execute the plan consequently is in social reality, and can be checked by the agent by monitoring the effects of the selected action(s) and comparing these effects with the intended effects. To note that actions may fail. Monitoring and diagnoses of actions are, e.g., addressed in the work of Boer *et al.* [19]). The explicit distinction of institutional powers and social abilities are not addressed in the nBDI framework.

Belief revision in nBDI is based on the received feedback from the environment. While in our extended N-BDI* framework, the belief revision happens with a higher frequency. In the nBDI framework, belief revision takes place every time after an action is taken, while in the N-BDI* framework belief revision takes place, first when a plan is selected and secondly when an action is taken.

Criado *et al.* [28] consider two types of norms, Constitutive and Deontic norms. In our work, we have a formalized norm representation based upon the work of Hohfeld [44], as this representation allows for a more granulated interpretation of norms. This representation formalism has shown to offer some useful features for normative reasoning (see Van Doesburg *et al.* [45]). The norms from GDPR (General Data Protection Regulation) described in this chapter are represented in this way.

In Fig. 4.2 we depict the N-BDI* framework. First, we present a deliberation cycle that allows an agent to take applicable actions permitted by the norms. The deliberation cycle of the N-BDI* framework using this initial control loop is presented in Algorithm 1.

The belief set B represents the agent's mental state and encodes the agent's knowledge about the world, i.e., what the agent holds to be true. The agent will act upon this, assuming they continue to hold. All observations O, including norms perceived, i.e., observed by the agent, are also included in the belief set by the revised function. We can separate the norms from the belief set by applying the $getNorms$ function. Goals correspond to the agent's "desires" and commitments

Algorithm 1: The Modified control loop of the normative BDI agent (N-BDI*), where O= set of observations, B= Belief set, G= Goal, P= Plan set, N= Norms, and A_p= Actions.

```
N=Norms, Input {Observe}
while True do
    /* While agent is alive and can observe the environment.
       */
    $O, Norms \leftarrow Observe$
    if $O \neq \emptyset$ then
        $B \leftarrow Setup(O)$;
        $N \leftarrow getNorms(Norms)$;
        $G \leftarrow GenerateGoal\ (B)$;
        /* Generate goals based on the belief set.            */
        $P \leftarrow GeneratePlansetP(g)|g \in G$;
        $U \leftarrow CalculateU(G, p)|p \in P$ ;
        $Pref_p \leftarrow Select(p, U)$;
        $B \leftarrow Revise(Pref_p)$;
        /* Revise the belief set of the agent based on the
           preferred plan.                                    */
        Generate $A_{Pref_p}$;
        /* Generate an action for the preferred plan.         */
        Result: Execute $A_{Pref_p}$;
    end
end
```

to plans with "intentions". So intentions are commitments to carry out certain plans in order to pursue certain goals. When intentions turn into actions, by executing those plans, agents expect certain effects of their actions. Consequently, executing plans come with new beliefs about the future world.

According to Algorithm 1, the agent has a set of plans P, where each is primarily characterized by goal G and a set of possible actions A_p. Each plan consists of an invocation, which is the event that the plan responds to and may contribute to G.

The N-BDI* framework belief set –B– contains the norms and observations. Based upon the belief set, the agent sets up the agent's G. For every goal, $g \in G$, a plan is selected and added to the plan set P. For every plan, $p \in P$, the agent calculates its utility using its knowledge about the world reflected in its belief set, its goals, and plans. The plans are then updated with their utility, after which the preferred plan is being selected $Pref_p$. Executing that plan requires the updating of the belief set B and taking the actions A_{Pref_p} that comes with that preferred plan.

Algorithm 1 allows an agent to execute plans with the highest utility. Executing the action may cause norm violations. Algorithm 2 describes how an action that is generated by Algorithm 1 can be assessed for norm compliance.

Inspecting P to find all the action recipes which have among their effects on G. The agent will examine the permission (power) to execute A_p. When an agent wants to execute A_p, it first checks the pre-condition of norms and if these pre-conditions are satisfied, then it will execute the A_p. After the execution of the A_p, the post-condition of that specific norm will be realized, and the agent updates its belief set B. This update includes the status information that the norm N has been applied.

As mentioned earlier, our goal is to use the N-BDI* framework to model and simulate the effect of applying different norms on STDMPs. Our architecture contains three phases: recognition, adoption, and compliance. In the recognition phase, the beliefs of an agent are revised and the norms become part of the belief set of the agent. This step equals to the RC in the nBDI framework. After the adoption phase (equal to the NC in the nBDI framework), the agent checks the compliance of actions being considered and executed. Norm violations may already be noticed during the adoption phase [90], as the agent could check the compliance of the actions within the set of (abstract) plans that are also stored in the agent's belief set. In the compliance phase, the agent checks the concrete actions' compliance rather than the abstract ones in the previous phase. We added another part to the normative phase in our architecture, i.e., monitoring. In the monitoring phase, the agent will reason about his actions and their consequences in the society he operates within.

Algorithm 2: The modified control loop that checks the norm violations before selecting the appropriate action.

```
Require: pre-conditions, post-condition
    /* pre-condition and post-condition are extracted from
    GDPR.                                                    */
Require: $A_p$ and $p \in P$
Require: $N$ = Permission
Require: $Pref_p$.
if The $pre - condition$ of the assign $Pref_p$ does not satisfy. then
    Select another plan from the set of plans.
    Check the pre-condition of that plan.
else if $pre - condition$ satisfies. then
    Permission is given to the agent.
    The permitted action is performed.
    $post - condition$ realized.
    Update $B + N$
end if
```

4.3 SECURE TRUSTWORTHY DIGITAL MARKETPLACES

An STDMP is a concept developed for data sharing in an open world, allowing for parties to exchange data in a secure way, while protecting the interests of the subjects whose data is exchanged. This way the interest of the data controllers, the data subjects right, the parties that created the data transformations and the parties that have an interest in applying those transformations to that data are being protected.

To reduce the complexity of the case study, we only consider three types of the STDMP agents:

- LH: license holding agents who hold data and can provide data to the market (the STDMPs);
- TEI agents who monitor the members' behavior;
- TRF: transformation agents who hold the algorithms that require the LH(s) data.

We have to mention that these parties are not necessarily interested in the data processing results. In this chapter, we consider that the TRF agent and the party who is interested in the results are the same. But, normally, any party could have the receiving party role.

The STDMPs society is a regulated environment which includes the policies and rules defined by goverments and managers.

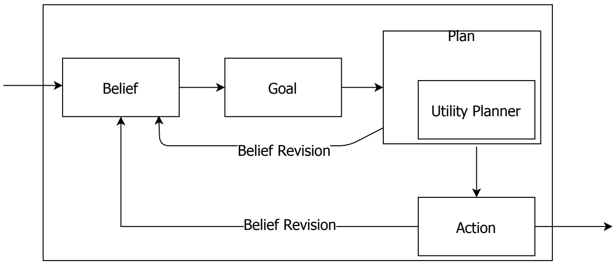


Figure 4.2: The N-BDI* framework. The agent will observe the environment and setup the belief set including the norm; next, the agent sets a goal and a set of plans are defined to achieve the goal. In the N-BDI* framework, the agent prioritizes the plans and selects a plan with the highest utility. The selected plan will be checked against the norm and in case of norm violation, the agent will choose the next plan in the list. After selecting the plan, the agent will have the associated action according to the selected plan and will execute the action.

4.3.1 *An example of the STDMP Scenario*

As we mentioned before, the STDMP consists of three main agents (LH, TEI, and TRF). Each of these agents can take the role of a processor and a controller, roles that are defined in GDPR. The processor role is responsible for processing data on behalf of the controller, which includes making data available, while the controller's role is to promptly process the requests made under the GDPR in a way that allows the processor to exercise its rights such as access or process data.

A secure and trustworthy data-sharing infrastructure build according to the STDMP architecture, can be used in various application domains. The STDMP will help the stakeholders to protect their interests and prevents data protection infringements.

To explain how the STDMP helps implements GDPR and other requirements derived from norms, we present a simplified scenario. The LH agent in its role as controller, receives informed consent for processing personal data from a data subject for a specific (set of) purpose(s). In its processor role, the LH agent asks permission from its controller to collect data and send it to the TEI agent. After giving permission, the LH processor agent collects and sends data to the TEI agent. The TEI agent asks the TRF agent for the algorithm to analyze the data. The TRF agent sends the algorithm after getting permission from its controller. The TEI agent combines the data with the algorithm and sends the result to the TRF agent. In the scenario depicted, data processing requires the protection of the interests of the stakeholders involved and compliance with GDPR. However, in this data processing infrastructure, trust needs to be organized among the stakeholders. One of the factors to organize trust is to apply the norms correctly to guide the data transactions. The TEI acts, as its name suggests – as a trusted third party. This agent's behavior should be completely determinant and transparent, and no human interference is part of that agent's actions. This allows stakeholders to check its behavior and adherence to the norms. In the scenario presented, the purpose of using the requested data must be fitted into the LH's interest and the request must meet the requirements of GDPR.

We formalize the mentioned scenario as a 4-tuple: Context = (LH, TRF, Contract, T), where the contract is the set of permissions and T is the set of allowed transformations. In general contracts contain/describe permissions, as well as definitions (ontology) and duties. There is a vast body of literature on the formal representation of norms and normative reasoning (see e.g. Van Doesburg *et al.* [44]). Norm representation and normative reasoning are not core to the research described in this thesis. And since for the Eduroam example we only need permissions, we will limit ourselves and use a simplified representation of permissions.

The TRF agent makes a transformation request (t_1) and the TEI agent receives the t_1 from the TRF agent. The TEI agent checks the eligibility of (t_1) by checking the condition of $t_1 \in T$ against the norms, where T denotes a set of allowed transformations. And, if the t_1 satisfies the condition, then the TEI agent will pass the request to the LH agent. Then, the LH agent checks the purpose of the transformation and processes the request. Note that, in the STDMP, the LH controller agent defined the set of licenses for each data set. Licenses have a defined set of conditions for using data. We visualize the scenario in Fig. 4.3. In the following section, we present a norm definition to express the contract among the members of the STDMPs.

4.4 REPRESENTATION OF NORMS

In this section, we present a recent general model of norms [98] that covers the concepts of norms and normative systems [50]. In Oren *et al.* [98], a norm n is modeled as a tuple:

Definition 1 (norm). A norm is defined as a tuple
$n = \langle role, normtype, conditions, action \rangle$ such that:

- role: indicates the organizational position;
- normtype is one of the four modal verbs "can" (which we formalize as a power), "can not" (disability), "must" (duty) and "must not" (which is not the same as a no-right, but the obligation to

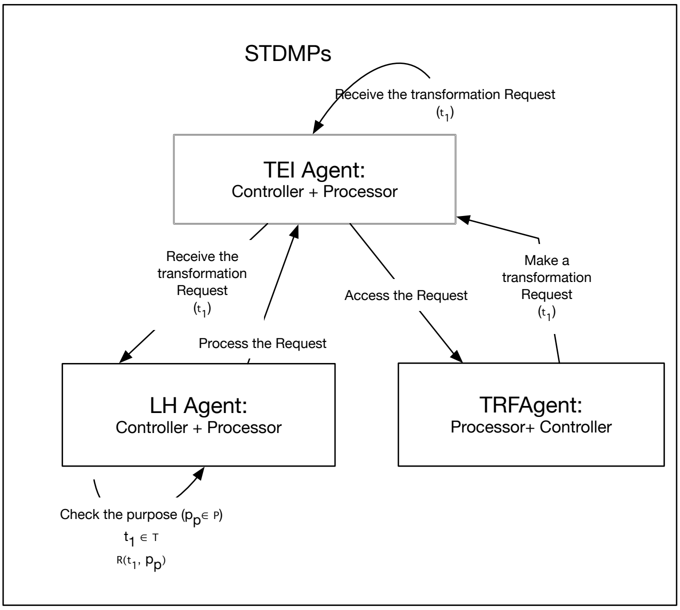


Figure 4.3: The TRF agent makes a transformation request (t_1) and the TEI agent receives the t_1 from the TRF agent. The TEI checks the eligibility of (t_1) by checking the condition of $t_1 \in T$ against the norms, where T denotes a set of allowed transformations. If the t_1 satisfies the condition, TEI will pass the request to the LH agent. And, the LH agent checks the purpose of the transformation and processes the request.

not do something!);[2] Permissions are equal to the Hohfeldian concept *power*.

- conditions (pre-conditions and post-conditions as extracted from GDPR.): describes when and where the norm holds (norm adoption);
- action: action specifies the particular action to which the normative relation is assigned (norm adoption);

Example: Consider the norm, **NormCollectData** that describes the permission to collect ***personal data*** from the ***data subject***, where the collector is the LH agent consisting of two sub-agents (the LH controller agent, the LH processor agent). The following norms are extracted from the GDPR terminology.

1. [Role: LH controller][Normative relation: ***Power***][condition: “if legitimate purpose of collecting data is specified explicit" ***AND***“ the LH controller agent has provided the data subject with the

[2] In this chapter, we define the contract as a set of permissions that when acted upon may result in other normative relations, including duties.

information on the processing of his data[3]"] [action: collecting data].

2. [Role: LH processor][Normative relation: ***Power***] [condition: "if processing of data is compatible with the purposes for which data was collected" ***AND*** "controller took appropriate measures to provide information relating to processing to the data subject ***AND*** the LH controller agent has provided the data subject with the information on the processing of his personal data"][action: process data].

4.5 IMPLEMENTATION

To implement the STDMP we used the Java Agent Development Framework based on BDI (Jadex) [104] platform. Jadex is an object-oriented software framework for the creation of goal-oriented agents following the BDI model. The Jadex reasoning engine tries to overcome the traditional limitations of the BDI agents by introducing an explicit representation of goals and a systematic way for the integration of goal deliberation mechanisms. The Jadex agent framework is built on the top of the JADE platform and provides an execution environment and an application programming interface (API) to develop the BDI agents. In this chapter, we propose to implement the STDMPs system using Jadex.

As an example of synthesis, we are now able to implement the N-BDI* framework illustrated in Algorithm 2. This is an excerpt of the code of the LH's controller agent where the agent checks the request and gives permission. The schema of the STDMP is presented in Appendix B, Fig. B.3.

4.6 RELATED WORK

Researchers have presented various extended BDI models with normative concerns for different purposes (e.g., [12, 28, 42, 130]); our N-BDI* framework is not the first model of a BDI agent architecture that considers norms, to some degree. We review some of the most influential normative BDI models, and compare these models with our model.

Kollingbaum [75] presented a Normative Architecture (NoA) with a language to define the normative concepts and a programming language to implement and reason about the norms. Kollingbaum used diagrams and text to describe the norms and implement them in

[3] Providing information to the data subject can be done before the collection of data (then it is part of the pre-condition, and the providing of information was part of a different action), or during the action of collecting data (then the result is part of the post-condition.)

Java, however, Kollingbaum (Chapter 3, p. 77, [76]) mentions that the NoA description is not sufficient to describe and test the norms. Contrastingly, Meneguzzi *et al.* [95] presented $v - BDI$ framework with a normative reasoning cycle for agents and agents can choose plans based on the given norm. Moreover, $v-$BDI specifies a norm via its constraints.

Broersen *et al.* present the Belief Obligation Intention Desire (BOID) framework [21] to include obligations in the BDI agents. In the BOID model, an agent's mental attitude (i.e., belief, desires, intentions, and obligations) is represented as a set of rules. Conceptually, BOID is well described, but because of computational complexity BOID has not been widely used in the practical environment.

The beliefs, intentions, and obligations (BIO-logical) agent framework by Governatori and Rotolo [57] present similar properties as the BOID architecture, but with less complexity. This framework uses three components of beliefs, intentions and obligations as the agent's mental attitude. Governatori *et al.* use Propositional Defeasible Logic to express the relations among agent's mental states. The authors claim that their framework has linear complexity and therefore, should scale up well. One of the limitations of the BIO-logical framework, however, is that due to using proposition calculus as representation formalism, the BIO-logical framework has limited applicability as the representation formalism is not expressive enough to express any practical case studies [95].

Meneguzzi and Luck [94] present a normative interpreter for AgentSpeak(L) that analyzes norms and modifies an agent's plan set to comply with the norms adopted by the agent. This normative AgentSpeak(L) however, only supports a specific type of norms, so the planner may only generate particular instances of plans or to all examples of a certain plan, which has been considered as serious shortcomings of the normative AgentSpeak(L) [95].

In contrast, in our approach, although norms are stored together with other beliefs, we can retrieve the norms directly by the $getNorms$ function (obtaining, specific norms from the belief set). The reason that we keep norms in the belief set is to avoid creating an external norms repository. The advantages are two-fold. First, norms are typically communicated (in the case of humans employing language) and may be interpreted differently by different agents. Second, this allows us to have a similar architecture as the original BDI model.

Hubner *et al.* [67] present the *Moiseb*^{2} model, which represents an organizational agent model. *Moiseb*^{2} model contains the functions, structures, and deontic [66]. An agent in the *Moiseb*^{2} model has a certain role, links, and certain tasks towards the achievement of collective goals. The agent's roles, tasks, and relations are expressed using deontic logic, which is the main difference with our model.

Framework	Base system	Normative-language	Normative Reasoning	References
NoA	NoA	Horn-Clause Logic	JAVA-described	[76]
$\nu-$BDI	AgentSpeak	Horn-Clause Logic	State-based	
BOID	Prop. logic	Propositional Logic	State-based	[21]
BIO	Prop.defeasible logic	Prop. defeasible logic	Proof-theoretical	[57]
Normative AgentSpeak	AgentSpeak	Horn-Clause Logic	Event-based	[94]
MaNEA	Magentix2	First-Order Logic	Rule-Based	[28]
N-2APL	2APL	Propositional Logic	Deadlines and priorities	[6]
Panagiotidi *et al.*	2APL	Ground Horn-Clauses	State-based	[99]
Lotzmann *et al.*	EMIL	First-Order Logic	State-based	[89]
Sadri *et al.*		Ground Horn-Clauses	State-based	[114]
N-BDI*		BDI	JAVA-described	[33]

Table 4.1: Summary of related frameworks compared to N-BDI*.

Garcia *et al.* [28] propose a method to manage the agent normative positions (i.e., permissions, prohibitions, and obligations). The Knowledge, Goals, Plan (KGP) model proposed by Sadri *et al.* [114], to support agents with normative concepts, based on the roles that the agent plays with the obligations and prohibitions that result from playing these roles.

The EMIL [89] architecture is one of the most detailed normative architectures described in the literature. This architecture defines two sets of components for each agent:

1. Epistemic, which is responsible for recognizing norms;
2. Pragmatic, which is responsible for guaranteeing that the institution creates some (usually normative) agent's behavior.

Applying the EMIL architecture in real scenarios can be complicated due to the detailed design of its' cognitive mechanisms.

Alechina *et al.* [6] present $N-2APL$, which is an extension of $2APL$. The $N-2APL$ language aims to provide a norm awareness mechanism for an agent.

Panagiotidi *et al.* [100] propose a planning framework under normative constraints. The norm representation in their framework includes the norms activation and expiration conditions with two norm-compliant and norm-violating mechanism presented in logic.

One of the pioneering architectures in the area of normative multi-agent systems was the deliberative normative agents' architecture [24]. Violating norms is acceptable in the architecture. Agents deliberate about the norms that are explicitly implemented in the model. Panagiotidi *et al.* presented a norm-oriented agent [100]; this agent considers operationalized norms during the plan generation phase and uses them as guidelines for the agent's future action path. Boella and van der Torre [16] introduced a defender and controller agent in their normative multi-agent system. In their models, defender agents should behave based on the current norms. Controllers monitor the behaviors of other agents and sanction violators, who can also change norms if needed.

We summarize the related work in Table 4.1. We compare three dimensions of each framework discussed in this section, namely the base system upon which norm reasoning is built (the base agent system where appropriate); the language in which norms are specified; and the type of normative reasoning performed by the system. Finally, we highlight the main references that detail these frameworks.

4.7 CONCLUSION

The STDMPs society is a regulated environment which includes different sorts of rules and policies issued by different parties. For these reasons, we consider the secure data sharing problem a representative example of a societal problem where norms impact the autonomous agents involved. Hence our case study, which we also used for evaluating the capability of the N-BDI* framework to identify the non-compliant member. The agents' behavior in our STDMP model is affected by different sorts of norms which are controlled by different mechanisms such as regimentation, enforcement and grievance and arbitration processes. Although this chapter focuses on the STDMPs, we believe our architecture is sufficiently general to study a variety of social scenarios.

We identify the main goals of the TEI agent as being twofold.

- First, it aims at supporting agent interaction as a coordination framework, making the establishment of business agreements more efficient.
- Second, it serves to provide a level of trust by offering an enforceable normative environment.

Our research focuses on modeling normative reasoning in a completely distributed environment. In particular, we are interested in how norms affect the STDMPs, which monitoring activities enable detection of (non-)compliance in networked societies of agents, and what enforcement activities would enhance compliance. To support this, we have implemented a prototype of the N-BDI* architecture.

By extending the agent control loop, generating a plan set by considering the norms and selecting a plan based on its utility, we present how the agent can act in a normative environment.

4.8 FUTURE WORK

We can identify several avenues for future work. In reality agents may choose to violate norms by executing a non-compliant plan. To represent this in an agent-based model, we should be able to model rational agents that can choose to violate norms if the benefits of violating the norms are higher than the benefits of complying with the norms (or the costs less than the costs of being compliant). We intend to extend our current framework so that we can represent all relevant aspects of normative reasoning. It is our ambition to create a generic normative framework which can be applied to various domains, such as data sharing in the Data Logistics for Logistics Data (DL4LD) and Enabling Personalized Interventions (EPI) project[4] where the autonomous members need to comply with the agreed contract while they want to achieve their own goals simultaneously. Presenting a mechanism that can capture the conflicts between the norms is also one of the future directions.

[4] https://delaat.net/

5 SOCIAL COMPUTATIONAL TRUST MODEL

In this chapter, we present the social computational trust model (SCTM), our agent based approach to computational social trust, which comply to the requirements/constraints we identified as vital in Chapter 2 and answer *RQ*3: "How can we express trust among members in a collaborative network in a dynamic computational model?" and sub-research question *RQ*3.2 "What generic risk factors can be identified?". Creating a cybersecurity alliance among organizations, as a means to minimize security incidents, has gained the interest of practitioners and academics in the last few years. To create and maintain stability within a cybersecurity alliance, members must trust each other. If organizations do not trust each other while interacting, the alliance is at risk to fail.

The SCTM model helps alliance members to select the right partner to collaborate with in order to perform collective tasks. Trust is a precondition for being willing to share data regarding the incident and intelligence. The SCTM combines benevolence and competence to estimate the interaction risk. Benevolence is computed from personal experiences gained through direct interactions, whereas competence of other alliance members (as the model only asks the neighbors of a trustee not 'the other members'). We propose an algorithm to evaluate the competence and benevolence of a compromised member. We developed a Belief Desire Intention– agent based model (BDI-ABM) as a part of our case study, which we present in this chapter to demonstrate our approach. The practicability of the proposed risk estimation approach is validated with a detailed experiment in Section 6.3 Chapter 6.

This chapter is based on:

- **Deljoo, A.**, Koning. R., van Engers,T., Gommans, L. & deLaat, C., Managing Effective Collaboration in Cyber-security Alliances Using Social Computational Trust, Cyber Security in Networking CSNet (2019). © IEEE.
- Koning, R., **Deljoo, A.**, Meijer, L., Grosso, P., & deLaat, C., Trust based collaborative defences in multi network alliances, Cyber Security in Networking CSNet (2019). © IEEE.
- **Deljoo, A.**, van Engers,T., Gommans, L. & deLaat C., Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners, New Generation Computing, [under review] © Springer.
- **Deljoo, A.**, van Engers,T., Gommans, L., & deLaat C., The impact of competence and benevolence in a computational model of trust, IFIP International Conference on Trust Management, 45-57 (2018) © Springer.

5.1 INTRODUCTION

Cyber attacks are serious threats to our networked society as organizations depend on the well functioning of the IT-infrastructure to guarantee vital processes. As attacks are becoming more and more organized, collaboration across public and private organizations is required to arrange technical countermeasures [3, 36]. Sharing cyber intelligence among different parties, such as internet & cloud service providers and enterprise networks, has become increasingly important.

In order to support the establishment of such collaboration, we need to organize and subsequently manage trust first, enabling organizations to let their trusted partners share cybersecurity information.

In this chapter, we explain how we support establishing partnerships within cybersecurity alliances by providing an ability to computationally evaluate potential partnerships among a community of alliance members [34].

We focus on the social aspect of trust and select the “right" partner to perform joint tasks. The term “right" implies that a member of the alliance is trustworthy to collaborate with. Traditionally, information sharing on a peer-to-peer basis is mostly established based on personal trust. But, the social network of organizations changes with time and becomes more complex, while the number of interactions may be far too many to manually decide on trust as a precondition for sharing data or intelligence. Therefore, it demands to define a more sophisticated and computationally executable method to select the “right" partner for sharing data and intelligence. In this work, we present the following contributions:

1. The Social Computational Trust Model (SCTM) represents social trust, which components are essential for evaluating the partners.
2. Risk assessment through the SCTM model. The SCTM model facilitates risk-based partner selection to select the "right" partner by combining the benevolence and competence factors. We identify two common risks for the members of the alliance.
3. The evaluation of a compromised member in the alliance.

Hereto we consider in our social computational trust model (SCTM) three aspects of trust as trustworthiness components. Based on these three trust components, our model aims to strengthen the effectiveness and stability of a cybersecurity alliance by enabling individual members to evaluate and select the most trustworthy partner for a particular situation at hand while keeping the risk of interaction at a minimum.

Therefore, there is a need to estimate the interaction risk for each organization in the alliance. This motivated us to propose a risk estimation framework, which provides a quantitative estimation of the risk for the partner involved in interacting with other members of the alliance. We present a mechanism to evaluate a compromised member's competence.

5.2 RELATED WORK

Different scholars have presented many computational trust models; nevertheless, only a few models are social computational models. One of the conceptual models of social trust developed by Adali et al. [5] is based on the model presented in [73], which takes ability, positive intentions, ethics, and predictability for the trustworthiness components.

They used a probabilistic approach in their model; however, by realizing the limits of the approach in the treatment of the social concepts, their model was not implemented [5]. Among all the presented computational trust models [103], the only computational approach that includes a complete set of established features based on the theory of trust is the socio-cognitive model developed by Castelfranchi and Falcone [23]. In their view, trust is established by considering the different beliefs that the trustor has about the trustee, both internal (beliefs on competence, disposition, and harmfulness) and external (opportunities and dangers). Meta-beliefs further adjust the importance of these beliefs about the relative strength of each belief. In practice, it is difficult to implement due to its richness.

Another social trust model is called situation-aware computational trust model (SOLUM), developed by Urbano et al. [125]. Their computational model consists of two parts. The first part is a general framework of computational trust, which is based on two fundamental character-

Table 5.1: Summary of the Trust Models.

Trust Models	Trust Components	Model Proposed	Mathematical Presentation	Risk Factor	Validation	
Adali's model	ability, positive intentions, ethics, and predictability	×	×		×	×
Castelfranchi and Falcone	internal and external belief of a trustee	×	✓	×	×	
SOLUM	ability, integrity and benevolence	✓	✓	×	×	
CoTAG	reliability, helpfulness, and local reputation	✓	✓	×	✓	
SELCSP	trustworthiness and competence	✓	✓	✓	✓	
SCTM (the proposed work)	competence, benevolence and integrity	✓	✓	✓		✓

istics of trust, the trustor's disposition and emotional state. For the second part, they propose a set of distinct techniques to extract information about the individual dimensions of a trustee's trustworthiness from the set of structured evidence available to a trustor. The main differences between our model and Urbano's model are that we employ the context definition with eight dimensions of context and consider different stages of relationships for the competence function. We extend Marsh's definition of competence [92] by considering three different situations for the trustor to decide about a (future) collaboration with the trustee. Furthermore, we introduce the risk estimation approach through the SCTM model.

The presented work by Gosh et al. [53] formalizes trust as competence and trustworthiness evaluation functions. They use the direct and indirect evidence to estimate the value of each component and their model recommends the most trustworthy cloud service provider to a cloud user. They do not address in their model the concept of eight dimensions of context we present. However, we compare the results of our study with the findings of their work (see Section 6.3.3, Chapter 6).

Fortino et al. [52] developed another social trust model, where they presented the CoTAG Algorithm (Cloud of Things Agent Grouping algorithm) by using the local trust metrics such as reliability, helpfulness, and local reputation to form the groups of agents in the IoT (Internet of Things) trusting scenario. In Table 5.1, we compare the presented trust models. It is evident from Table 5.1 that most of the models have not provided the mathematical formulation of their trust model or did not consider the risk estimation approach. In [53], the authors present some results according to the partner selection concept; however, the motivation of their approach is different.

5.3 TRUST

Trust is seen as an essential precondition for any interactions in the social system. Trust is studied in different areas, from sociology to psychology [93][1]. Trust among the members of alliances has been

[1] An elaborated overview of the concepts used in the organizational context can be found in studies performed by Bachmann [8, 93].

empirically demonstrated to be important for alliance formation [108]. Trust has some benefits for alliances, being a substitute for formal control mechanisms, reducing transaction costs, facilitating dispute resolution, and allowing for more flexibility.

Also, we consider that trust is a multidimensional construct. Trust encompasses as a number of characteristics, including expectation and belief that the other members act with goodwill and behave as expected. Therefore, in this work, we conform to the following description given by Mayer [93] *"Trust is the willingness* of a trustor to be vulnerable to the actions of a trustee based on the expectation that the trustee will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other parties". We define trust as the expectation held by a trustor that a trustee will not exploit any vulnerabilities when faced with an opportunity to do so [10, 36, 83, 93]. This expectation is justified when the given member:

- Has the potential ability to perform a given (sub-)task (competence).
- Adheres to a set of rules agreed upon and acts accordingly to fulfill the commitments (integrity).
- Acts and does good even if unexpected contingencies arise[2] (benevolence).

Fig. 5.1 depicts the trust framework. Essentially, the framework says that a member is trustworthy if he has an ability to perform a (sub-)task in a given situation, has integrity, and has a positive relationship with the trustor. Once trust is established, the trustor is willing to take the risk, and the outcome of the risk estimation block serves as a feedback to update the perception about the trustee's factors (i.e., competence, integrity, and benevolence). Therefore, it is important to estimate the trustee's trustworthiness by considering each of these three dimensions individually, and dynamically combine them by considering different situations and stages of relations. However, most of the computational trust approaches evaluate the trustee's trustworthiness as a black box and do not consider different trustworthiness's dimensions such as competence, benevolence and integrity [103]. Our computational trust model is based on the multidisciplinary literature on trust [8, 23, 93], describing the evaluation of competence, integrity and benevolence of the given trustee. To reduce the complexity of our initial simulation approach, we motivate that benevolent behavior could also imply integrity.

[2] Acts toward the interest of the alliances.

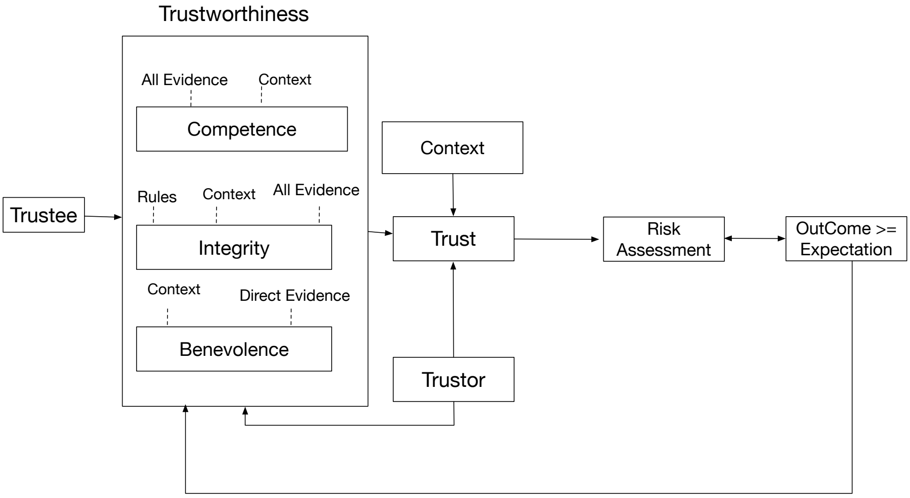


Figure 5.1: The framework with the three trustworthiness components: benevolence, integrity, and competence. The combination of these three components determine trust.

5.4 OUR SOCIAL COMPUTATIONAL TRUST MODEL

In this section, we introduce the SCTM. This model provides the basis for the decision-making process that each member has to perform when deciding on collaborating or not with other members. This process can be broken down into two sub-processes: (1) evaluate trust based on three distinct factors (integrity, benevolence, and competence), and (2) evaluate the interaction risk of entering the alliance according to the trust value. In the following, we present our SCTM model to evaluate trust based on benevolence, competence and integrity factors.

5.4.1 *Notation*

The SCTM model is applied to environments where trustor agents choose the right trustees to interact with, with or without the posterior establishment of detailed agreements between partners. Therefore, to evaluate trustees and select the right trustee to collaborate with, we present the extension of the trust framework by adding the context dimensions to the context component in Fig. 5.2. In the following, we present the notations and mathematical models to evaluate a trustee's trustworthiness.

We denote an agent society by A, and it includes a trustor x and trustee y, $x, y \in A$. In this research, each member of an agent society can be represented as a trustee or trustor. $Tr(x, y, s_i, \tau)$ represents the amount of trust that trustor x has in y to perform a task τ in a situation

$s_i \in S$, where $S = \{s_1, s_2, ..., s_n\}$ is the set of all the possible situations in the society.

In order to define the situations that lead to an agreement between a trustor and a trustee, the authors in [4] define the context which consists of four main dimensions: identity, time, location, and activity. We extend the Urbano *et al.* [125] context concept and identify eight dimensions of context $\{d_1, d_2, ..., d_8\}$. The dimensions d_1 and d_2 represent the agents, the trustor and the trustee, respectively, dimensions d_3 represents time of agreement and d_4 defines a situation. The remaining dimensions d_5, d_6, d_7 and d_8 characterize the task type, its complexity, deadline, and outcome of the task, respectively. For d_8, we distinguish three different types of outcomes $d_8 \in \{Fd, Fdd, V\}$. Fd denotes a fulfilled duty; i.e. the trustor, based on the evidence, concludes that the trustee performed the given task on time. Fdd (fulfilled duty with delay) means the trustee performed the given task with an (un)expected delay, and V means the trustee did not perform the agreed task (violation).

In this thesis, we assume that all agreements between a trustor and a trustee refer to the same task type τ, but with different deadlines (d_7) and outcomes (d_8) for the (sub-)tasks.

- (d_5): The task type refers to the agreed task between a trustor and a trustee. Here, we assume that one task, τ being negotiated by all members, is composed of different sub-tasks $\tau_{s1}, ...\tau_{sn}$.
- (d_7): The trustee needs to answer the trustor's request within a specific time window Δt_w.
- (d_8): We employ Algorithm 3 to calculate the outcome d_8 of each (sub-)task.

Algorithm 3: Calculates the outcome d_8 of a (sub-)task.

```
Input  : $d_7$: deadline of a (sub-)task
         $t_{request}$: request time of a (sub-)task
         $t_{report}$: report time of a (sub-)task.
Output: $d_8 \in \{Fdd, Fd, V\}$

$d_8 := V$ /* Initialize the return value */
$\Delta t_w := t_{report} - t_{request}$
if $0 < \Delta t_w < d_7$ then
    $d_8 := Fd$
else if $\Delta t_w \geq d_7$ then
    $d_8 := Fdd$
end
return $d_8$
```

If trustor x wants trustee y to perform a task $\tau = \{\tau_{s1}, ...\tau_{sn}\}$ consisting n sub-tasks, trustor x creates for each sub-task $\tau_{si}, i = 1, ..n$, a record in its database Kb_x. Each record contains a deadline d_7 for the corresponding sub-task, within which trustor x expects a report from trustee y. When trustor x sends the task to trustee y, it is the

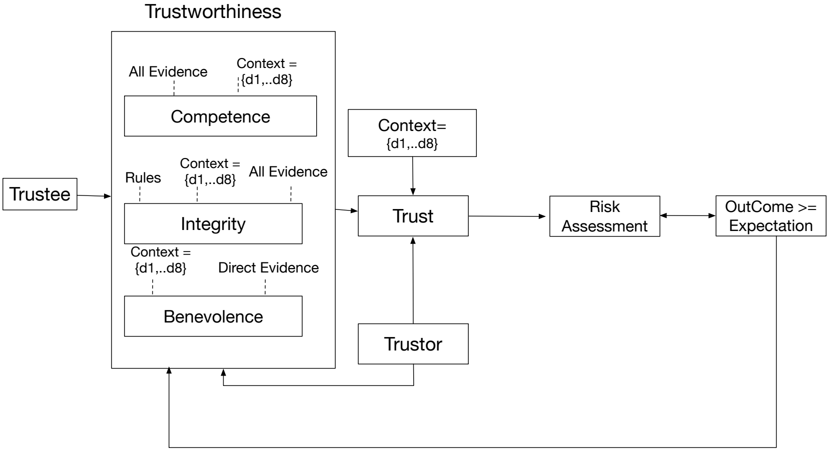


Figure 5.2: The three trustworthiness components, benevolence, integrity and competence, are presented in this framework. Integrity and Competence components get the context with its dimensions and all the available evidence to evaluate a trustee. Benevolence component has the context and direct evidence as inputs to calculate the trustee's benevolence.

responsibility of trustee y how to schedule the sub-tasks and to report for each completed sub-task.

For each receiving report from trustee y, trustor x applies Algorithm 3 to update the corresponding record (i.e., d_8) in its Kb_x. Note that in our work, we consider that sub-tasks are independent of each other. But, if the outcome of sub-tasks are depended on each other, then trustor x should add a task graph (or dependency graph) to his request, (i.e., a Directed Acyclic Graph (a DAG)) that defines the dependencies of the sub-tasks [123].

Table. 5.2 we summarize the notations that we use in the rest of the thesis.

Table 5.2: Notations and values

Description	Representation	Value Range
Society of Agents (trustor, trustee)	A	
Agent	$x, y \in A$	
Knowledge-base of agent x	Kb_x	
`/* Also a trustee can have a knowledge base. */` Set of Situations	$S = \{s_1, s_2, ..s_m\}$	
Task of Sub-tasks	$\tau_{s1}, ...\tau_{sn}$	
Context	$D = \{d_1, d_2, ...d_8\}$ 3	
d_8	$\{Fd, Fdd, V\}$	1, 0.5, 0
All the direct evidence of trustor x on trustee y in the situation s_i	$Ed(x, y, s_i, \tau; Kb_x)$	
All the available evidence (indirect) on y from y's neighbors in the situation s_i in Kb_x	$Ec(nbr_y, y, s_i, \tau)$	
Trustee's trustworthiness toward trustor x in the situation s_i	$TW(x, y, s_i, \tau)$	[0,1]
Trust x on y to perform τ in the situation s_i	$Tr(x, y, s_i, \tau)$	[0,1]

The outcome of interactions between trustor x and trustee y is called evidence (E). In the current SCTM model, we consider all the available evidence on a trustee. And, each trustor has a knowledge-base (Kb) that contains all the interactions with its neighbors. The trustor stores the following information from its interactions in its Kb, the message Id, Id of a trustee (e.g., sender), $t_{request}$, t_{report}, a task type and outcome of tasks (see Fig. 5.4). We consequently define evidence (E) as the outcome of the interaction between trustor x and trustee y performing a (sub-)task τ by trustee y in a situation $s_i \in S$. According to Algorithm 3 we denote this evidence by $d_8(x, y, s_i, \tau)$. Next, we define function $val(.) : d_8 \rightarrow [0, 1]$ that assigns a value in the interval $[0, 1]$ to d_8:

$$val(d_8) = \begin{cases} 1 & \text{if } d_8 = Fd \\ 0.5 & \text{if } d_8 = Fdd \\ 0 & \text{if } d_8 = V \end{cases}$$

5.5 EVIDENCE GATHERING

The direct evidence ($Ed(x, y, s_i, \tau; Kb_x)$) of the interaction between trustor x and trustee y to perform a (sub-)task τ in situation s_i is defined by the set:

$$\begin{gathered} Ed(x, y, s_i, \tau; Kb_x) = \\ \{d_8(x, y, s_i, \tau) \in Kb_x | d_1 = x, d_2 = y, d_4 = s_i, d_5 = \tau\}, \end{gathered} \tag{5.1}$$

that is the set of d_8 values from all entries in the knowledge-base Kb_x of trustor x that deal with the interaction between x and y to perform a (sub-)task τ in situation s_i. Here and in the following the notation for τ as a set of sub-tasks means that it refers to an aggregation over all the sub-tasks. To extract the evidence of the other dimensions of context, we can replace d_8 by other dimensions such as d_5 or d_6 to extract the evidence of that specific dimension.

We define the function $val_d(.) : Ed \rightarrow [0, 1]$ that assigns a value to the set of direct evidence $Ed(x, y, s_i, \tau; Kb_x)$ in the interval $[0, 1]$ as:

$$\begin{gathered} val_d(Ed(x, y, s_i, \tau; Kb_x)) = \\ \frac{1}{N_x} \sum_{d_8(x,y,s_i,\tau) \in Ed(x,y,s_i,\tau;Kb_x)} val(d_8(x, y, s_i, \tau)), \end{gathered} \tag{5.2}$$

where N_x is the number of entries in Kb_x that deal with x and y in situation s_i and task τ. val_d denotes a function that assigns a value to the direct evidence. We assume that at least there are two entries in the Kb_x, i.e., $N_x >= 2$ [125].

[3] Dimensions are: d_1 = trustor, d_2= trustee , d_3 = time, d_4= location, d_5= task, d_6= complexity, d_7= deadline, d_8= Outcome.

Likewise, we define the available evidence (indirect) of the interactions between the neighbors of y as a trustor and y as a trustee for situation s_i as the set:

$$\begin{gathered} Ec(nbr_y, y, s_i, \tau) = \\ \{Ed(u, y, s_i, \tau; Kb_u) | u \in nbr_y\}, \end{gathered} \tag{5.3}$$

where $Ec(nbr_y, y, s_i, \tau)$ denotes all the available evidence from the set of y's neighbors (nbr_y). In our network schema (Fig. 5.3b) trustee y has four direct neighbors, Z, A, M, W.

For this set we define a function $val_c(.) : Ec \rightarrow [0, 1]$ that assigns a value in the interval $[0, 1]$ to the available evidence (indirect) set ($Ec(nbr_y, y, s_i, \tau)$) as:

$$\begin{gathered} val_c(Ec(nbr_y, y, s_i, \tau)) = \\ \frac{1}{N_{nbr}} \sum_{Ed(u,y,s_i,\tau;Kb_u) \in Ec(nbr_y,y,s_i,\tau)} val_d(Ed(u, y, s_i, \tau; Kb_u)), \end{gathered} \tag{5.4}$$

where N_{nbr} is the number of neighbors that contribute to val_c. val_c represents a function that assigns a value to all the available evidence on trustee y stored in the Kbs of trustee y's neighbors.

5.5.1 *Social Computational Trust Model*

The social computational trust model (SCTM) that we explain in this chapter combines three distinct functions, namely a competence, a benevolence, and a integrity function. These functions are presented in this section as illustrated in the SCTM model in Fig. 5.4.

5.5.1.1 *The benevolence evaluation function*

Several scholars have considered the benevolence as one of the key elements of trust and the trustworthiness's antecedent (e.g., [82, 86]). Benevolence shows the intention and integrity of a trustee towards a trustor. The benevolence and integrity aspects are somehow complementary to each other. In some scenarios, where the number of interactions between a trustee and trustor are limited or the trustor has only one neighbor, the trustee's trustworthiness can be evaluated by the integrity functions where we consider the experiences of all the members. In similar cases, when most of the members are malicious, and evidence tends to be misleading, benevolence can help to evaluate the trustworthiness by relying more on the available direct evidence between the trustee and the trustor. The benevolence value, $Ben(x, y, s_i, \tau)$, of trustee y toward trustor x is computed from their mutual interactions in the situation s_i. Trustor x calculates the benevolence of y by extracting the direct evidence of its interactions with trustee y

from its knowledge-base Kb_x (see Figure. 5.3a). $val_d(Ed(x, y, s_i, \tau; Kb_x))$ assigns a value to the extracted direct evidence between x and y in situation s_i. The benevolence function equals the function val_d, which is in the interval of $[0, 1]$, given by:

$$Ben(x, y, s_i, \tau) = val_d(Ed(x, y, s_i, \tau; Kb_x)). \tag{5.5}$$

The trustee's intention is a state of mind and is not directly observable, accessible by the trustor, and may be different from the actions taken by the trustee towards the members of the society. Nevertheless, a trustee's intentions can be inferred by looking at the trustee's attempts to follow through with an action it has agreed to perform [9, 31]. The successfulness of such actions are reflected by the competence of the trustee.

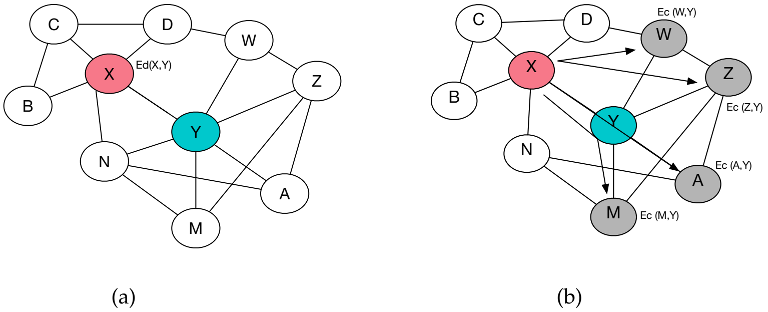
5.5.1.2 *The competence evaluation function*

Marsh [92] stated that a trustee's competence could be evaluated based on the different situations, and we derive from his definition an enumeration of situations that the trustor needs to evaluate the trustee's competence. In other words, the competence of trustee y is described from the viewpoint of trustor x under specific situations. These situations are:

1. The trustee is not known to the trustor, in the current or similar context.
2. The trustee is known to the trustor, but not in the current or similar context.
3. The trustee is known to the trustor and trusted in the current or similar context.

If the trustee is known to the trustor and trusted, then the trustee has proven his ability to perform the agreed task [35, 92], and the trustee's competence can be derived from the previous interaction. However, when the trustee is unknown to the trustor (in case the trustee is a new member to the network or the trustee and the trustor have not yet interacted), the trustee's competence needs to be evaluated. Therefore, we propose the following functions to gather the required evidence. To collect the evidence, the trustor requests the trustee's direct neighbors to provide him the available evidence on the trustee.

The competence function $Com(nbr_{y \backslash x}, y, s_i, \tau)$ evaluates the given trustee's ability in performing a given (sub-)task τ in the specific situation s_i. The competence function takes all the evidence available on the trustee under evaluation as inputs. To gather the evidence, the trustor will request the evidence from the direct neighbors of the trustee about the ability of the trustee on performing the (sub-)task τ,



(a) (b)

Figure 5.3: Gathering all the available evidence on trustee y. (a) Gathering the direct evidence on trustee y, trustor x extracts the evidence from its Kb_x. (b) Gathering the indirect evidence on trustee y from y's direct neighbors.

in our model. The set $Ec(nbr_{y\setminus x}, y, s_i, \tau)$ represents the complementary evidence on trustee y from all its direct neighbors about trustee "y"'s performance on a (sub-)task τ and where $nbr_{y\setminus x}$ denotes the direct neighbors of trustee y except the trustor x. Figure. 5.3b shows trustor x sending a request to the neighbors (Z, A, M, W) of trustee y to gather the evidence about the performance of trustee y.

Similar to the benevolence function, the value for the competence of the trustee is in the interval of $[0, 1]$:

$$Com(nbr_{y\setminus x}, y, s_i, \tau) = val_c(Ec(nbr_{y\setminus x}, y, s_i, \tau)), nbr_{y\setminus x} = nbr_y \setminus \{x\}. \tag{5.6}$$

The competence, integrity, and benevolence functions are normalized to be in the interval $[0, 1]$. To gather the evidence, x has been excluded from the neighbors of y. In the SCTM model, integrity and benevolence are behavioral properties, while competence depends on the ability of the trustee to perform the given (sub-)task τ.

5.5.2 *The Integrity evaluation function*

Integrity refers to the consistency of trustees' behaviors to adhere to a set of norms (agreed contract) [93]. Therefore, in our model, we define a trustee's integrity as his consistency in his past actions, which means that the trustee is consistent in fulfilling his promises and performing the given task successfully (promises are regarded as the agreed contract). Like the direct evidence Eq.5.1, we define the restricted direct evidence for a successfully fulfilled task τ in situation s_i as:

$$Ed(x, y, s_i, \tau; Kb_x, d_8 = Fd) = \{d_8(x, y, s_i, \tau) \in Kb_x | d_8 = Fd\}, \tag{5.7}$$

$$Int(nbr_{y\setminus x}, y, s_i, \tau) = \frac{1}{N_{Ec}} \sum_{u \in nbr_{y\setminus x}} |Ed(u, y, s_i, \tau, kb_u, d_8 = Fd)| \quad (5.8)$$

where $N_{Ec} = |Ec(nbr_y, y, s_i, \tau))|$.So, the integrity is the weighted sum of the number of entries in the knowledge-bases Kb_u of neighbors u of y about task τ in situation s_i under the condition that the task is fulfilled successfully.

5.5.3 *The Trustworthiness evaluation function*

The trustworthiness evaluation function $TW(x, y, s_i, \tau)$ estimates the trustworthiness of a trustee towards a trustor from the combination of the competence, the integrity, and the benevolence function, such that:

$$\begin{gathered} TW(x, y, s_i, \tau) = \\ \frac{1}{3}(Ben(x, y, s_i, \tau) + Com(nbr_{y\setminus x}, y, s_i, \tau) + Int(nbr_{y\setminus x}, y, s_i, \tau)) \end{gathered} \quad (5.9)$$

.

It should be mentioned that, unlike our trustworthiness evaluation function, in the work of Guo *et al.*[61] and Liu *et al.*[88] the multiplication operation is used instead of the summation operation. The reason is that they presented a rank estimation approach instead of evaluating the trustworthiness of trustees. Nevertheless, they claimed that the value could be small due to the multiplication operations.

5.5.4 *The trust evaluation function*

The trust evaluation function $Tr(x, y, s_i, \tau)$ estimates the trust that trustor x has in trustee y in the situation s_i. As illustrated in Fig. 5.4, this function takes the estimated value of the trustee's trustworthiness, which is given by Eq. 5.9. This means that, so far it concerns the work of this chapter, $Tr(x, y, s_i, \tau) = TW(x, y, s_i, \tau)$. Therefore, $Tr(x, y, s_i, \tau)$ returns the value of the trust that trustor x has in trustee y in the situation s_i, and is presented by:

$$\begin{gathered} Tr(x, y, s_i, \tau) = \\ \frac{1}{3}(Ben(x, y, s_i, \tau) + Com(nbr_{y\setminus x}, y, s_i, \tau) + Int(nbr_{y\setminus x}, y, s_i, \tau)). \end{gathered} \quad (5.10)$$

We conclude that we can evaluate trust between any trustee and trustor, no matter whether they have direct interactions or not.

As we mentioned, the trust factors have potentially unique effects on evaluating the members. Moreover, Mayer *et al.* [93] stated that trust computes from different factors in different situations. Therefore, in this

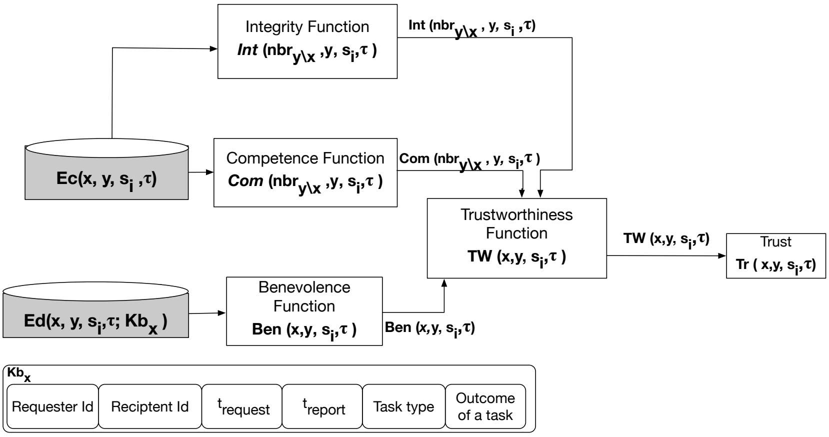


Figure 5.4: **S**ocial **C**omputational **T**rust **M**odel (SCTM). SCTM uses the competence, benevolence, and integrity functions to evaluate a trustee's trustworthiness.

chapter to determine a proper set of values for the three components of the SCTM model, we employ three different weights for the benevolence, competence and integrity functions to determine the impact of each factor to select the right partners. Therefore, $Tr(x, y, s_i, \tau)$ in Eq. 5.11 returns the trust value that trustor x has in trustee y in the situation s_i and a τ needs to be performed by y, and is presented by:

$$\begin{gathered} Tr(x, y, s_i, \tau) = \frac{1}{3}(\alpha * Ben(x, y, s_i, \tau) + \\ \beta * Com(nbr_{y\backslash x}, y, s_i, \tau) + \gamma * Int(nbr_{y\backslash x}, y, s_i, \tau)) \end{gathered} \tag{5.11}$$

where α, β and γ are values in the interval $(0, 1]$. The SCTM model aims to provide a quantitative estimation of the interaction risk. Few trust models explicitly take the risk factors into account [103]. In most of the computational trust models, the risk is considered as a factor that a user must derive from different ingredients recommended by the system. In many cases, the relation between the risk and trust are left in the shadow [31]. Most of the trust models acknowledge the intuitive observation where the trust and risk are in an inverse relationship, i.e., a low/high trust value is associated with a high/low risk, or, risk and trust pull in opposite directions to determine a member's acceptance by a partner [58, 70].

5.6 RISK ESTIMATION THROUGH THE SCTM MODEL

The SCTM model aims to help the alliance members to select a "right" partner to collaborate with on joined tasks. The term "right" implies that a member of the alliances has enough benevolence and competence. This reflects in a low interaction risk. Das *et al.* [31] stated that benevolence reduces a partner's relational risk in an alliance. As we mentioned before (see Section 5.5.1.1), benevolence is a behavioral property and shows that an alliance member has a reputation for dealing fairly and caring about its partner's interest in the alliance [30, 49, 131]. The scholars believe that benevolence [4] of a trustee reduces the perceived likelihood of opportunistic behavior occurring, which leads to low transaction costs [31]. The opportunistic behavior has been viewed as the "dark side" of inter-organizational relationships and defines as "one firm may not abide by the terms of the agreement in order to exploit the other for short-term gains"[101]. The opportunistic behavior counts as a reason that increases the relational risk. The relational risk in any alliance increases if one of the partners finds it difficult to protect its own resources from other members, this act can be seen as the opportunistic behavior [31].

Competence of a trustee refers to the ability of the trustee to perform the task. Therefore, the competence of the trustee reduces the performance risk in the alliance. It is essential to distinguish between the relational and the performance risk. For example, depending on whether the relational and the performance risk is more of a threat to the stability of an alliance, members may decide on the strategy to mitigate the risk [30, 92, 93]. Therefore, we propose to assign the weights to the performance and relational risk factor (w_1, w_2) that can be adjusted for different situations. In Fig. 5.5, we present the trust framework and risk estimation factors. The trust is calculated based on the context and indirect or direct evidence on a trustee. After the trust relation is established, the trustor will calculate the interaction risk by combining the relational and performance risk. We conclude that the relational and performance risks are independent events.

The SCTM model computes the interaction risk between a trustor and a trustee in the situation s_i using the computation of benevolence and competence of a given trustee. In the following, we explain our risk estimation framework as shown in Fig. 5.6.

When a trustor decides to estimate the interaction risk of the given trustee, the trustor sends a request to the risk estimation block regarding the estimation of the interaction risk of the given trustee. According to different scenarios and contexts, we can assign different weights to the competence and the benevolence functions; w_1 and w_2 are their corresponding weights. We define the relational and the performance risks as follows:

[4] Some of the scholars use goodwill trust instead of benevolence-based trust.

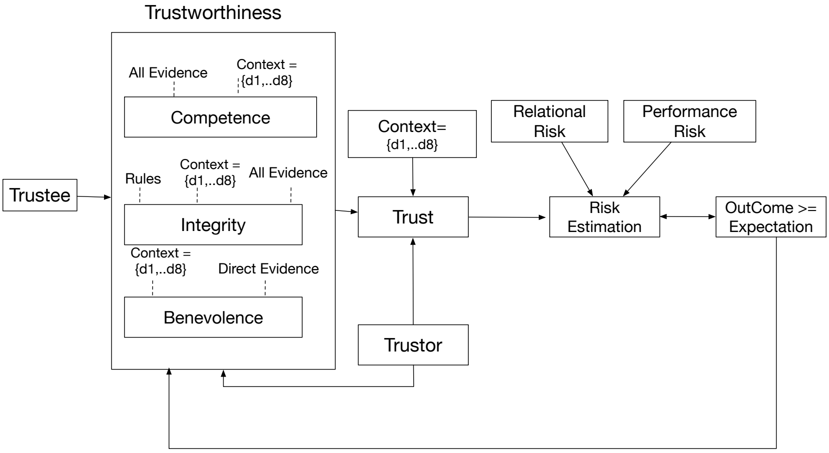


Figure 5.5: Trust framework and its three trustworthiness components. Each component gets the (direct and indirect) evidence and context to calculate the trustee's competence, integrity and benevolence and combines them to evaluate the trustee's trust. Next, the trustor estimates the interaction risk based on relational and performance risk.

- Relational risk. Probability[5] and consequence[6] of not having successful cooperation. Therefore, because of potential opportunistic behavior, the relational risk will increase.
- Performance risk. The probability and consequences that alliance objectives are not realized despite satisfactory cooperation between two partners.

We demonstrate our assumption through the following propositions.

Proposition 5.1: *Benevolent[7] behavior of partners increases trust and reduces former relational risk in the alliance. The benevolent behavior of a member shows that the members will cooperate in good faith, rather than behave opportunistically.*

We formulate this proposition as follows:

$$Rr(x, y, s_i, \tau) \propto 1 - Ben(x, y, s_i, \tau), \tag{5.12}$$

where $Ben(x, y, s_i, \tau)$ is the benevolence of trustee y toward trustor x in situation of s_i.

We define competence as the member's ability within the alliance to perform according to the specified agreement or contract. Higher

[5] Probability is defined as the possibility of an adjusted asset. In the cooperative network, it depends on parameters such as opportunistic behavior (relational risk) of the alliance or commercial/technological/strategic hazards (performance risk).

[6] Consequence is translated as the effect of unfavorable occurrence (like an unauthorized usage of resources) on the organization revenue.

[7] Some of the scholars consider faith and good intentions instead of benevolence.

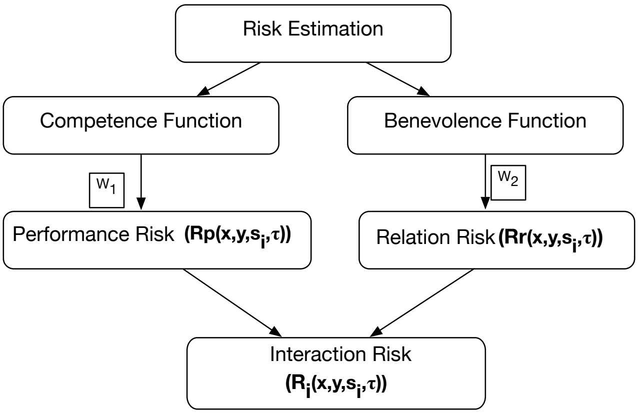


Figure 5.6: Functional relationships between the modules of the SCTM model that are used to estimate the interaction risk. The interaction is a sum of the performance risk and interaction risk.

competence will consequently result in a lower performance risk. Competence is key to achieving the goal of the alliances. Therefore, we formulate the following proposition to show this relation.

Proposition 5.2: *The performance risk will be reduced if the competence of the given member is high. High competence gives a sense of confidence that a member is capable of accomplishing a given task successfully, which can be observed from the available evidence on the given member. Moreover, a member with a high competence suggests a high probability of performing the given task successfully, which is tantamount to low performance risk.* Proposition 5.2 can be represented as:

$$Rp(x, y, s_i, \tau) \propto 1 - Com(nbr_{y \setminus x}, y, s_i, \tau), \tag{5.13}$$

where $Com(nbr_{y \setminus x}, y, s_i, \tau)$ is the trustee y's competence to perform the given (sub-)task in situation s_i. Competence is concerned only with the ability of a member to perform the task, not the member's intention to do so.

The interaction risk value for each task is calculated as the summation of consequences[8] and likelihood values of competence and benevolence of the given member. By given the benevolence and competence of a trustee, we estimate the relational and performance risks, and we derive the total interaction risk. Eq. 5.14 defines the interaction risk $Ri(x, y, s_i, \tau)$:

$$Ri(x, y, s_i, \tau) = w_1 Rr(x, y, s_i, \tau) + w_2 Rp(x, y, s_i, \tau), \tag{5.14}$$

[8] We consider the weighting factors for each risk.

where $Rr(x,y,s_i,\tau)$ denotes the relational risk and $Rp(x,y,s_i,\tau)$ is the performance risk. Therefore, by defining the weighting factors w_1 and w_2 to each risk component, we can emphasize on the main discriminator's factor. Marsh [92] described how decision-makers apply weighting such as potential gains and losses to estimate the risk (i.e., consequences). In the SCTM model, we assign two weight factors to the risk components. The weight w_1 indicates the opportunistic behavior and the poor performance is indicated by w_2.

Therefore, we define the interaction risk $Ri(x,y,s_i,\tau)$ in the alliance as the summation of two defined risks:

$$Ri(x,y,s_i,\tau) = \alpha(1 - Ben(x,y,s_i,\tau)) + (1-\alpha)(1 - Com(nbr_{y\setminus x},y,s_i,\tau)), \quad 0 \leq \alpha \leq 1 \tag{5.15}$$

Here, $\alpha = w_1$ and $1 - \alpha = w_2$. If we set $\alpha = 1$, that means that the weight of poor performance risk is 0, and the SCTM model uses the relational risk to estimate the interaction risk in situation s_i. On the other hand, if we set $\alpha = 0$, the SCTM model uses the performance risk to estimate the interaction risk $Ri(x,y,s_i,\tau)$.

5.6.1 *Selecting a right partner to collaborate with*

This section concentrates on selecting the right partner to perform a task. After estimating the interaction risk, competence and benevolence of each member in the alliance, we propose Algorithm 4 for a trustor to select the right trustee to collaborate with to perform a (sub-)task. As mentioned in Section 5.5.1, we evaluated the competence and benevolence of trustees based on all the available evidence by incorporating $d_8(x,y,s_i,\tau)$ about the performance of trustees based on the different tasks. Algorithm 4 applies the competence and benevolence, to select a trustee (domains / members) with the lowest interaction risk for a (sub-)task τ in situation s_i.

Algorithm 4: Selecting the right partner (trustee) to perform a (sub-)task τ in situation s_i

```
Result: Selected trustee and interaction risk value
Input  : trustor $x \in A$
Input  : $w_1 = \alpha$ , $w_2 = (1 - \alpha), s_i, \tau$
St:= null /* Initialize selected trustee.                    */
Rst:= 2 /* Initialize interaction risk between trustor $x$
    and selected trustee.                                     */
foreach $y \neq x \in A$ do
    $Ri(x,y,s_i,\tau) =$
     $\alpha(1 - (Ben(x,y,s_i,\tau)) + (1 - \alpha)(1 - (Com(nbr_{y\setminus x},y,s_i,\tau))$
    if $Ri(x,y,s_i,\tau) < Rst$ then
        St = y
        Rst = $Ri(x,y,s_i,\tau)$
    end
end
return (St, Rst)
```

In the following section, we present mechanisms to calculate the value of competence of a compromised trustee, where a trustee plays the role of a malicious entity in the alliance.

5.7 COMPROMISING THE ALLIANCE

One of the most difficult threats that an alliance has to defend itself against is compromised partners. Like in any social network, one or more partners of such networks may conspire against the interest of the other members. The challenge for alliance partners is to be able to detect and mitigate adversarial behavior in the alliance. Before being detected, other members will perceive a compromised member as being benevolent. Compromised partners can disrupt the entire alliance network by showing undesirable behavior. Like the other members, compromised partners are supposed to act on behalf of the others. Those members may expect certain responses on their requests, and the consequence of not being able to recognize the behavior of compromised partners from trustworthy ones will lead to false trust in compromised partners. Once compromised partners are detected, their trust values will decrease, and their neighbors can use the trust values to avoid cooperating with them in joint tasks [84].

5.7.1 *Calculating the competence of a compromised member*

As we presented in Section 5.5.1, when a trustor wants to calculate a trustee's competence, the trustor will request the available evidence on the given trustee from its direct neighbors. After receiving evidence, he will aggregate all the available evidence according to the defined method in Section 5.5. Given all the available evidence set on trustee y, which is provided by its direct neighbors nbr_y (i.e., nbr_y represents the direct neighbors of y), we use the Welch's t-test [126] to compare two sets of evidence on y. Because two *Kb*s can have different sample sizes, and we want to test whether they have equal mean. Trustor x selects one of the neighbors of y by employing Algorithm 5, to compare two independent nbr_y's *Kb*s. We have to mention that trustor x interacted with the selected neighbor before. Therefore, trustor x can calculate the value of the competence based on the previous interactions.

Algorithm 5: Select a trustworthy neighbor of y, to compare its *Kb* with the *Kb* of trustor x, to capture the differences in evidence on y.

```
Result: A selected trustworthy neighbor of $y$, Trustworthiness
        value
Input: trustor $x \in A$
Input: trustee $y, nbr_y$
Input: $s_i$
$St := null$ /* Initialize selected trustee.                    */
$TW_{st}$ := -1 /* Initialize trustworthiness of selected
   trustee.                                                      */
foreach $z \neq x \in nbr_y$ do
    /* Calculate the value of benevolence and competence
       for trustor $x$ and trustee $z$.                          */
    $Ben(x,z,s_i,\tau) = val_d(Ed(x,z,s_i,\tau;Kb_x))$
     $Com(x,z,s_i,\tau) = val_c(Ec(x,z,s_i,\tau))$
     $TW(x,z,s_i,\tau) = \frac{1}{2}(Ben(x,z,s_i,\tau) + Com(x,z,s_i,\tau))$
    if $TW(x,z,s_i,\tau) > TW_{st}$ then
        $St = z$
        $TW_{st} = TW(x,z,s_i,\tau)$
end
return $(St, TW_{st})$
```

After selecting a trustworthy neighbor z of y, we will compare two samples of evidence, $Ed(z,y,s_i,\tau;Kb_z) \in Kb_z$ and $Ed(x,y,s_i,\tau;Kb_x) \in Kb_x$. We use the Welch's t-test as follows:

$$t = \frac{\overline{Q_1} - \overline{Q_2}}{\sqrt{\frac{{S_1}^2}{N_1} + \frac{{S_2}^2}{N_2}}} \tag{5.16}$$

where $\overline{Q_1} = \overline{Ed(z, y, s_i, \tau; Kb_z)}$ is the mean of direct evidence of trustor z on trustee y and $\overline{Q_2} = \overline{Ed(x, y, s_i, \tau; Kb_x)}$ is the mean of direct evidence of trustor x on trustee y. S_1 and N_1 are the sample variance and the sample cardinality of $Ed(z, y, s_i, \tau; Kb_z)$, respectively. S_2 and N_2 are the sample variance and the sample cardinality $Ed(x, y, s_i, \tau; Kb_x)$, respectively. And, the degrees of freedom ν associated with this variance using the Welch–Satterthwaite equation [117] is approximated by:

$$\nu \approx \frac{(\frac{{S_1}^2}{N_1} + \frac{{S_2}^2}{N_2})^2}{\frac{{S_1}^4}{{N_1}^2\nu_1} + \frac{{S_2}^4}{{N_2}^2\nu_2}} \tag{5.17}$$

where $\nu_1 = N_1 - 1$ is the degrees of freedom associated with N_1. And, $\nu_2 = N_2 - 2$ is the degrees of freedom associated with N_2. Therefore, if the mean values of $\overline{Q_1}$ and $\overline{Q_2}$ are equal then to calculate the competence of trustee y, trustor x can use Eq. 5.6. If the mean values are not equal, then trustor x will decide based on the number of interactions in the Kbs, because the smaller sample size leads to the larger variance. For example, if $N_1 > N_2$ then trustor x computes trustee y's competence by employing $Ed(z, y, s_i, \tau; Kb_z)$. Other advanced techniques such as a designated power and cost level can be used to determine the optimal sample [69]. Benevolence is calculated based on the agent's experiences in previous joint tasks. If a comprised agent has not yet violated his agreement, his benevolence value is still unaffected. Compromised behavior will lead to lower benevolence at the trustor's side as soon as the compromised member starts to perform oscillating attacks, as such behavior implies violation of agreements.

In the next chapter, we will demonstrate how the SCTM model can be used in deciding on collaboration among alliance members. For this demonstration, we used a Belief-Desire and Intention-Agent-Based Modeling (BDI-ABM) based simulation. BDI-ABM is selected because of the following reasons. First, agents are autonomous entities and able to reproduce complex human or system behavior into a scenario. They can adapt to different situations based on their observations. Second, agents have a common goal and to achieve the goal, they are able to cooperate and coordinate with each other.

5.8 CONCLUSION

In this chapter, we answered RQ 3, RQ 3.1 and RQ 3.2 by presenting the SCTM model and its components.

We proposed an evidence-gathering approach by introducing eight dimensions for each context to gather a variety of evidence on a trustee. Two types of evidence are taken into account to evaluate the trustworthiness of a given member. The trust value is computed by three trust factors, namely competence, integrity, and benevolence, where benevolence is computed from direct evidence between a trustee and a trustor gained through direct interactions, and competence and integrity are assessed on the base of the received feedback from the other alliance members (a trustee's direct neighbors). Benevolence and competence are combined to model and estimate the level of risk involved in each context. We introduced the weights to the trust factors to indicate the main trust factor for different scenarios. In the next chapter, we evaluate the performance of the SCTM model through different sets of simulations and the SARNET research environment (emulation).

6 EVALUATION OF THE SCTM MODEL

In this chapter, we present the evaluation of our SCTM model described in Chapter 5. We evaluated the SCTM model by conducting three sets of ABM simulations.

In the first set of simulations, we answer the sub-research question (RQ 3.1), in which we investigate the impact of the individual components of the SCTM model, namely competence and benevolence on the trust value. This set of simulations is described in Section 6.1. The results show that the benevolence and competence are impacted by the evolving relations between any trustor–trustee pair.

In the second set of simulations, we answer the sub-research question (RQ 3.2): "Can the identified risk factors be evaluated by an automated process, and do these factors have a unique impact on the trust value?" To answer RQ (3.2), we integrated the presented risk estimation model in Chapter 5 with the SCTM model in a scenario to select the "right" partner. Through the ABM simulation, we study the behavior of collaborating partners that collaborate in typical cyber-defense tasks. In order to have realistic trust values when evaluating the SCTM component we use the Epinion dataset[1], which also allows us to compare the SCTM performance against two alternative computational trust models, namely the Situation-aware and Social Computational Trust model (SOLUM) and the Selection of Cloud Service Providers model (SelCSP). This set of simulations is presented in Section 6.3.

In the third set of simulations, we answer RQ (4): "How can the computational trust model practically facilitate the selection of partners in the SARNET emulation?". This simulation aims to evaluate the practical applicability of the SCTM model in the SARNET emulation. This simulation was joint work with another researcher that focused on the cyberattacks and cyber defense mechanisms, while we focused on the collaboration, i.e., the social aspect thereof. The SCTM was used in this SARNET simulation to determine the right (capable and effective) partners to collaborate with, in the case of an attack. The SARNET simulation results that included the SCTM model were compared to the results of a simulation that did not include the SCTM model. This comparison shows the effectiveness of the SCTM model in this practical context. This set of simulations are described in Section 6.4.

[1] http://www.trustlet.org/epinions.html

This chapter is based on:

- **Deljoo, A.**, van Engers,T., Gommans, L. & deLaat C., Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners, IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS), (2018) © IEEE.
- Deljoo, A., van Engers,T., Gommans, L. & deLaat C., The impact of competence and benevolence in a computational model of trust, IFIP International Conference on Trust Management, 45-57 (2018) © Springer.
- **Deljoo, A.**, Koning, R., van Engers,T., Gommans, L. & deLaat C., Managing Effective Collaboration in CybersecurityAlliances Using Social Computational Trust, Journal of Annals of Telecommunications [Under Review], © Springer.

6.1 FIRST SET OF SIMULATIONS: THE IMPACT OF BENEVOLENCE AND COMPETENCE ON THE SCTM

In this section, we present the SCTM model. We define a mechanism for evaluating the trustworthiness of a trustee that can be used by the trustor to evaluate trust and make decisions about the future relationship with the trustee. Extracting trustworthiness of the trustee based on Mayer *et al.*'s model as described in [93] has been only implemented by only a few scholars such as Urbano and Guo [61, 125]. Most of these computational trust approaches estimate trustees' trustworthiness using individual items of evidence about these trustees' behavior in the past interactions, either with the trustor or with third-party agents [2, 96, 113]. However, none of these approaches are able to estimate the benevolence of the trustee. We claim that understanding the benevolence and competence of the trustee towards the trustor at the moment of the trust decision is fundamental for accurately estimating the latter's trustworthiness. Therefore, we present the following proposition:

Proposition 6.1: Close and long-term relationships have a direct impact on the competence and benevolence of partners. With this in mind, we present the main hypothesis of this work as follows:

Hypothesis 6.1. The extraction of benevolence-competence based on the information from the set of evidence on the trustee under evaluation and its use inadequate stages of the relationship between trustor and trustee shows that trustee's trustworthiness improves by increasing the number of interactions between trustor and trustee.

6.1.1 *Simulation setup*

In this set of experiments, we want to test **Hypothesis 6.1**, which we formulated as follows: The extraction of benevolence–competence based on the information from the set of evidence on the trustee under evaluation and its use inadequate stages of the relationship between trustor and trustee shows that the trustee's trustworthiness improves by increasing the number of interactions between the trustor and the trustee.

The experiments were conducted in the Jadex [20] environment. This collaborative network is comprised of participating organizations shown in Fig. 6.1 to represent our network in adequate stages. Each node represents an autonomous organization that needs to trust other parties and share sensitive information with them. For simplicity, we assumed that there is only one task being negotiated by all nodes that mitigate an attack and share the attack information with other parties[2].

We define four different sub-tasks as: τ_1 to monitor a certain type of network traffic, τ_2 to providing resources, τ_3 to block certain IP numbers, and τ_4 to give the information about the current status of the network.

This model starts after the establishment of an agreement between the trustor and the selected trustee, thus excluding the selection process itself. It focuses on both types of agents' decisions concerning the fulfillment of the established agreement: the trustees may opt to fulfill the agreement (the trustors will report the outcome fulfilled duty with delay (Fdd)) or to delay its realization. Accordingly, the trustors may respond to a delay by either retaliating, denouncing the breach (reporting an outcome violation (V)) or forgiving the contingency (reporting the outcome fulfilled duty (Fd)).

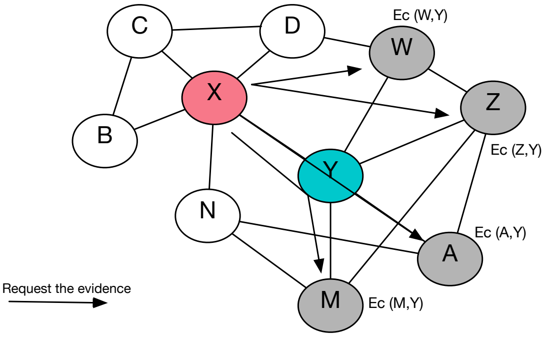


Figure 6.1: Social network schema.

[2] The technical details and code of this research can be found in http://delaat.net/sarnet/index.html

Table 6.1: $Ben(X, Y, s_i, \tau)$ evaluation for the number of rounds

No. of rounds →	20	50	100
$Ben(X, Y, s_i, \tau)$	*0.22*	0.31	**0.86**
SD	*0.113*	0.105	**0.081**
M	*0.762*	0.777	**0.810**

6.1.2 *Result and Discussion*

Our result consists of two parts. First, we calculate the benevolence of the trustee (agent) Y by considering all the evidence (i.e., direct interactions) that X has on Y. Second, we evaluate the competence of the given trustees A, Z and y in four mentioned situations from the trustor's X view. Hence, we perform four different types of situations simultaneously, each with six agents. We assume that agents are honest and there is no conflict on the evidence and messages are encrypted (the interminable agents cannot manipulate the message). In order to compare all approaches, we measure and average the number of agreements with outcomes Fd, Fdd and V.

We are able to calculate the benevolence and competence of each trustor by equation 5.5 and 5.6. To calculate the $Ben(X, Y, s_i, \tau)$, we extract all the evidence that X has on Y. We perform the simulation for 20, 50 and 100 rounds of interactions. We also consider that each agent can freely fulfill its duty, fulfill duty with delay and violate the agreement. We have summarized the result in Table 6.1 (including the mean, M, and standard deviation, SD).

To evaluate the competence function, we select three agents A, Z and Y from the set of agents and calculate the competence of these three agents from the trustor X perspective. Agent X will collect the evidence by sending a query to each agent's direct neighbors and asking their opinions. For example, all the available evidence on Y is collected from its neighbors, (which are expressed M, A, Z, W) is reported in Table 6.2. To calculate the competence of A, Z, and Y, agent X will perform the same procedure similar to agent Y. The simulation has been repeated for four different situations and three different rounds. As a result, we compared the mean (M) and standard deviation (SD) and the competence of each trustee for different rounds and the details are shown in Table 6.3 and Table 6.1. In Table 6.3 and Table 6.1, the highest values are in bold and the lowest values are in italics.

Table 6.2: Agent X asks the direct neighbors of agent y about its performance in the four situations.

Situation →	s_1	s_2	s_3	s_4
Z	Fdd,Fd	Fd,Fd	Fdd,Fd	Fdd,Fdd
M	Fdd,Fd	Fd,Fd	Fdd,Fd	Fdd,Fdd
W	Fdd,Fd	Fd,Fd	Fdd,Fd	Fdd,Fdd
A	Fdd,Fd	Fd,Fd	Fdd,Fd	Fdd,Fdd

We notice that in Table 6.1 the benevolence of the trustees increases as the number of interactions increases. For instance, with only 20 rounds, when the number of interactions between any trustee–trustor is not large, the benevolence is small. By increasing the number of rounds, the benevolence increases significantly. Indeed, this confirms that the number of interactions are, in fact, impacting the benevolence existing between any pair of trustor–trustee.

In the case of competence (see Table 6.3), we also observed the same behavior from the simulation. The competence of the agent is influenced directly by the number of interactions.

As we mentioned before, the simulation has been repeated for 50 and 100 rounds. The benevolence value reaches the maximum value of one by increasing the number of interactions between partners and the partners that are considered to be in a close relationship (see Proposition 6.1). Indeed, each trustor can conclude the trustworthiness of a trustee in each round and make a decision. Overall, we were able to confirm the truthfulness of Hypothesis 6.1.

6.2 SECURE AUTONOMOUSE RESPONSE NETWORK (SARNET) IMPLEMENTATION

Koning *et al.* use the SARNET emulation to create a multi-domain overlay network using virtual machines (VM) and virtual network (VNET) functions [78, 79]. SARNET's domain-agents in the overlay cooperate by requesting certain tasks to be executed by other members in response to an emulated attack. Each domain acts autonomously using its own agent that uses the domain's resources to defend against attacks. Since defending against distributed attacks benefits from cooperation,we decided to facilitate this by applying the model in this chapter to the defense against a Distributed Denial of Service (DDoS) attack, which now asks help from its most trusted nodes, using Algorithm 4 presented in Chapter 5. The SARNET emulation is a framework for the

Table 6.3: Competence evaluation for number of agents A, Z and Y and the number of rounds. SD and M are the mean and standard deviation of Kb_Y.

No. of rounds →	20	50	100
$Com(Y, A, S, \tau)$	*0.21*	0.40	**0.65**
$Com(Y, Z, S, \tau)$	*0.28*	0.43	**0.88**
$Com(Y, X, S, \tau)$	*0.18*	0.33	**0.54**
SD_A	*0.081*	0.113	**0.105**
SD_Z	*0.067*	0.095	**0.091**
SD_Y	*0.090*	0.074	**0.069**
M_A	*0.817*	0.803	**0.770**
M_Z	*0.835*	0.850	**0.845**
M_Y	*0.762*	0.777	**0.735**

detection and mitigation of attacks on network infrastructures[3]. To perform the detection and mitigation, SARNET provides multiple defense strategies, where the SARNET selects a strategy with the highest efficiency and executes it (see Koning *et al.* [79] Chapter 4, Section 6). A defense consists of multiple tasks. These tasks can be performed locally at the domain itself, or the domain delegates the task to other members. In this respect, we identify two distinct types of tasks for the SARNET alliance members during the attack period, *informative* and *executive tasks*. The *informative task* concerns the behavior of the given member, while the *executive task* represents the ability of the given member to perform the task. The informative task is based on the information flow and requesting threat information locally or delegating the responsibility to a member to act on its partner's behalf. On the other hand, executive tasks are the actions and tasks that need to be performed in order to have an effective collaboration in the defense and mitigation of the attack. In the SARNET emulation, each domain has its own SARNET agent, which is responsible for coordinating the activities such as sharing information among the members in the alliance.

- *Informative task*: provides and responds to the requested information.
  - *Informative tasks performed locally*: trust a member to identify a traffic source or request information.

[3] The SARNET has been implemented on the top of VNET to provide a testbed for the research in the secure networking domain.

  - *Informative tasks delegated to a member*: trust a member to continuously send threat information.
- *Executive task*: trusts a member to implement a countermeasure to reduce the impact of an attack.
  - *Executive tasks performed locally*: trust direct neighbors to perform actions.
  - *Executive tasks delegated to a member*: trust a member to act on your behalf.

The SARNET domain agents' behavior can be changed according to some pre-defined parameters. For example, we can give the agents a pre-filled database of evidence, or, set the probability that an agent executes the task. Table 6.4 shows the pre-filled values that we used for the SARNET agents.

Table 6.4: Pre-filled evidence and the success probability of executing a task are given as an input in the SARNET emulation.

member	Fd	Fdd	V	success probability
M	3	1	1	0.0000001
Y	1	1	1	1.0
Z	1	1	1	0.6
A	1	1	1	0.1
W	1	1	1	0.0000001

6.3 SECOND SET OF SIMULATIONS: EVALUATION OF THE SCTM MODEL USING THE EPINIONS DATASET

The second set of simulations evaluates the SCTM model's performance to select the "right" partner to perform a given task. For this demonstration, we used a Belief-Desire and Intention-Agent-Based Modeling (BDI-ABM) simulation. BDI-ABM is selected because of the following reasons. First, agents are autonomous entities and able to reproduce complex human or system behavior into a scenario. Second, they can adapt to different situations based on their observations. Third, agents have a common goal and to achieve the goal, they are able to cooperate and coordinate with each other.

6.3.1 *Result and Discussion*

We implemented the proposed framework using BDI-ABM in the Jadex platform [20]. The Jadex platform provides two levels of security mechanisms: the platform and application level [68]. The platform level is concerned with mechanisms for protecting a platform against unauthorized access, and the application level deals with security aspects for services. The Jadex framework can guarantee a decentralized environment where different platforms can connect while the authorization is involved. Therefore, we built a case study using the features of the Jadex platform, which has the main characteristics of a typical alliance.

Each domain (member) in an alliance network (see Fig. 6.2) is configured in one Jadex platform and different services have been defined for the domains. The alliance topology is known to the members, therefore, each domain knows its direct and indirect neighbors. Each domain is able to receive a message from both direct and indirect neighbors. However, the monitoring service is only available to the direct neighbors. Each member has a knowledge base (*Kb*), which contains the interactions with the other members. In a *Kb*, we store message Ids (Id of the requester, Id of the recipient), a (sub-)type, the time when the requester sends a request ($t_{request}$), and the time on which the requester receives a message from the recipient – the report time (t_{report})– then we can calculate the outcome of the corresponding (sub-)task(s).

Since evaluating a domain's trustworthiness depends on fulfilling the agreements within a time window (Δt_w), we need to implement a method of time tracking the messages. The Jadex framework supports an asynchronous programming style with a future abstraction, where the future allows the requester to check if the result of the (sub-)task is available to the requester [20].

To do the time tracking, we need to map the request time and report time to the message Id. We include the Id of the messages between a sender and a recipient. Therefore, the sender and the receiver of the message can look up the corresponding request in its *Kb*. Each domain is able to monitor the actions of its direct neighbors using the monitoring service provided by the Jadex framework.

To evaluate the SCTM model, we define four different sub-tasks. Each sub-task evaluation was repeated ten times and averaged.

6.3.2 *Simulation Setup*

We have set up an alliance network– collaborative network of organizations– shown in Fig. 6.2. Each trustee represents an autonomous organization that needs to build trust with other parties and share incident information within the alliance network based on the agreement among the members. We considered only one task being negotiated by all members, which mitigates and defends the alliance against a certain attack type (d_5).

Let us consider that at present, six members of the SARNET Alliance collaborate to perform the task τ (defend and mitigate the attack). The members are denoted as $N, X, Y, M, A, Z, C, B, D$ and W. Assume that member N is under attack and requests its neighbors to perform

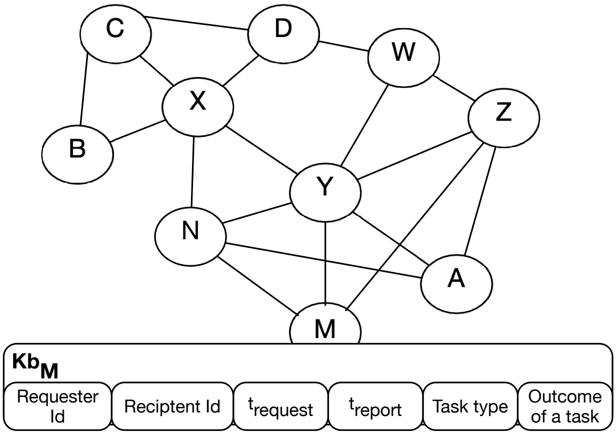


Figure 6.2: In the network, each trustee has a *Kb* which contains the message Id, the direct and indirect neighbors' identifiers, the task type, $t_{request}$, t_{report} and the outcome of a (sub-)task. We extract the related evidence to calculate d_8, indirect evidence (Ec), and direct evidence (Ed) from the trustees' *Kb*s.

certain tasks during the attack. N wants to choose ideal domains for collaboration in order to mitigate and defend against the attack. The task contains four different sub-tasks $\tau_{s1} \ldots \tau_{s4}$, where the sub-tasks are: τ_{s1} provides resources within a specific time window, τ_{s2} monitors certain traffic, τ_{s3} blocks a certain link, and τ_{s4} implements a certain countermeasurement. As mentioned in Section 5.4.1, each sub-task contains the type d_5, deadline d_7 and outcome d_8 of a sub-task. We calculate outcome d_8 of each of sub-task by using Algorithm 3 presented in Chapter 5.

In the network, each trustee has a *Kb* which contains the message Id, the direct and indirect neighbors' identifiers, the task type, $t_{request}$, t_{report} and the outcome of a (sub-)task. We extract the related evidence to calculate d_8, indirect evidence (Ec), and direct evidence (Ed) from the trustees' *Kb*s. There are multiple entries in the trustees' *Kb* s from all its direct neighbors for each sub-task. Fig. 6.2 presents the network schema of the SARNET Alliance[4]. Table 6.5 gives the details of the settings of the simulation. We use the BDI-ABM simulation to test proposition *5.1* and proposition *5.2*, presented in Chapter 5, Section 5.6. The SCTM model focuses on three possible outcomes for a trustee, being the agent who supposes to respond to the trustor's requests. This trustee may fulfill its duty according to the agreement (Fd) or fulfilled the agreement with delay (Fdd) or retaliate, resulting in a breach of the agreement (V). The trustor can decide that delivery with delay should

[4] The source code for the SARNET Alliance can be found in https://github.com/Adeljoo/Collaborative-network.

Table 6.5: Simulation settings and their illustrations.

Parameters	Values	Illustrations
A	Fixed	number of agents in an alliance
τ	Fixed	task type (defend and mitigate the attack)
$t_{request}$	Initiate the simulation	Request time
t_{report}	Receive the feedback on the request	Report time
Δt_w	10 s	Time window
α	0.3	Weight factor
s_i	4	$s_i \in \{s_1, ..s_4\}$
τ_{si}	4	$\tau_{si} \in \{\tau_{s1}, ..\tau_{s4}\}$

be treated as delivery Fd, a failure V or it can leave it as is for Fdd when communicating about the experience to others.

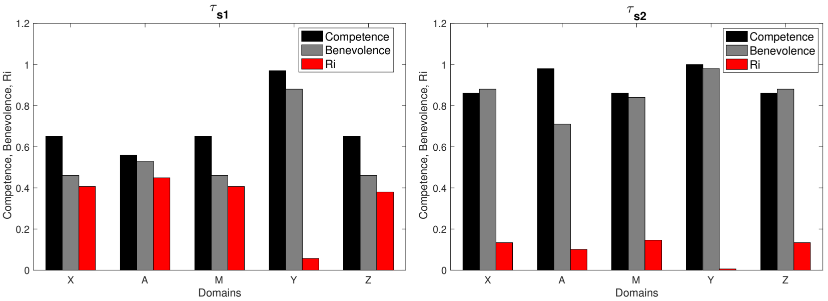


Figure 6.3: Variation of competence, benevolence, Ri of domains X, A, M, Y, Z under sub-tasks: τ_{s1}= provide resources within a specific time window and τ_{s2}= monitor certain traffics.

To select the right partner, the SCTM model computes the competence (see Chapter 5, Section 5.5.1.2) and benevolence (see Chapter 5, Section 5.5.1.1) of each member (the trustee) for each τ_s. The interaction risk is evaluated using Eq. 5.15 presented in Chapter 5, Section 5.6. As it was described in Section 5.4.1, Chapter 5, to evaluate the trustworthiness, multiple parameters of the context are required. However, to estimate trustworthiness, we do not have to define all these parameters, since the majority of these parameters are subjective or selective – they are given to a model from initially defined inputs set.

As explained before, one of the goals of our trust framework is to estimate the interaction risk Ri for any alliance's member based on the member's benevolence and competence value.

Based on the given (sub-)task such as $\tau_{s1}, \tau_{s2}, \tau_{s3}$, and τ_{s4}, the SCTM model recommends the member to collaborate with those who have the minimum interaction risk. In Fig. 6.3 and 6.4 the interaction risk Ri,

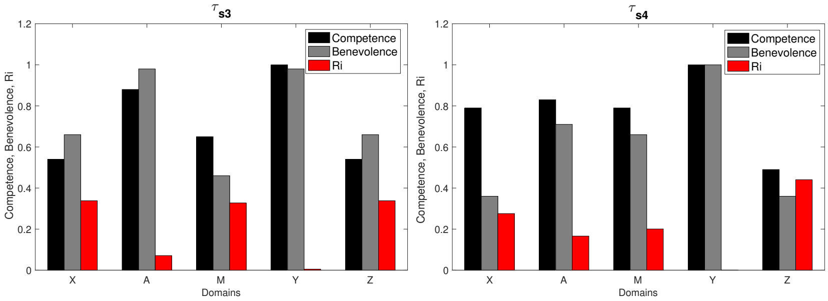


Figure 6.4: Variation of competence, benevolence, *Ri* of domains X, A, M, Y, Z under sub-tasks: τ_{s3}= block a certain link and τ_{s4} = implement a certain countermeasurement.

the benevolence, and the competence values for $\tau_{s1}, \tau_{s2}, \tau_{s3}$, and τ_{s4} are presented.

By considering the recommendations by the SCTM model for the member with the minimum interaction risk, the following members will be chosen for the sub-tasks:

- Y, for sub-tasks $\tau_{s1}, \tau_{s2}, \tau_{s3}$, and τ_{s4}.
- X, Y, Z, A, for sub-task τ_{s2}.
- Y and A, for sub-task τ_{s3}.
- Y and A, for sub-task τ_{s4}.

We can observe from Fig. 6.3 and 6.4 that domains with the higher benevolence and competence values having the minimum risk and the SCTM model will recommend them to other alliance members.

As illustrated in Fig. 6.3 and 6.4 the risk value rises/falls if benevolence and/or competence of domains decreases/increases. However, the significant observation is that even if two members have the same competence value (e.g., M and X in Fig. 6.3 for τ_{s2}), the one with higher benevolence value results in the lower interaction risk. Therefore, Z is selected to perform τ_{s2}. Another example of such behavior is seen in Fig. 6.4 for τ_{s4}, where X and M have similar competence values. We conclude that for interactions, benevolent members with less competence are favorable over those (non-benevolent) with higher competence. This is explained by the fact that competent members with lower benevolence values are capable of damaging their partner with higher impact than those who are benevolent but incompetent.

6.3.3 *Evaluation of the SCTM model with the Epinions dataset*

The evaluation section's major objective is to validate, analyze and compare the proposed SCTM with other trust models. The validation aims to compare the benevolence and the competence values generated by the SCTM model with those generated by some existing trust models. We have selected the SOLUM model presented by Urbano *et al.* [125] and the SELCSP model introduced by Ghosh *et al.* [53], trust models based on some similarities with the SCTM model. Both the SOLUM and the SELCSP model are based on social trust. The SOLUM [125] model calculates trust by combining ability, integrity, and benevolence. The SELCSP model [53] takes competence and trustworthiness as two trust factors and combines them to evaluate a member's trust. The SCTM model has the following advantages compared to the SELCSP model and the SOLUM model. First, we present the context definition with eight dimensions of context and consider different stages of relationships for the competence function. Second, we introduce the risk estimation approach through the SCTM model.

All the three models aim to evaluate the trust values for trustees, and then provide a recommendation to trustors, therefore, we chose them to validate the SCTM model result. We apply three trust ranking-based measures to investigate the performance of the trust models. These ranking-based measures are precision and recall (cutoffs at 5 and 10, and denoted by $Pre@5/10$, $Rec@5/10$) and mean average precision (MAP) [60, 120]. The higher the value of these measures indicate the better performance of the models. We use the Epinions dataset [135] to validate our model. The validation will show the performance of the SCTM model against the SOLUM and the SELCSP model.

6.3.3.1 *Epinions*

Epinions is one of the most popular datasets[5] that consists of a pool of individual reviewers who write a review on products/items and assign a trust value between -1 and $+1$ towards other users and a feedback rating between 1 and 5 for the products/items to facilitate the decision-making process for the users to select the most trustworthy product from the website. To have the trust value between $[0, 1]$ (see Table 5.2, Chapter 5) and deal with the trust value -1, we shift the rating values to $1-6$ then normalize the values to $[0, 1]$ divided by the maximum value 6[6]. The Epinions dataset contains $49,290$ users, $1,139,738$ items, and $841,372$ statements ($717,667$ trusts and $123,705$ negative experiences).

In the SCTM model, we estimate the trustworthiness of a member (a trustee) based on direct and indirect evidence from the trustee's neighbors. We assign the value of trust to each trustee and recommend

[5] *http : //www.trustlet.org/extended_epinions.html*

[6]

$$Trust_{SCTM} = \frac{Trust_{Epi} + 1}{2} \tag{6.1}$$

Table 6.6: The Trust Ranking Performance of All The Trust Models In Epinions.

	Pre@5	*Pre*@10	*Rec*@5	*Rec*@5	**MAP**
The SELCSP model	0.3010	0.1865	0.4840	0.5823	0.5591
The SCTM model	0.3556	0.2390	0.5520	0.6243	0.5641
The SOLUM model	0.3509	0.2061	0.5478	0.5824	0.5595

a trustee with the maximum value to a trustor. In addition, the SCTM model aims to recommend a member (a trustee) to collaborate based on the interaction risk. By considering these similarities, we chose the Epinions dataset to validate the SCTM model. The Epinions dataset reflects the real-world rating scenarios and often adopted by previous studies [53, 60, 61].

To validate our model, we select five arbitrary items from the Epinions dataset. These five arbitrary items correspond to the members X, Y, Z, A, and M considered in our case study (see Section 6.3.2). Each item is rated by multiple users, by a value between $1 \ldots 5$. We adopt these rating values in our model in the following way: $V = 1$, $Fdd = 2$ and $Fd = 3, 4, 5$. Trust values of all five items are evaluated using the competence and the benevolence evaluation functions as explained in Sections 5.5.1.1 and 5.5.1.2. Through the SCTM model, we employ the direct and indirect (witness) evidence to compute the trustworthiness (benevolence and competence) value of each trustee (item/domain). As suggested by [53], we also consider the ratings of:

- Items 1 and 2 as direct evidence.
- Items $3, 4$, and 5 as the indirect evidence/referrals, which are available from the trustee's neighbors.

After considering the direct and indirect evidence value in the Epinions dataset, we run the simulation to evaluate the competence and benevolence of the SCTM model, the SOLUM model, and the SELCSP trust model. Fig. 6.5 and Fig. 6.6 compare the results obtained from the SCTM model, the SOLUM and the SELCSP trust models. In Fig. 6.5, the benevolence value of all three models is depicted. From Fig. 6.5 and 6.6, we can observe that in all the domains, our SCTM model yields better results compared to the SELCSP model and the SOLUM model. By proposing a better formalization, employing different types of evidence, and assigning different values based on the outcome of a task to evidence (see Algorithm 3), our model can perform better. On the other hand, the SELCSP model and the SOLUM model work similarly to non-contextual evidence-gathering mechanisms and evaluations.On the other hand, the SELCSP model and the SOLUM model work similarly due to non-contextual evidence–gathering mechanism and evaluation.

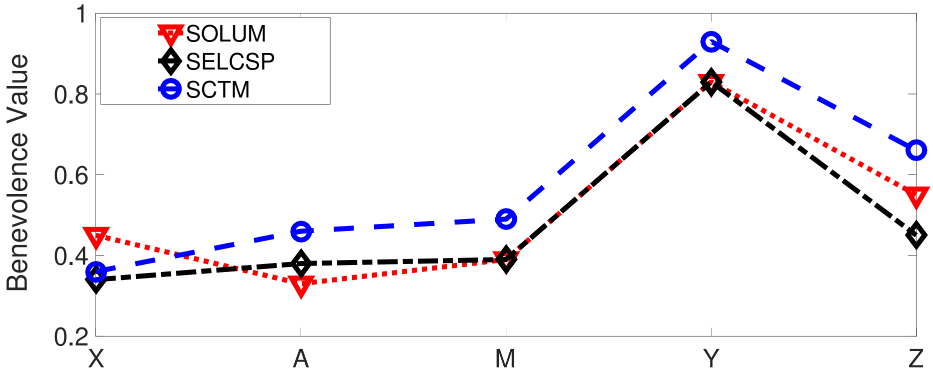


Figure 6.5: Comparison of the value of benevolence generated by the SCTM model with the SOLUM model and the SELCSP model.

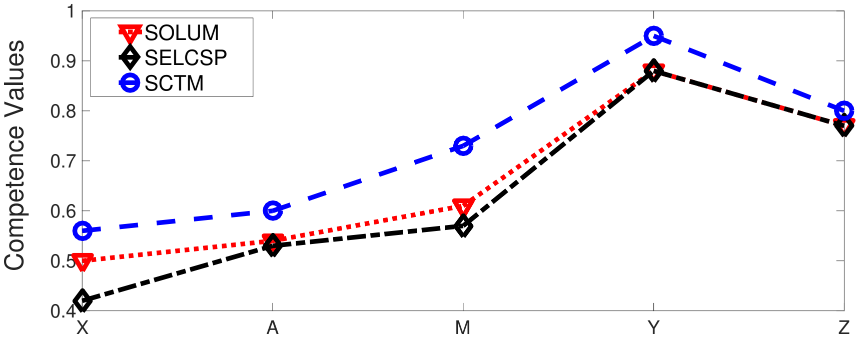


Figure 6.6: Comparison of the value of competence generated by the SCTM model with the SOLUM and the SELCSP models.

Fig. 6.6 shows the competence value based on the indirect evidence for three different trust models. The competence value of the SELCSP is mapped to the SOLUM trust model which makes the SOLUM model performs slightly better. This difference can be related to the proposed scoring algorithm in the SELCSP model. The SELCSP model has a different approach of assigning a value to the evidence, where they give 0 to a negative experience and 1 to a positive experience. Indeed, they eliminated the dis-confirming (bad experience) evidence from their evidence-gathering method in their model. The SCTM model captures better competence values by considering different stages of relation and context in its evaluation.

We present the result of trust ranking-based measures for the three trust models in Table 6.6.

The SELCSP model performs the worst among all the methods due to a non-context-based trustworthiness evaluation and removing the bad experience from the set of evidence. The SOLUM model works better than the SELCSP model by adopting the four dimensions of context. Finally, our SCTM model outperforms the SOLUM model. In other words, we propose a better formalization of trustworthiness factors and context-based evidence.

6.4 THIRD SET OF SIMULATIONS: EVALUATION OF THE SCTM MODEL WITH THE SARNET EMULATION

In this set of simulations, we aim to validate the result of the SCTM model implemented in the simulation environment with the SCTM model's implementation in SARNET emulation. Ralph Koning developed the SARNET emulation and presented the details of implementation in his thesis [77]. In the following, we will present the simulation setup for both environment and case study to validate the SCTM model.

6.4.1 *Simulation Setup*

To validate and test the SCTM model, we implement the SCTM model in two distinctive environments. First, we use an Agent-Based Model (ABM) to setup the SARNET Alliance network[7], shown in Fig. 6.7 in the Jadex platform [20], where the nodes represent the alliance members. Second, the SARNET emulation, as described in Section 6.2, is used to validate the SCTM model.

Basically, in the case of attacks such as a DDoS, collaboration and coordination amongst the organizations is essential to save their businesses and mitigate the attack on time with the minimum cost [25]. In the case of a DDoS attack, the victim can start to mitigate and defend against the DDoS attack locally within its own domain or delegate the responsibility to a member [25, 71]. In both cases, the victim needs to trust a member to perform the given action or provide the requested information. In our simulation, we present the evaluation and ranking algorithms that evaluate and rank the members based upon their shown behaviors, capabilities, and integrity. The ranking algorithm will help the victim to select the right partner to resolve the situation. We implemented the partner selection algorithm that uses the SCTM model in two different environments (simulation and emulation) and studies their behaviors.

[7] The technical details about the SARNET project can be found in http://delaat.net/sarnet/index.html, and the code is published in https://github.com/Adeljoo/Collaborative-network

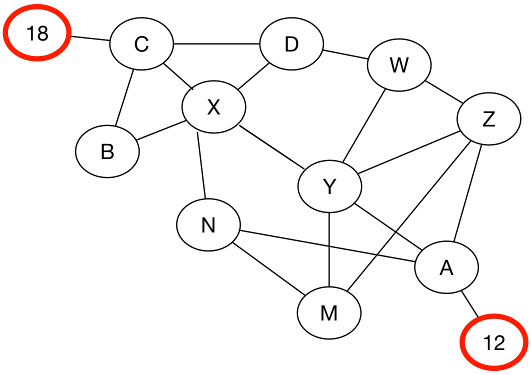


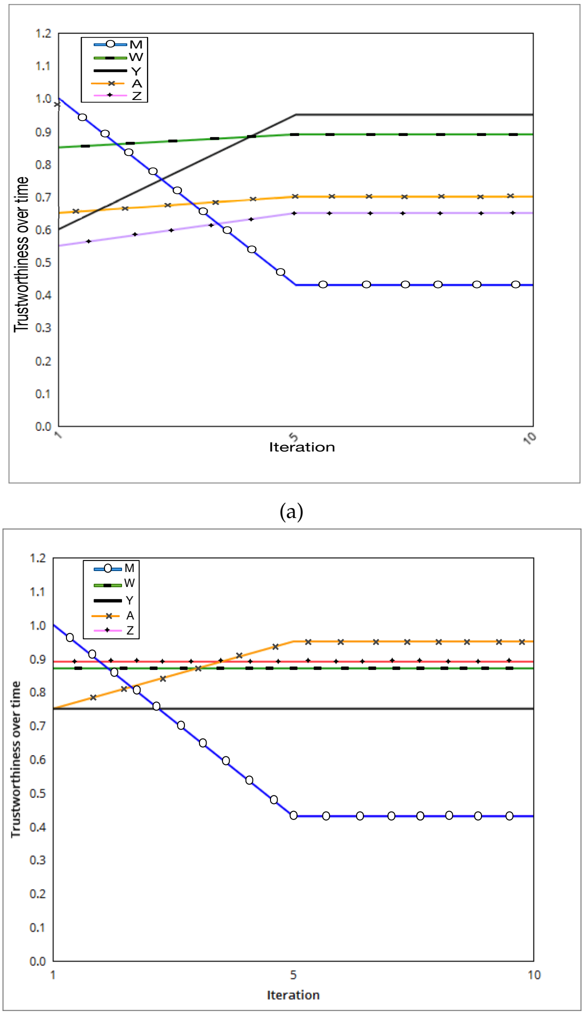
Figure 6.7: The figure illustrates the single attacker close to the victim (“N") and the attacker in position “12"; and the single attacker far from the victim, with the attacker in position “18".

Let us consider the two following scenarios that present two distinct attack situations:

- *single attacker close to the victim (“N”)* and the attacker in position “12";
- *single attacker far from the victim*, with the attacker in position “18".

We implement these scenarios in the Jadex platform. Then, we rank the members based upon their trustworthiness, which computes the trustee’s capabilities, behaviors, and integrity to perform the given context of a task.

In the network, each node has a *Kb* (*Kb* is presented in Chapter 5, Fig. 5.4) which contains the evidence of its interactions with the direct and indirect neighbors. We extract the related evidence to calculate d_8, *Ec* and *Ed* from the node’s *Kb*. For each task, there are multiple entries in the node’s *Kb* from all its direct neighbors. The time, location and complexity of a task is given to the SCTM model from the pre-defined values. We calculate the deadline d_7 and the task outcome d_8 by employing Algorithm 3, which is presented in Chapter 5. Each simulation was repeated 10 times and we gathered the direct and indirect evidence from the *Kb*s. Table 6.7 gives the details of the simulation settings. Fig. 6.7 presents the network schema of the SARNET Alliance and the attackers’ position in the network.



(a)

(b)

Figure 6.8: The members' trustworthiness over time in performing different tasks in the ABM simulation (a) when the attacker is in position “12"(see **??**), “Y" is selected as the right (capable and effective) partner to perform the given tasks; and (b) when the attacker is in position “18", “A" is recommended to the victim to perform the task.

Table 6.7: Simulation settings and their illustrations.

Parameters	Values	Illustrations
A	Fixed	number of agents in an alliance
τ	3	Informative locally
		Executive locally and Executive delegated
N_x	6	Number of entries in the *Kb*s
$t_{request}$	-	Request time
t_{report}	-	Report time
Δt_w	10 s	Time window
t_{delay}	5 s	Acceptable delay
α, β, γ	0.4, 0.7, 0.3[8]	Weight factors
S	2	a set of situations

6.4.2 *Result and Discussion*

The topology shown in Fig. 6.7 is used to implement the scenarios in Section 6.4.1. We illustrate the result of our simulation in Fig. 6.8. The horizontal line indicates the number of iterations and the vertical line shows the members' trustworthiness over time. To calculate the trustworthiness of a member (a trustee), based on the task type (i.e., informative or executive), we extract the evidence from the trustee's *Kb*s and use the presented Eq. 5.9 in Chapter 5 to evaluate the members' trustworthiness based on the task type. Algorithm 5 recommends a trustee, which has the maximum trustworthiness value. We implement two mentioned scenarios (Section 6.4.1 with the following tasks: *informative tasks performed locally*, *executive tasks performed locally*, *executive tasks delegated to a member*. In both scenarios, we have the combination of the informative and executive tasks in our simulation; therefore, the trustworthiness is computed from the available evidence on both local and delegated task outcomes.

Figs. 6.8a and 6.8b illustrate the trustworthiness of each member over time for the two mentioned scenarios in the ABM simulation. Each line represents the trustworthiness value for a member. We take 10 snapshots of the iteration and rank the members based on the evidence for these 10 iterations, at the moment the member starts to perform a task. The different colors or indication symbols represent the different nodes in the alliance. We present Algorithm 5, Chapter 5 to recommend the right partner to collaborate with in an attack situation. Therefore, we compared the result of this algorithm with the results of

[8] The corresponding values for α, β, γ are adopted from Guo *et al.* "An Extended Trust Antecedents Framework for Trust Prediction"[61].

the implementation of the same algorithm in the SARNET emulation. As Fig.6.8a shows, when the attacker is in position “12" (see **??**), member “Y" is recommended to the victim (i.e.,“N") to perform the task. On the other hand, when the attacker is located at position “18" (see Fig. 6.8b), then “A" is selected to help the victim (i.e.,“N").

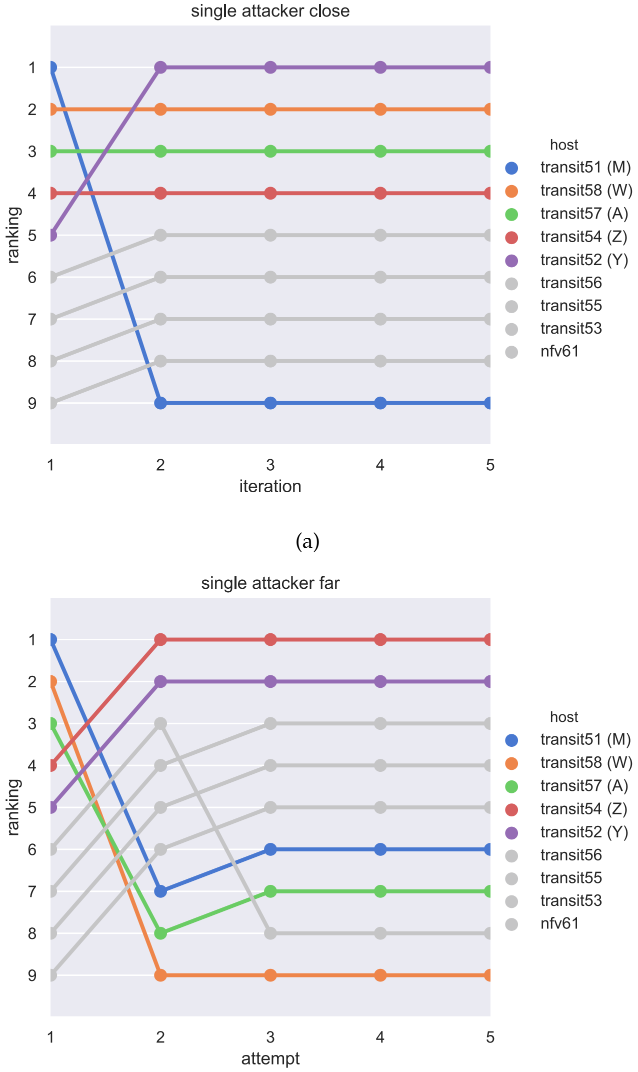
In Figs. 6.9a and 6.9b, shows the results of the implementation of the SCTM model in the SARNET emulation (i.e., explained in Section 6.2), for two mentioned scenarios. The x-axis indicates the iteration numbers and the y-axis shows the trustworthiness of each node in the alliance. As we can observe from the result, when the attacker is in position 12, the SCTM model simulation environment and the implementation of the SCTM model in the SARNET emulation are in good agreement with each other. In the both environments (simulation and emulation), select the member For the second part of the evaluation, when the attacker is far, the SCTM model in the simulation environment and SARNET emulation varies in some cases like member “W". The variations can be explained because of the success probability and pre-defined evidence that is used for each member in the SARNET emulation.

In Koning *et al.* [79], they used the SARNET emulation to study three collaborative non-trust-based defense approaches such as counteract everywhere, minimize countermeasures, and minimize propagation. We compare the performance of these approaches using the efficiency from [78, 105] as a defense performance metric as it includes both the performance loss of each observed metric weighted by their importance and the costs of implementing countermeasures. These approaches were evaluated under ideal circumstances where every member was capable and willing to cooperate.

Yet, as we mentioned in Section 5.3, Chapter 5, members are not always able to perform a given task or may not be willing to take the risk to help. To perform under these circumstances, we developed a new approach, *approach 4*, that uses the SCTM model to select a right partner to perform a task.

To validate the SCTM model in the SARNET emulation, we initialize the SARNET emulation with the values from Table 6.4 to create a sub-optimal situation for collaborative defense, and we evaluate all of the approaches under these circumstances. The implemented approaches are:

- approach 1 - we locate a defense strategy on all the neighbors that sees attack traffic and starts near the victim. Therefore, the defense starts to work till it reaches the attackers.
- approach 2 - As we can observe from the result, when the attacker is in position 12, the SCTM model simulation environment and the SCTM model’s implementation in the SARNET emulation are in good agreement.



(a)

(b)

Figure 6.9: Employing the SCTM model and the partner selection algorithm in the SARNET emulation to rank the members based on their capabilities and benevolence (a) ranking of the members when the attacker is in position “12" and (b) when the attacker is in position “18". The x-axis indicates the iteration numbers and the y-axis shows the trustworthiness of each node in the alliance. As we can observe from the result, when the attacker is in position 12, the SCTM model simulation environment and the SCTM model’s implementation in the SARNET emulation are in good agreement. For the second part of the evaluation, when the attacker is far, the SCTM and SARNET emulation varies in some cases like member “W". The variations can be explained because of the success probability and pre-defined evidence is used for each member in the SARNET emulation.

- approach 3 - we minimize propagation; this approach behaves similarly to approach 1; in this approach, a defense starts at a member, which is the closest to the victim. The difference is, in this approach, we wait for a period to monitor the effect while the victim is still under attack.
- approach 4 - employing the SCTM model in the SARNET emulation to select an appropriate member to perform the given task (i.e., defense on the victim's behalf).

Fig. 6.10 shows the implementation of the SCTM model in the SARNET emulation in comparison to the approaches (1,2, and 3) from Koning *et al.* [79] in the four situations that are described in Koning *et al.* [79] as well.

On the $x-$axis (see Fig. 6.10), *attempt* indicates the defense attempt using the same *Kb*. The different colors or symbols indicate different approaches. For the first defense attempt, when the *Kb* is in the initialized state, the trust-based approach 4 is not as efficient as the other approaches. On the subsequent attempts, when the *Kb* continues to be populated based on the member's actual behavior, we see that the efficiency of the trust-based approach converges and matches the efficiency of the best performing non-trust-based approach.

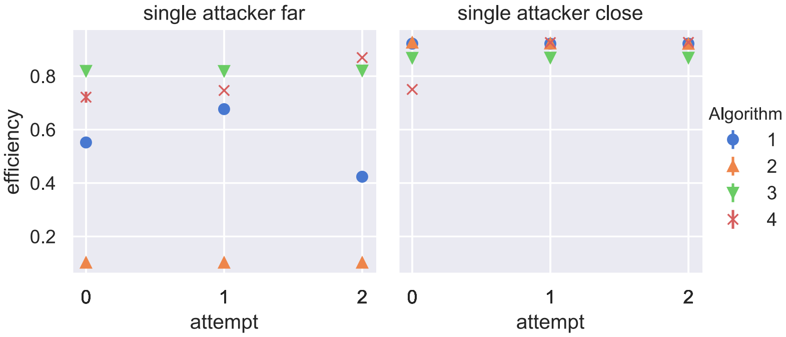


Figure 6.10: Comparison of the trust-based approach (4) (employing the SCTM model in the SARNET emulation) to the non-trust-based approaches (1–3) presented in Koning *et al.* [79]. In "single attacker far" the attacker is connected to the member furthest from the victim. In "single attacker close" the attacker is connected to the direct neighbors of the victim. As expected, the first approach ranks the members based upon the initialization. For the following approaches the ranking evolves until it converges to the set member behavior and become stabilized as expected.

6.5 CONCLUSION

This chapter presented the SCTM model's evaluation to prove its effectiveness in supporting the selection of the right partners for collaborative cyber defense operations and we compared its performance with some alternative computational trust models. We performed this evaluation through a set of simulations, which we presented in this chapter.

In the first set of simulations, we answered the sub-research question (RQ 3.1), in which we investigated the impact of the individual components of the SCTM model, namely competence and benevolence on the trust value. The results showed that the benevolence and competence values are impacted by the evolving relations between a trustor and a trustee.

In the second set of simulations, we answered the sub-research question (RQ 3.2): "Can the identified risk factors be evaluated by an automated process, and do these factors have a unique impact on the trust value?" To answer (RQ 3.2), we integrated the presented risk estimation model in Chapter 5 with the SCTM model in a scenario to select the "right" partner. Through the ABM simulation, we studied the behavior of collaborating partners that collaborate in typical cyber-defense tasks. In order to have realistic trust values when evaluating the SCTM component we used the Epinion dataset, which also allowed us to compare the SCTM performance against two alternative computational trust models, namely the Situation-aware and Social Computational Trust model (SOLUM) and the Selection of Cloud Service Providers model (SelCSP). We demonstrated that our SCTM model outperformed both trust models in terms of trust ranking performance. By employing the Welch's t-test, the SCTM model can detect a malicious member, while the SELCSP and SOLUM models cannot identify such a member.

In the third set of simulations, we answered RQ (4): "How can the computational trust model practically facilitate the selection of partners in the SARNET emulation?". This simulation's goal was to evaluate the practical applicability of the SCTM model in the SARNET emulation. The SCTM was used in this SARNET simulation to determine the right (capable and effective) partners to collaborate with, in the case of an attack. Based on this evaluation, we proposed the ranking algorithm that ranked the members based on their trustworthiness by considering two distinctive tasks: the informative and executive tasks. The SARNET simulation results included the SCTM model were compared to the results of approaches that did not include the SCTM model. This comparison showed the effectiveness of the SCTM in this practical context.

7 CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

Our modern digital society has become increasingly dependent on a well-functioning infrastructure. A downside of this dependency is its vulnerability to cyber attacks. Organizations observe attacks that increase in volume and variety. As hackers collaborate in creating these attacks, organizations also benefit from collaboration to protect their networks and customers. Sharing cyber intelligence (studied in Chapter 2) is key when learning how to most effectively defend against current and future attacks by recognizing its pattern and select the most capable and effective partner(s) (studied in Chapter 5) when creating an automated defense [77].

Sharing cybersecurity information helps organizations create solutions for future attacks. However, establishing collaboration among different organizations facilitating information sharing requires trust, which needs to be organized, maintained, and evaluated. In this thesis, we showed that trust can be computationally described, maintained and evaluated as shown by the frameworks and models described in Chapters 3, 4 and 5. These frameworks help to create collaborative networks, orchestrate collaborative networks, whilst supporting organizations in finding the right partners who must coordinate their tasks that implement mitigation against cyber attacks.

Our exploratory study on the cybersecurity alliances provided us with new and innovative insights that help us define the requirements, technologies, and metrics required to create orchestrated defenses within collaborative networks operated within alliances.

Next to the answers to the research questions, we presented the following scientific contributions:

- A normative Agent-based model N-BDI* to reason about the behavior of alliances' members.
- A simulation environment using N-BDI* to model collaborative networks, part of the simulation contributed to the Digital market place and presented at Amdex group.
- A Social Computational Trust Model SCTM to evaluate trust among alliances' members.
- A risk estimation framework to decide on the most appropriate action.

- The validation of the SCTM model in practice is demonstrated in the operational layer of SARNET for decision-making in a multi-domain defense orchestration.

Using these contributions and the knowledge we gained during the studies performed in this thesis, we can answer the research questions that we described in Section 1.2.

7.2 ANSWERING THE RESEARCH QUESTIONS

The goal of this thesis was to study What dynamic computational trust models facilitate the decision-making process on selecting the right partners for collaborative cyber defense operations? Considering the goal, we first need to understand what trust means in this context, and this leads us to our first research question:

- ***RQ1*** *What does trust mean, and how can the defined concept of trust be applied in collaborative networks?*
  As we assumed that a group of autonomous organizations will hereto create an alliance, therefore, in Chapter 2 we explained the requirements to create such an alliance. To create an alliance, there are three requirements:
  1. creating common benefits for the members,
  2. define what trust means and present a trust framework that a member (trustor) can use to evaluate the other parties' (trustees) trustworthiness, And
  3. a governance model to define common policies.

  To answer RQ1, we provided in Chapter 2:
  - A Trust definition: "*Trust is the willingness* of a trustor to be vulnerable to the actions of a trustee based on the expectation that the trustee will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other parties". This expectation is realized when the given member has:
    * *Competence*: potential abilities to perform a given task (e.g., a task could be to defend against this type of attack).
    * *Integrity*: adheres to a set of rules agreed upon and acts accordingly to fulfill the commitments.
    * *Benevolence*: acts and does good even if unexpected contingencies arise[1].

[1] Acts toward the interest of the alliances.

  - A trust framework: depicted in Fig. 5.1 and presenting how the given member's trustworthiness can be evaluated through the competence, benevolence, and integrity components.
  - A governance model: the governance model is based upon the SPG framework by Leon Gommans [55], which we combined with our trust framework. From this governance model, we derived a control framework that we employed to define a set of rules for the SPG alliance members. In this thesis, we assumed that the policies are defined as the basis for the interaction among the members.

  Once trust is presented during the interaction among the members, we need to understand how to describe the trust interactions, which led us to our second research question.

- ***RQ2*** *How can we model the interactions of a collaborative network?*

  The collaborative network is the result of the interactions among different members. Therefore, it is necessary to describe and model the interactions within the collaborative network. Our studies show (Chapters 3 and 4) that describing and modeling collaborative networks' interactions can contribute to this research question in three ways:

  1. To describe the interactions of the collaborative network, we need to identify factors such as parties (members), roles, activities, and exchanged messages between the parties in the network. After identifying these factors, we used them to model the interactions.
  2. To model the interactions among different parties by employing the identified factors in the first item, the following approach is used. First, we analyzed interactions in the network at the signal layer, i.e., the message exchange between actors. Second, we modeled them with the components of the BDI agent. Third, we identified the implicit actions, intentions, and conditions which are necessary for the interaction to occur. In Chapter 3, we extended the model with a mechanism for expressing both compliant as well as non-compliant behavior of the agents, as alliance members, although bounded by the SPG rules, may, as an autonomous party, have conflicting interests and some may show non-compliant behavior. As a consequence, we need to monitor the members' behavior and be able to identify non-compliment members. To support this, we included some norms, representing guidance to the members of collaborative networks and expressed as rules into the BDI agent reasoning cycle. This leads to

the N-BDI* architecture that includes mechanisms to reason about those rules.
In Chapter 4, we employed the N-BDI* to implement a secure digital market place case study using GDPR as the source of rules and simulated the interaction among different parties. We extended the N-BDI* architecture planner by employing the utility function in the N-BDI* planner. This allows agents to select a plan that maximizes their utility while complying with the GDPR rules simultaneously. The simple and effective N-BDI* reduced the complexity of the normative reasoning and enabled the agents to filter the plans and reason about uncertainty.

3. Throughout the thesis we demonstrated that an agent-based model (ABM) is an effective modeling approach for representing collaborative networks where autonomous members need to collaborate to achieve a goal and where each member has his own desires and goals. Each agent in the society of agents represents a member of the collaborative network which has its own characterizations, goals, and intentions to collaborate with the other parties. Later on in Chapter 5, the members of this collaborative network played the role of trustee and/or trustor.

Through the SPG and the trust framework, we explored and developed case studies that are described in Chapter 3 and Chapter 4. These case studies guided us to model the interaction of the collaborative network and provided an answer to RQ 2. The trust framework that we explained in Chapter 2 has been converted to the computational trust model that we presented in Chapter 5.
After modeling the interactions of a collaborative network, trust emerges from the interactions, and then we needed to express trust in a computational model that members can use to evaluate the given member's trustworthiness and update the trustworthiness over time. Therefore, we defined RQ 3 as follows:

- ***RQ 3*** *How can we express trust among members in a collaborative network in a dynamic computational model?*
  To define a computational trust model and evaluate members' trustworthiness, we examined different trustworthiness factors and we defined the following sub-research questions:

  - ***RQ 3.1*** *What are the trustworthiness factors? Do these factors have a unique impact on the trust value?*

  To answer RQ3.1, in Chapter 5, we formalized and developed the social computational trust model (SCTM). The SCTM model has three independent trustworthiness factors namely: competence, benevolence, and integrity (see Chapter 2). We proposed an evidence-gathering approach by introducing eight dimensions for each interaction context to gather a variety of evidence on a given member (a trustee). To calculate a member's (a trustee's) competence and integrity, we use indirect evidence, while calculating the member's benevolence, direct evidence is used. In the SCTM, we evaluated the member's trustworthiness even when such direct evidence is not available to the evaluating member (the trustor). To gather indirect evidence, the trustor will request the evidence about the given member (trustee) from the trustee's direct neighbors.

The evaluation of the SCTM has been done by different case studies implemented in the ABM simulations, (see Chapter 6). From the obtained results, we were able to conclude that first, each of these trustworthiness factors has the unique impact on the evaluation of the given trustee or in other terms they are independent of each other, in other words, benevolence delineates only the trustee's intention, rather than the trustee's ability to perform the given task. The competence of a trustee refers to the ability of the trustee to perform the task. Thus, we showed that each of these factors has a different impact on the trust value based on the task type. Second, the trust value can be changed over time when new evidence becomes available to the trustor. Third, considering the eight dimensions of the context of the interactions helped the trustor make a better decision even when the direct evidence on the given member was not available to the trustor.

Furthermore, we evaluated the performance of the SCTM model against two other trust models using real-world datasets: SelCSP, and SOLUM. We applied three trust ranking-based measures to investigate the performance of each trust model. These ranking-based measures are precision and recall and mean average precision, and we demonstrated that the SCTM model outperformed both trust models in terms of trust ranking performance.

During our development, we realized that compromised members are one of the most important threats to the alliance. Therefore, we applied Welch's t-test to compare the set of evidence and identified the inconsistencies within the set of evidence. By combining competence, benevolence, and integrity, agents can evaluate a given member's (or trustee's) trustworthiness.

In most of the computational trust models, the risk is considered as a factor that a user must derive from different ingredients recommended by the system. In many cases, the relation between risk and trust are not evaluated [31]. Most of the trust models acknowledge the intuitive observation where the trust and risk are in an inverse relationship, i.e., a low/high trust value is associated with a high/low risk, or, risk and trust pull in opposite directions to determine a member's acceptance by a partner. Therefore, to make a clear distinction between risk and trust and estimate the interaction risk for the collaborative network members, we defined RQ 3.2 as:

- ***RQ 3.2*** *What generic risk factors can be identified, and can they be evaluated by an automated process?*

The SCTM model aimed to help the alliance members select a "right" partner to collaborate with on joint tasks. The term "right" implies that a member has enough benevolence and competence. This reflects a low interaction risk.

In Chapter 5, we identified two types of risks that directly impact the success of a collaborative network. These two risks, relation and performance risk, are combined into the interaction risk, which can be estimated by summing the relation and performance risk. The relation and performance risks are estimated based on a combination of the benevolence and competence value. Regarding the context of interactions, we considered the use of weight factors to show that each risk has a different impact on partner selection in a different context.

By considering the answers on RQ3.1 and RQ 3.2, we can now see how to express trust among the members as a dynamic trust that can be expressed with three independent trustworthiness factors. Thus, the SCTM model contributes to the existing body of knowledge in seven ways:

- By introducing the context, we can gather different evidence on a given trustee and evaluate such a trustee based on the requested evidence.
- Each member in the society has a knowledge-base, which stores its interactions with the other parties and provides a suitable approach to investigate the member's past behavior.
- Trust can change over time, and we prove that by receiving the new evidence, the trustor can reevaluate the given trustee over time.
- The SCTM takes into account the current stage of the relation between the interacting parties, and our simulations show that trust increases as the trustee and the trustor are working closely together, which confirms our intuitions about the development of relationships.
- Trust depends on the context of interactions, and trustworthiness factors are independent of each other. Weight factors

allow us to adjust the relative importance of trustworthiness depending on the context of interactions between the parties.

- The SCTM model recommends that a member collaborates with others by considering the interaction risk and trust value.
- The SCTM model is able to identify the compromised members using the Welch's t-test.

Now that we have defined and validated the SCTM model with the simulation, we need to validate our model in the SARNET research environment, which lead to RQ4:

- ***RQ 4*** *How can the computational trust model practically facilitate the selection of partners in the SARNET emulation?*

In Chapter 6, we evaluated the applicability of the SCTM model in the SARNET Emulation testbed. The computational model was used to determine the right (capable and effective) partner to collaborate with, in the case of an attack. To select the right member,

- first, we evaluated a trustee's competence, benevolence, and integrity based on the different tasks types, and
- second, we ranked the trustees based on the trustworthiness values.

To mitigate and defend against the attack, the member under attack (i.e., victim) asks for help from the set of neighbors and selects the most trustworthy neighbor to perform certain actions. We showed that applying the SCTM model in this way leads to effective mutual defense strategies against cyber attacks in the SARNET research environment (emulation). The obtained results showed that the SCTM model from the simulation and the SARNET emulation are comparable in ranking the members. The ranking of members helps the victim to select capable and effective members based on the task type to collaborate and defend against the attack. The trust based defense strategy is more effective compared to the other defense strategies that Ralph Koning implemented in the SARNET emulation testbed.

7.3 MAIN RESEARCH QUESTION

We asked the main research question:

What dynamic computational trust models enable cyber-intelligence sharing through partner selection for collaborative cyber defense operations?

In this thesis, we presented such a computational trust model and showed that it worked in the practical context of the SARNET project. To achieve the goal of this thesis:

- First, we provided the trust definition and identified the requirements, such as a governance model, to create a collaborative network.
- Second, we modeled the interactions among the members of the collaborative network to understand the behavior of members, to do so. We presented a normative framework called N-BDI*. By using the N-BDI* framework, we were able to model the interaction of collaborative networks and reason about the member's behaviors. We have implemented and tested the N-BDI* framework by employing the agent-based model.
- Third, we proposed, developed, and evaluated the SCTM model. The members of this collaborative network employ the SCTM model to evaluate the other members based on three independent trustworthiness factors and selected the most capable and effective partners based on these factors for the context of the interaction. Through the different case studies and using the real-world dataset, we have proved that the SCTM model out-performs other trust models.
- Fourth, the SCTM model has been implemented and evaluated in the operational layer of SARNET for decision-making in a multi-domain defense orchestration, using the SCTM model to improve the efficiency of the defense strategies.

7.4 FUTURE RESEARCH AND CHALLENGES

Each chapter in this thesis is concluded by a list of research challenges and potential future research avenues. In this section, we explore them in a broader context.

The list can be categorized into three groups: group 1) related to acquiring appropriate datasets to validate the model; group 2) related to the choice of theories; and group 3) related to modeling challenges.

- *The Alliances dataset*

  During this research, we realized that the common benefits of sharing data were never identified. Creating the common benefits for the data providers such as the Cyber threat alliance, Cyber Leadership Alliance, NATO Communications and Information Agency (NCIA), and the various third-party organizations could facilitate data sharing in the cybersecurity domain. In addition, the thesis employed real-world trust-based datasets such as trustlet data to evaluate the computational trust model. Although the trustlet dataset is not a cybersecurity data source, and this is a limitation, we should mention that this dataset is well-known among the trust researcher community. Therefore, it allows researchers to validate and compare their trust frameworks. Providing a reliable and available dataset from the providers would help researchers to understand the incentive behind the cybersecurity alliances and propose greater solutions to the studied problems in this thesis. In the future research, a new dataset from the SARNET alliance can help researchers to further develop their risk mitigation strategies.

- *Modeling research opportunities*

  During the development of this thesis, we proposed the use of the BDI agent and extended the agent planner cycle and added the norms to the belief set of the agent called N-BDI*. Future work can, however, extend the N-BDI* to capture the conflicts of norms.

  Moreover, we were limited in the use of the current version of BDI agents and extend the agents to meet the requirements of the project. Hence, the N-BDI* framework is a general normative framework that can be applied to other normative scenarios. Although, in this research, we did not elaborate on the representational aspects of norms, this topic is addressed by other researchers.
  For future research, topics such as social influences and tipping points that affect the stability of alliances can be investigatedas well as the effects of changes in policies and other norms. During this research, we realized that the alliances' stability depends on well-planned agreements that encourage members to cooperate. For instance, by introducing a normative framework, we were able to monitor the member's behavior and identify the non-compliant member in the network. However, the question is, what is an acceptable number of violators that an alliance can survive and deliver the agreed service? Because the violators will not fulfill their commitments and other parties will lose trust in these violating trustees. Therefore, the stability of alliances will be at risk. More research into applying BDI models aimed at the study of tipping points of collaborative networks seems appropriate.

- *Theoretical Challenges*

  We focused on theories related to trust evaluation and trustworthiness factors in a collaborative network, rather than concepts of authentications and operational approaches, which has been studied in the SARNET project at the tactical and operational layers. This was mainly because trust evaluation was needed as an essential first step to shed light on the concept of creating an alliance. The goal here was to provide solutions to facilitate the collaboration not only as technical solutions but also of solutions and approaches related to the governance of this collaborative network.

Two more general future research directions that we touched upon briefly in this thesis are: first) identified the common benefits for joining an alliance. Some existing model such as value-net can be used as a future direction for this topic to investigate research questions such as how the value creation or identifying the benefits influence the performance of the alliance.

Moreover, this thesis contributes to a new field of research such as trustworthy AI, by looking at different aspects such as trustworthiness factors and governance frameworks, where the N-BDI* and SCTM model can be applied in such trustworthy AI projects. For instance, in data sharing projects such as DL4LD or E-health, selecting the most trustworthy data suppliers and datasets is an interesting research direction for the developed method in this thesis to be employed. The study on the Secure Digital Market Place (SDMP), described in Chapter 4 may help advance all these projects that in essence all are about data sharing within alliances.

Second research direction that could be investigated in the future is the bootstrapping problem of new members to the network, where there is no evidence available on the new member, therefore, the trustors can not evaluate the new member's trust and never will be asked to share or help the other parties in the network.

The obtained results, alongside the potential future research, hold the promise of creating an alliance to effectively increase the information sharing among different parties in the scope of cybersecurity, which has been formulated as the research problem at the beginning of this project, ultimately has a direct impact on the cybersecurity domain.

Part I

APPENDICES

A EXTENSION OF THE BDI AGENT: A CASE STUDY

In Chapter 3 and Chapter 4 we presented the N-BDI* framework. Additional to these chapters, in this Appendix, we present a case study as a proof of concept for the computational model that enables agents to choose an appropriate plan based on received information about the current environment's state. We employed Algorithm 1 presented in Chapter 4 to simulate the BDI agent planner.

A.1 ILLUSTRATIVE USE CASE

The example case is a scenario that goes as follows:
Bob is a security manager at company A. For his company's sake, he is looking for a way to collaborate with Alice, who is a security manager at company B. Alice and Bob that are represented as agents in our simulation, are not part of a collaborative group (i.e., an SP) and have had no previous experience with sharing data before the first iteration. To establish this collaboration, each agent needs to plan its actions based on the estimated risks and benefits, which means maximizing the benefits while minimizing the risks.
Bob has three options. The options are:

1. (Plan A) Give complete access over the company's data to Alice;
2. (Plan B) Request certification from her company;
3. (Plan C) Deny Alice's request.

In this scenario, Bob's goal is "sharing with Alice" and his subgoals are "calculate risks" and "estimate benefits". These risks and benefits are depending on the situation. For example, suppose that Bob and Alice have not yet collaborated before, then Bob would take a risk if he would select plan (A), as he can not be sure about Alice's trustworthiness. Whereas, if he chooses Plan (C) Bob knows right away that he will not be able to gain benefits of this collaboration. One may think that selecting plan (B) is the most appropriate plan for this scenario. However, each plan is associated with a particular response time and requires a different amount of work; e.g., requesting a certification from the company implies completing many processes.

This scenario exposes the following problem: how can agent Bob select the most appropriate plan to achieve its goal based on its current state (that includes what Bob knows of his environment)?

	Plans and sub-plans	Probabilities	Contribution value
		$Pr \to [0,1]$	$val_c \to [0,1]$
Plan A	Give overall access	0.35	0.06
	Start to share data	0.65	0.0
Plan B	Request a certification	0.95	0.08
	Check the certification (if the certificate is valid and the requester (i.e., Bob) is authenticated then start to share data)	0.05	1.0
Plan C	Deny Alice's request	0.40	0.05
	Use the resources for own purpose	0.60	0.0

Table A.1: Plans and sub-plans contributions values and Probabilities.

After selecting the goal an agent selects plan with highest utility from a set of possible plans to achieve that goal. In Chapter 4, we present Algorithm 1 which implements this approach; the algorithm has linear complexity and is simple as well as effective.
Note that more than one plan may have the same plan utility, and in this case, a plan is selected randomly from those with the maximum utility or the agent will select the plan with the least the execution cost.

A.2 EXPERIMENT SETTINGS

Our experiment consists of a simulation of the accumulated satisfaction of an agent after executing a plan to achieve a goal. We compared the plan selection described above against selecting plans randomly from a set of possible plans. The satisfaction of an agent is calculated based on the extend to what his goals are satisfied. Our experiment consists of running a number of iterations in which we perform the following steps.

1. Randomly generate the probability for each event in the interval $[0,1]$.
2. Instantiate the ascribed scenario for each plan, according to the given probability of events (see Table. A.1).
3. Compute the utility for each plan.

For planning we tested three different strategies:

1. The agent selects the plan with the highest utility from the set of alternative plans leading to the same goal.
2. The agent selects a plan randomly from the set of alternative plans leading to the same goal.
3. The agent always selects the same plan over and over again (constant plan selector).

For each plan after execution thereof we store the satisfaction of the scenario associated with the selected plan.

In our experiment, we ran 1000 iterations of the steps described above, each of which takes less than 1 second to run. After obtaining the results, we calculated and compared the average satisfaction and the accumulated satisfaction of all iterations for each plan selector (utility-based selection, randomized selection and fixed plan). The fixed plan selection was split into always "share everything" and always "deny". The average satisfaction, the standard deviation and minimum and maximum values, and accumulated satisfaction obtained this way are detailed in Table A.2. In Table A.2, the highest values are in bold and the lowest values are in italics.

Plans	M	SDV	Min	Max
Randomly	*0.38*	1.59	0.0001	0.44
AskCTA	**0.93**	*0.54*	0.0001	0.98
Deny	0.41	0.72	0.0002	0.21
Share everything	0.53	**2.76**	0.0001	0.60

Table A.2: Satisfaction by the Plan Selector used in our simulation (n = 1000). The utility based plan selector corresponds to AskCTA (which stands for Ask for a Certification) and both Share everything and Deny are based on fixed plan selection. We compare Share everything with Deny as the first satisfies our goal while deny doesn't, while both bring their own risks.

As can be seen in Table A.2 the plan selector with the best results is the plan "AskCAT" using the utility-based plan selector, while the constant plan selector has the worst results with always Deny performing even worse than Share everything. The simulations show that on average the utility-based planner performs best in uncertain situations. However, this is not the case for every individual iteration, since the extended planner selects the plan with the higher expected value, but an undesired event, such as a crash or being selfish, could cause other plans to be more successful. This uncertainty is clearly seen in the results of selecting the "Deny" plan, which is associated with high standard deviation that can also be observed in Table A.2 As a consequence of choosing the fixed plan Deny, the agent may get very satisfied (a very good performance if the transaction would have involved high risks at low costs) or very unsatisfied (as the agent does not gain any benefits).

The impact of using a plan selector over time is shown in Fig. A.1 where the accumulated satisfaction is plotted from the first to the 1000th iteration. The difference in performance between the different plan selecting strategies can be clearly seen as the iterations progress, but during the first iterations, this difference is small, due to the uncertainty of the scenario that arises from selecting a plan.

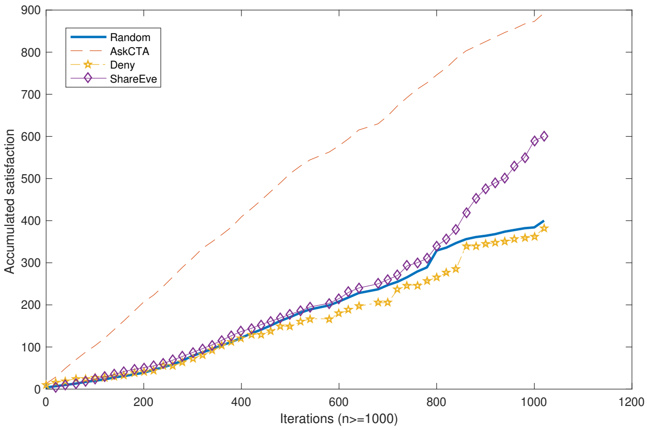


Figure A.1: Accumulated Satisfaction.

We used the Jadex [20] platform to implement the plan selection algorithm[1].

A.3 DISCUSSION

As discussed above, our plan selector performed significantly better compared to the alternative plan selectors. Our experiment helped us to identify a limitation of our approach: the representation of dependent probabilities. This dependency is not captured in the model presented here. Including such dependencies is future work.

In our selection plan algorithm, we do not assume that each action that is executed will succeed. Observations of the agent (monitoring) will enable the agent to learn about its effectiveness. Considering the consequences of other agent's actions in a time-dependent environment is one of the fundamental problems in open systems. The work presented here is just one step towards a model capable of capturing the knowledge necessary for agents to understand *"what is going on"* when they meet each other [111, 112].

A.4 CONCLUSION

Utility-based model development is a promising approach for taking the appropriate action when an agent has to select a plan in uncertain circumstances. In our model we didn't allow for actions to fail. This is a limitation of our model and including probabilities of failure is planned for future research. That research should result in a simple but effective algorithm that chooses a plan based on the plan utility also

[1] This plan selection code can be found in https://github.com/Adeljoo/Planselector

considering an uncertain outcome of the plan execution.
While having these limitations, with our modest first series of simulation experiments we were able to demonstrate the effectiveness of our approach using empirical evaluation.

Part II

APPENDICES

B CASE STUDIES IMPLEMENTATION WITH JADEX FRAMEWORK

In this Chapter, we presented the screen shot of the implemented case studies such as Eduroam and STDMP with the Jadex environment. The case studies is built in Jadex environment and developed in Java by Ameneh Deljoo.

- A complete scheme about the process can be drawn unifying procedural and institutional descriptions which have been shown in Fig. B.1. In this process, the university acts as offeror and a student acts as offeree, we presented the institutional relation that attach to each role in this process.
- In Fig. B.2, we presented the screen shot of the Eduroam case study in the Jadex framework.
- In Fig. B.3, the SDTMP presented in the Jadex framework with different parties and modeling their interactions.

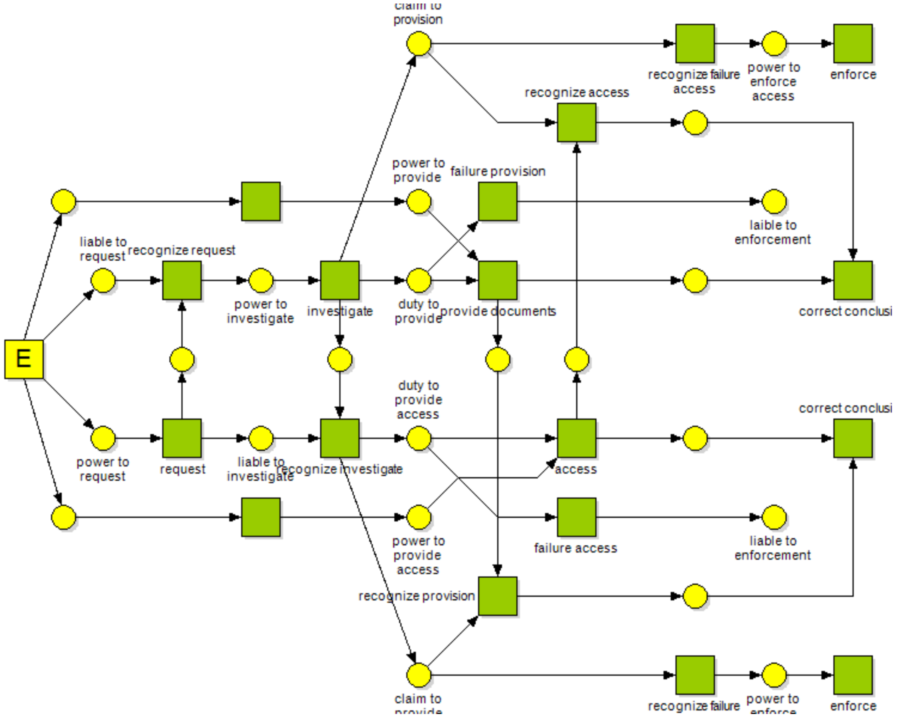
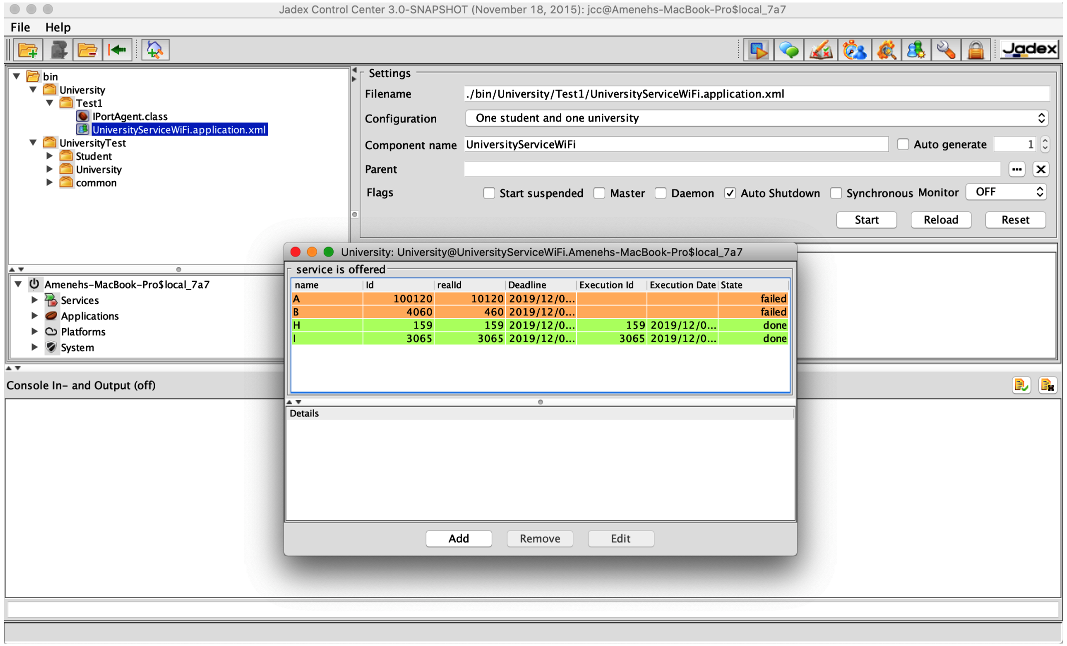


Figure B.1: Full action pattern associated to an Eduroam WiFi connection



Jadex Control Center 3.0-SNAPSHOT (November 18, 2015): jcc@Amenehs-MacBook-Pro$local_7a7

Settings

Filename ./bin/University/Test1/UniversityServiceWiFi.application.xml

Configuration One student and one university

Component name UniversityServiceWiFi ☐ Auto generate 1

Parent

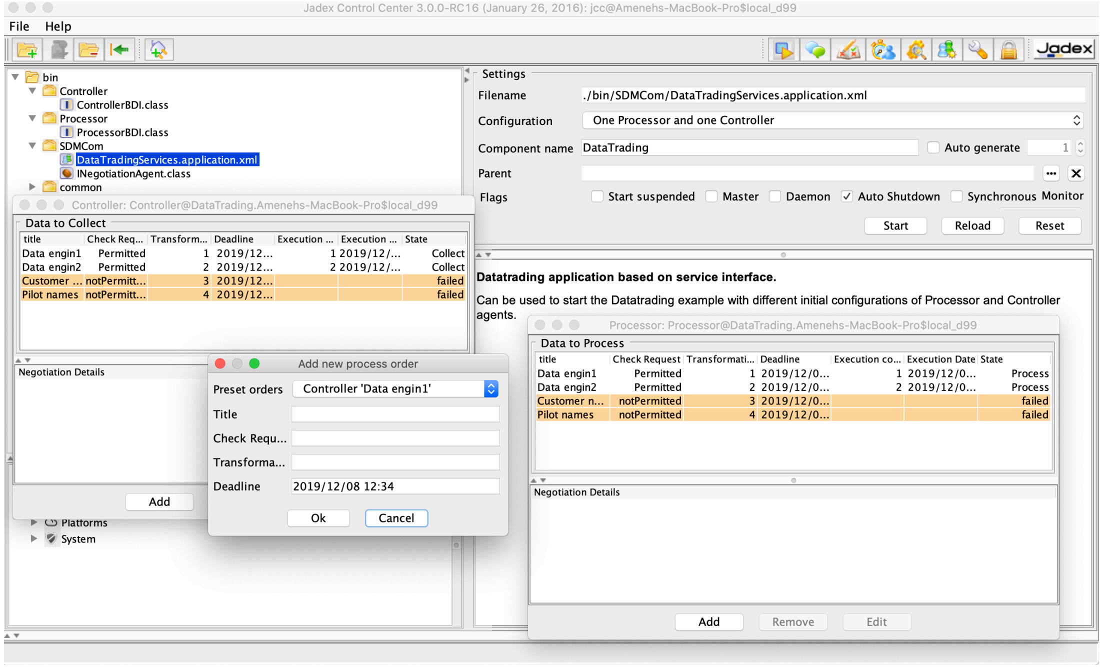
Flags ☐ Start suspended ☐ Master ☐ Daemon ☑ Auto Shutdown ☐ Synchronous Monitor OFF

University: University@UniversityServiceWiFi.Amenehs-MacBook-Pro$local_7a7

service is offered

name	Id	realId	Deadline	Execution Id	Execution Date	State
A	100120	10120	2019/12/0...			failed
B	4060	460	2019/12/0...			failed
H	159	159	2019/12/0...	159	2019/12/0...	done
I	3065	3065	2019/12/0...	3065	2019/12/0...	done

Figure B.2: Implementation of Eduroam WiFi connection with the Jadex Software, where the students need to provide an ID and university checks the ID to provide the service if the students are providing the valid ID.



Settings

Filename ./bin/SDMCom/DataTradingServices.application.xml

Configuration One Processor and one Controller

Component name DataTrading □ Auto generate 1

Parent

Flags □ Start suspended □ Master □ Daemon ☑ Auto Shutdown □ Synchronous Monitor

Datatrading application based on service interface.

Can be used to start the Datatrading example with different initial configurations of Processor and Controller agents.

Controller: Controller@DataTrading.Amenehs-MacBook-Pro$local_d99

Data to Collect

title	Check Req...	Transform...	Deadline	Execution ...	Execution ...	State
Data engin1	Permitted	1	2019/12...	1	2019/12/...	Collect
Data engin2	Permitted	2	2019/12...	2	2019/12/...	Collect
Customer ...	notPermitt...	3	2019/12...			failed
Pilot names	notPermitt...	4	2019/12...			failed

Negotiation Details

Add new process order

Preset orders Controller 'Data engin1'

Title

Check Requ...

Transforma...

Deadline 2019/12/08 12:34

Processor: Processor@DataTrading.Amenehs-MacBook-Pro$local_d99

Data to Process

title	Check Request	Transformati...	Deadline	Execution co...	Execution Date	State
Data engin1	Permitted	1	2019/12/0...	1	2019/12/0...	Process
Data engin2	Permitted	2	2019/12/0...	2	2019/12/0...	Process
Customer n...	notPermitted	3	2019/12/0...			failed
Pilot names	notPermitted	4	2019/12/0...			failed

Negotiation Details

Figure B.3: Implementation of STDMP the Jadex framework, where the ETI, LH and TRF are implemented. In this market place, we provided an ability to add a new member to the market. We check the purpose of contract and check the permission to use/ analyze the requested data.

BIBLIOGRAPHY

[1] G. Abdelkader. “Requirements for achieving software agents autonomy and defining their responsibility.” In: *Proc. Autonomy Workshop at AAMAS 2003*. Vol. 236. 2003.

[2] A. Abdul-Rahman and S. Hailes. “Supporting trust in virtual communities.” In: *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*. IEEE. 2000, 9–pp.

[3] M. Abomhara et al. “Cyber security and the internet of things: vulnerabilities, threats, intruders and attacks.” In: *Journal of Cyber Security and Mobility* 4.1 (2015), pp. 65–88.

[4] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles. “Towards a better understanding of context and context-awareness.” In: *International Symposium on Handheld and Ubiquitous Computing*. Springer. 1999, pp. 304–307.

[5] S Adali, W. Wallace, Y Qian, P Vijayakumar, and M. Singh. “A unified framework for trust in composite networks.” In: *Proc. 14th AAMAS W. Trust in Agent Societies, Taipei* (2011), pp. 1–12.

[6] N. Alechina, M. Dastani, and B. Logan. “Programming norm-aware agents.” In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems. 2012, pp. 1057–1064.

[7] A. Argandoña. “Sharing out in alliances: Trust and ethics.” In: *Journal of Business Ethics* 21.2-3 (1999), pp. 217–228.

[8] R. Bachmann. “Trust, power and control in trans-organizational relations.” In: *Organization studies* 22.2 (2001), pp. 337–365.

[9] K. S. Barber, K. Fullam, and J. Kim. “Challenges for trust, fraud and deception research in multi-agent systems.” In: *Workshop on Deception, Fraud and Trust in Agent Societies*. Springer. 2002, pp. 8–14.

[10] J. B. Barney and M. H. Hansen. “Trustworthiness as a source of competitive advantage.” In: *Strategic management journal* 15.S1 (1994), pp. 175–190.

[11] J. Barney, M. Wright, and D. J. Ketchen Jr. “The resource-based view of the firm: Ten years after 1991.” In: *Journal of management* 27.6 (2001), pp. 625–641.

[12] G. Beavers and H. Hexmoor. "Obligations in a BDI agent architecture." In: *International Conference on Artificial Intelligence (IC-AI 2002) Los Vegas, NV*. 2002.

[13] M. Berman, J. S. Chase, L. Landweber, A. Nakao, M. Ott, D. Raychaudhuri, R. Ricci, and I. Seskar. "GENI: A federated testbed for innovative network experiments." In: *Computer Networks* 61 (2014), pp. 5–23.

[14] C. Bicchieri. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2005.

[15] P. Blau. *Exchange and power in social life*. Routledge, 2017.

[16] G. Boella and L. van der Torre. "Norm governed multiagent systems: The delegation of control to autonomous agents." In: *Intelligent Agent Technology, 2003. IAT 2003. IEEE/WIC International Conference on*. IEEE. 2003, pp. 329–335.

[17] A. Boer. "Legal theory, sources of law and the semantic web." In: *Proceedings of the 2009 conference on Legal Theory, Sources of Law and the Semantic Web*. IOS Press. 2009, pp. 1–316.

[18] A. Boer and T. Van Engers. "An agent based legal knowledge acquisition methodology for agile public administration." In: *Proceedings of the 13th International Conference on Artificial Intelligence and Law*. ACM. 2011, pp. 171–180.

[19] A. Boer, T. Van Engers, and G. Sileno. "A problem solving model for regulatory policy making." In: *Modelling Policy-making (MPM 2011)* (2011), p. 5.

[20] L. Braubach, W. Lamersdorf, and A. Pokahr. "Jadex: Implementing a BDI-infrastructure for JADE agents." In: (2003).

[21] J. M. Broersen, M. Dastani, Z. Huang, J. Hulstijn, and L. Torre. "The BOID architecture; Conflicts Between Beliefs, Obligations, Intentions and Desires." In: (2001).

[22] A. Carter and L. Weber. "Trust and interpersonal relations: implications for nonviolence." In: *A Paper Presented at the Symposium on Nonviolence. Utica, NY*. 1992.

[23] C. Castelfranchi and R. Falcone. *Trust theory: A socio-cognitive and computational model*. Vol. 18. John Wiley & Sons, 2010.

[24] C. Castelfranchi, F. Dignum, C. M. Jonker, and J. Treur. "Deliberative normative agents: Principles and architecture." In: *International Workshop on Agent Theories, Architectures, and Languages*. Springer. 1999, pp. 364–378.

[25] Y. Chen, K. Hwang, and W.-S. Ku. "Collaborative detection of DDoS attacks over multiple network domains." In: *IEEE Transactions on Parallel and Distributed Systems* 18.12 (2007), pp. 1649–1662.

[26] J. A. Colquitt, B. A. Scott, and J. A. LePine. "Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance." In: *Journal of applied psychology* 92.4 (2007), p. 909.

[27] D. Cox, M. La Caze, and M. Levine. "Integrity." In: (2001).

[28] N. Criado, E. Argente, and V Botti. "A BDI architecture for normative decision making." In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1–Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems. 2010, pp. 1383–1384.

[29] J. Cruz and S Costa-Silva. "Trust: theoretical framework and underlying disciplines, conceptualization, antecedents and consequences." In: *EURAM–European Academy of Management, St Andrews, Scotland* (2004), pp. 5–8.

[30] T. Das and B.-S. Teng. "Risk types and inter-firm alliance structures." In: *Journal of management studies* 33.6 (1996), pp. 827–843.

[31] T. K. Das and B.-S. Teng. "Trust, control, and risk in strategic alliances: An integrated framework." In: *Organization studies* 22.2 (2001), pp. 251–283.

[32] A. Deljoo, L. Gommans, T. van Engers, and C. de Laat. "An Agent-Based Framework for Multi-Domain service networks: Eduroam case study." In: *The 8th International Conference on Agents and Artificial Intelligence (ICAART'16)*. 2016, pp. 275–280.

[33] A. Deljoo, T. M. van Engers, R. van Doesburg, L. Gommans, and C. de Laat. "A Normative Agent-based Model for Sharing Data in Secure Trustworthy Digital Market Places." In: 2018, pp. 290–296.

[34] A. Deljoo, T. van Engers, L. Gommans, and C. de Laat. "Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners." In: *2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)*. IEEE. 2018, pp. 45–54.

[35] A. Deljoo, T. van Engers, L. Gommans, and C. de Laat. "The Impact of Competence and Benevolence in a Computational Model of Trust." In: *IFIP International Conference on Trust Management*. Springer. 2018, pp. 45–57.

[36] A. Deljoo, T. van Engers, R. Koning, L. Gommans, and C. de Laat. "Towards Trustworthy Information Sharing by Creating Cyber Security Alliances." In: *IEEE TrustCom-18*. IEEE. 2018, pp. 1506–1510.

[37] A. Deljoo, L. Gommans, C. de Laat, and T. van Engers. “The Service Provider Group Framework.” In: *Looking Beyond the Internet: Workshop on Software-defined Infrastructure and Software-defined Exchanges,2016*. Flux Research Group, University of Utah, 2016.

[38] A. Deljoo, L. Gommans, C. de Laat, and T. van Engers. “What Is Going On: Utility-Based Plan Selection in BDI Agents.” In: *The AAAI-17 Workshop on Knowledge-Based Techniques for Problem Solving and Reasoning WS-17-12*. 2017.

[39] G. G. Dess, G. McNamara, A. B. Eisner, and S.-H. Lee. *Strategic Management: Creating Competitive Advantages*. McGraw-Hill Education, 2019.

[40] G. Dietz and D. N. Den Hartog. “Measuring trust inside organisations.” In: *Personnel review* 35.5 (2006), pp. 557–588.

[41] F. Dignum. “Autonomous agents with norms.” In: *Artificial Intelligence and Law* 7.1 (1999), pp. 69–79.

[42] V. Dignum and F. Dignum. “Towards an agent based infrastructure to support virtual organisations.” In: *Collaborative business ecosystems and virtual enterprises*. Springer, 2002, pp. 363–370.

[43] V. Dignum, J. Vázquez-Salceda, and F. Dignum. “Omni: Introducing social structure, norms and ontologies into agent organizations.” In: *International Workshop on Programming Multi-Agent Systems*. Springer. 2004, pp. 181–198.

[44] R. v. Doesburg and T. v. Engers. “Perspectives on the Formal Representation of the Interpretation of Norms.” In: *Legal Knowledge and Information Systems: JURIX 2016: The Twenty–Ninth Annual Conference*. Vol. 294. IOS Press. 2016, p. 183.

[45] R. van Doesburg and T. van Engers. “The False, the Former, and the Parish Priest.” In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*. 2019, pp. 194–198.

[46] K. M. Eisenhardt. “Agency theory: An assessment and review.” In: *Academy of management review* 14.1 (1989), pp. 57–74.

[47] A. Elangovan and D. L. Shapiro. “Betrayal of trust in organizations.” In: *Academy of management review* 23.3 (1998), pp. 547–566.

[48] N. E. Elgohary, A. A. Elfetouh, and S. I. Barakat. “Developing a reputation model for electronic markets.” In: *International Journal of Electrical and Computer Science (IJECS-IJENS)* 10.6 (2010).

[49] P. M. Fandt and G. R. Ferris. “The management of information and impressions: When employees behave opportunistically.” In: *Organizational Behavior and Human Decision Processes* 45.1 (1990), pp. 140–158.

[50] A. D. Farrell, M. J. Sergot, M. Sallé, and C. Bartolini. "Using the event calculus for tracking the normative state of contracts." In: *International Journal of Cooperative Information Systems* 14.02n03 (2005), pp. 99–129.

[51] R. Fehling. "A concept of hierarchical Petri nets with building blocks." In: *International Conference on Application and Theory of Petri Nets*. Springer. 1991, pp. 148–168.

[52] G Fortino, F Messina, D Rosaci, and G. Sarné. "Using trust and local reputation for group formation in the Cloud of Things." In: *Future Generation Computer Systems* 89 (2018), pp. 804–815.

[53] N. Ghosh, S. K. Ghosh, and S. K. Das. "SelCSP: A framework to facilitate selection of cloud service providers." In: *IEEE transactions on cloud computing* 3.1 (2015), pp. 66–79.

[54] L. Gommans, J. Vollbrecht, B. Gommans-de Bruijn, and C. de Laat. "The Service Provider Group framework: A framework for arranging trust and power to facilitate authorization of network services." In: *Future Generation Computer Systems* 45 (2015), pp. 176–192.

[55] L. H. M. Gommans. "Multi-domain authorization for e-Infrastructures." In: (2014).

[56] A. Gouaich. "Requirements for achieving software agents autonomy and defining their responsibility." In: *International Workshop on Computational Autonomy*. Springer. 2003, pp. 128–139.

[57] G. Governatori and A. Rotolo. "BIO logical agents: Norms, beliefs, intentions in defeasible logic." In: *Autonomous Agents and Multi-Agent Systems* 17.1 (2008), pp. 36–69.

[58] T. Grandison and M. Sloman. "A survey of trust in internet applications." In: *IEEE Communications Surveys & Tutorials* 3.4 (2000), pp. 2–16.

[59] M. Granovetter. "Economic action and social structure: The problem of embeddedness." In: *American journal of sociology* 91.3 (1985), pp. 481–510.

[60] G. Guo, J. Zhang, and D. Thalmann. "A simple but effective method to incorporate trusted neighbors in recommender systems." In: *International Conference on User Modeling, Adaptation, and Personalization*. Springer. 2012, pp. 114–125.

[61] G. Guo, J. Zhang, D. Thalmann, and N. Yorke-Smith. "Etaf: An extended trust antecedents framework for trust prediction." In: *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press. 2014, pp. 540–547.

[62] R. Hardin. "Distrust: Manifestations and management." In: *Distrust* 8 (2004), pp. 3–33.

[63] R. A. Heiner. "Origin of predictable behavior: Further modeling and applications." In: *The American economic review* 75.2 (1985), pp. 391–396.

[64] C. Hewitt, P. Bishop, and R. Steiger. "A universal modular actor formalism for artificial intelligence." In: *Proceedings of the 3rd international joint conference on Artificial intelligence*. Morgan Kaufmann Publishers Inc. 1973, pp. 235–245.

[65] W. N. Hohfeld. "Some fundamental legal conceptions as applied in judicial reasoning." In: *Yale Lj* 23 (1913), p. 16.

[66] J. F. Hübner, O. Boissier, R. Kitio, and A. Ricci. "Instrumenting multi-agent organisations with organisational artifacts and agents." In: *Autonomous agents and multi-agent systems* 20.3 (2010), pp. 369–400.

[67] J. F. Hübner, J. S. Sichman, and O. Boissier. "A model for the structural, functional, and deontic specification of organizations in multiagent systems." In: *Brazilian Symposium on Artificial Intelligence*. Springer. 2002, pp. 118–128.

[68] Jadex. *Jadex Active Components User Guide, Security*. URL: `https://download.actoron.com/docs/nightlies/latest/jadex-mkdocs/guides/ac/08%20Security/`.

[69] S.-L. Jan and G. Shieh. "Optimal sample sizes for Welch's test under various allocation and cost considerations." In: *Behavior research methods* 43.4 (2011), pp. 1014–1022.

[70] A. Josang and S. L. Presti. "Analysing the relationship between risk and trust." In: *International Conference on Trust Management*. Springer. 2004, pp. 135–145.

[71] L. Kagal, T. Finin, and A. Joshi. "Trust-based security in pervasive computing environments." In: *Computer* 34.12 (2001), pp. 154–157.

[72] H. H. Kelley. "Attribution theory in social psychology." In: *Nebraska symposium on motivation*. University of Nebraska Press. 1967.

[73] K. Kelton, K. R. Fleischmann, and W. A. Wallace. "Trust in digital information." In: *Journal of the American Society for Information Science and Technology* 59.3 (2008), pp. 363–374.

[74] T. Kiyonari, T. Yamagishi, K. S. Cook, and C. Cheshire. "Does trust beget trustworthiness? Trust and trustworthiness in two games and two cultures: A research note." In: *Social Psychology Quarterly* 69.3 (2006), pp. 270–283.

[75] M. J. Kollingbaum and T. J. Norman. "NoA-a normative agent architecture." In: *IJCAI*. 2003, pp. 1465–1466.

[76] M. J. Kollingbaum. "Norm-governed practical reasoning agents." PhD thesis. University of Aberdeen Aberdeen, 2005.

[77] M. J. Kollingbaum. "Automating network security." PhD thesis. University of Amsterdam, 2020.

[78] R Koning, B de Graaff, G Polevoy, R Meijer, C de Laat, and P Grosso. "Measuring the efficiency of sdn mitigations against attacks on computer infrastructures." In: *Future Generation Computer Systems* 91 (2019), pp. 144–156.

[79] R Koning, G Polevoy, L Meijer, C de Laat, and P Grosso. "Approaches for collaborative security defences in multi network environments." In: *The 6th IEEE International Conference on Cyber Security and Cloud Computing (IEEE CSCloud 2019)/(IEEE Edgecom 2019)*. 2019.

[80] R. Koning, B. De Graaff, R. Meijer, C. De Laat, and P. Grosso. "Measuring the effectiveness of SDN mitigations against cyber attacks." In: *IEEE Conference on Network Softwarization (NetSoft), 2017*. IEEE. 2017, pp. 1–6.

[81] R. Koning, A. Deljoo, S. Trajanovski, B. de Graaff, P. Grosso, L. Gommans, T. van Engers, F. Fransen, R. Meijer, R. Wilson, et al. "Enabling E-science applications with dynamic optical networks: Secure autonomous response networks." In: *Optical Fiber Communications Conference and Exhibition (OFC), 2017*. IEEE. 2017, pp. 1–3.

[82] T. R. Koscik and D. Tranel. "The human amygdala is necessary for developing and expressing normal interpersonal trust." In: *Neuropsychologia* 49.4 (2011), pp. 602–611.

[83] R. Krishnan, X. Martin, and N. G. Noorderhaven. "When does trust matter to alliance performance?" In: *Academy of Management journal* 49.5 (2006), pp. 894–917.

[84] N. Labraoui, M. Gueroui, and L. Sekhri. "A risk-aware reputation-based trust management in wireless sensor networks." In: *Wireless Personal Communications* 87.3 (2016), pp. 1037–1055.

[85] J.-N. Lee, M. Q. Huynh, and R. Hirschheim. "An integrative model of trust on IT outsourcing: Examining a bilateral perspective." In: *Information Systems Frontiers* 10.2 (2008), pp. 145–163.

[86] D. Z. Levin, R. Cross, L. C. Abrams, and E. L. Lesser. "Trust and knowledge sharing: A critical combination." In: *IBM Institute for Knowledge-Based Organizations* 19 (2002).

[87] J. D. Lewis and A. Weigert. “Trust as a social reality.” In: *Social forces* 63.4 (1985), pp. 967–985.

[88] H. Liu, E.-P. Lim, H. W. Lauw, M.-T. Le, A. Sun, J. Srivastava, and Y. Kim. “Predicting trusts among users of online communities: an epinions case study.” In: *Proceedings of the 9th ACM conference on Electronic commerce*. ACM. 2008, pp. 310–319.

[89] U. Lotzmann, M. Möhring, and K. G. Troitzsch. “Simulating the emergence of norms in different scenarios.” In: *Artificial intelligence and law* 21.1 (2013), pp. 109–138.

[90] M. Luck, S. Mahmoud, F. Meneguzzi, M. Kollingbaum, T. J. Norman, N. Criado, and M. S. Fagundes. “Normative Agents.” In: *Agreement Technologies*. Ed. by S. Ossowski. Springer Netherlands, 2013, pp. 209–220.

[91] N. Luhmann. *Trust: And, Power : Two Works*. John Wiley & Sons, 1979.

[92] S. P. Marsh. “Formalising trust as a computational concept.” In: (1994).

[93] R. C. Mayer, J. H. Davis, and F. D. Schoorman. “An integrative model of organizational trust.” In: *Academy of management review* 20.3 (1995), pp. 709–734.

[94] F. Meneguzzi and M. Luck. “Norm-based behaviour modification in BDI agents.” In: *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems. 2009, pp. 177–184.

[95] F. Meneguzzi, W. Vasconcelos, N. Oren, and M. Luck. “Nu-BDI: norm-aware BDI agents.” In: *The 10th European Workshop on Multi-agent Systems (EUMAS), London, UK*. 2012.

[96] L. Mui, M. Mohtashemi, and A. Halberstadt. “A computational model of trust and reputation.” In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. IEEE. 2002, pp. 2431–2439.

[97] D. North and D. Smallbone. “The innovativeness and growth of rural SMEs during the 1990s.” In: *Regional studies* 34.2 (2000), pp. 145–157.

[98] N. Oren, S. Panagiotidi, J. Vázquez-Salceda, S. Modgil, M. Luck, and S. Miles. “Towards a formalisation of electronic contracting environments.” In: *Coordination, organizations, institutions and norms in agent systems IV*. Springer, 2009, pp. 156–171.

[99] S. Panagiotidi, S. Alvarez-Napagao, and J. Vázquez-Salceda. "Towards the norm-aware agent: Bridging the gap between deontic specifications and practical mechanisms for norm monitoring and norm-aware planning." In: *International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*. Springer. 2013, pp. 346–363.

[100] S. Panagiotidi, J. Vázquez-Salceda, and F. Dignum. "Reasoning over norm compliance via planning." In: *International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*. Springer. 2012, pp. 35–52.

[101] A. Parkhe. "Strategic alliance structuring: A game theoretic and transaction cost examination of interfirm cooperation." In: *Academy of management journal* 36.4 (1993), pp. 794–829.

[102] A. Picot, R. Reichwald, and R. T. Wigand. *Information, organization and management*. Springer Publishing Company, Incorporated, 2008.

[103] I. Pinyol and J. Sabater-Mir. "Computational trust and reputation models for open multi-agent systems: a review." In: *Artificial Intelligence Review* 40.1 (2013), pp. 1–25.

[104] A. Pokahr, L. Braubach, and W. Lamersdorf. "Jadex: A BDI reasoning engine." In: *Multi-agent programming*. Springer, 2005, pp. 149–174.

[105] G. Polevoy. "Defence Efficiency." In: *CoRR* abs/1904.07141 (2019). arXiv: `1904.07141`. URL: `http://arxiv.org/abs/1904.07141`.

[106] A. Preece, K. Hui, and P. Gray. "Kraft: Supporting virtual organisations through knowledge fusion." In: *Artificial Intelligence for Electronic Commerce: Papers from the AAAI–99 Workshop*. 1999, pp. 33–38.

[107] W. Reisig. *A primer in Petri net design*. Springer Science & Business Media, 2012.

[108] P. S. Ring and A. H. Van de Ven. "Structuring cooperative relationships between organizations." In: *Strategic management journal* 13.7 (1992), pp. 483–498.

[109] J. B. Rotter. "A new scale for the measurement of interpersonal trust 1." In: *Journal of personality* 35.4 (1967), pp. 651–665.

[110] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer. "Not so different after all: A cross-discipline view of trust." In: *Academy of management review* 23.3 (1998), pp. 393–404.

[111] M. Rovatsos. "Interaction frames for artificial agents." In: (2001).

[112] M. Rovatsos, G. Weiss, and M. Wolf. "An approach to the analysis and design of multiagent systems based on interaction frames." In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. ACM. 2002, pp. 682–689.

[113] J. Sabater and C. Sierra. "REGRET: reputation in gregarious societies." In: *Proceedings of the fifth international conference on Autonomous agents*. ACM. 2001, pp. 194–195.

[114] F. Sadri, K. Stathis, and F. Toni. "Normative KGP agents." In: *Computational & Mathematical Organization Theory* 12.2-3 (2006), p. 101.

[115] M. Sako. *Price, quality and trust: Inter-firm relations in Britain and Japan*. 18. Cambridge University Press, 1992.

[116] M. Sánchez, G. López, O Cánovas, and A. F. Gómez-Skarmeta. "A proposal for extending the eduroam infrastructure with authorization mechanisms." In: *5th International Workshop on Security in Information Systems (submitted 2007)*. 2007.

[117] F. E. Satterthwaite. "Synthesis of variance." In: *Psychometrika* 6.5 (1941), pp. 309–316.

[118] J. R. Searle and J. R. Searle. *Speech acts: An essay in the philosophy of language*. Vol. 626. Cambridge university press, 1969.

[119] J. Searle. *Making the social world: The structure of human civilization*. Oxford University Press, 2010.

[120] G. Shani and A. Gunawardana. "Evaluating recommendation systems." In: *Recommender systems handbook*. Springer, 2011, pp. 257–297.

[121] G. Sileno, A. Boer, T. van Engers, et al. "Towards a computational model for institutional scenarios." In: *BNAIC 2013: Proc. of the 25th Benelux Conf. Artificial Intelligence*. Citeseer. 2013, pp. 183–190.

[122] F. Skopik, D. Schall, and S. Dustdar. "Modeling and mining of dynamic trust in complex service-oriented systems." In: *Socially Enhanced Services Computing*. Springer, 2011, pp. 29–75.

[123] A. Taal, J. Wang, C. de Laat, and Z. Zhao. "Profiling the scheduling decisions for handling critical paths in deadline-constrained cloud workflows." In: *Future Generation Computer Systems* 100 (2019), pp. 237–249.

[124] T. R. Tyler. "Trust within organisations." In: *Personnel review* 32.5 (2003), pp. 556–568.

[125] J. Urbano, A. P. Rocha, and E. Oliveira. "The impact of benevolence in computational trust." In: *Agreement Technologies*. Springer, 2013, pp. 210–224.

[126] B. L. Welch. "The generalization ofstudent's' problem when several different population variances are involved." In: *Biometrika* 34.1/2 (1947), pp. 28–35.

[127] G. White. "Strategic, tactical, & operational management security model." In: *Journal of Computer Information Systems* 49.3 (2009), pp. 71–75.

[128] K. Wierenga and L. Florio. "Eduroam: past, present and future." In: *Computational methods in science and technology* 11.2 (2005), pp. 169–173.

[129] O. E. Williamson. "The institutions of governance." In: *The American Economic Review* 88.2 (1998), pp. 75–79.

[130] P. Yan-bin, J. Gao, J.-Q. Ai, C.-H. Wang, and G. Hang. "An extended agent BDI model with norms, policies and contracts." In: *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*. 2008.

[131] A. Zaheer, B. McEvily, and V. Perrone. "Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance." In: *Organization science* 9.2 (1998), pp. 141–159.

[132] L. G. Zucker. "Production of trust: Institutional sources of economic structure, 1840-1920." In: *Research in organizational behavior* 8 (1986), pp. 53–111.

[133] cvedetails.com. URL: `https://www.cvedetails.com/index.php`.

[134] merriam-webster. URL: `https://www.merriam-webster.com/dictionary/trust`.

[135] trustlet.org. *Epinions dataset*. URL: `http://www.trustlet.org/epinions.html`.

[136] van Engers, Tom M, Meijer, Lydia, Nieuwenhuis, Kees and Gommans, Leon. *NWO/STW Workshop "ICT with Industry2016"Lorenz Centre Leiden*. Tech. rep. 2016.

PUBLICATIONS

- Journal
  - Deljoo, A., van Engers, T., Taal, A., Gommans, L., & de Laat, C., "Social Computational Trust Model (SCTM): A Framework to Facilitate Selection of Partners", Under review, New Generation Computing.
  - Deljoo, A., Koning, R., van Engers, T., Taal, A., Gommans, L., & de Laat, C., "Managing Effective Collaboration in Cybersecurity Alliances Using Social Computational Trust", Under Review, Annals of Telecommunications.
- Conference Proceedings
  - Deljoo, A., Koning, R., van Engers, T., Gommans, L., & de Laat, C., (2019, October). "Managing Effective Collaboration in Cybersecurity Alliances Using Social Computational Trust". In 2019 3rd Cyber Security in Networking Conference (CSNet) (pp. 50-57). © IEEE. DOI: 10.1109/CSNet47905.2019.9108949
  - Deljoo, A., van Engers, T., Koning, R., Gommans, L., & de Laat, C., (2018, August). "Towards trustworthy information sharing by creating cyber security alliances". In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 1506-1510).© IEEE. DOI: 10.1109/TrustCom/BigDataSE.2018.00213
  - Deljoo, A., van Engers, T., Gommans, L., & de Laat, C., (2018, November). "Social computational trust model (sctm): A framework to facilitate selection of partners". In 2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS) (pp. 45-54). © IEEE. DOI: 10.1109/INDIS.2018.00008
  - Deljoo, A., van Engers, T., Gommans, L., & de Laat, C., (2018, July). "The impact of competence and benevolence in a computational model of trust". In IFIP International Conference on Trust Management (pp. 45-57). Springer, Cham. © Springer. DOI:10.1007/978-3-319-95276-5_4
  - Deljoo, A., van Engers, T. M., van Doesburg, R., Gommans, L., & de Laat, C., (2018). "A Normative Agent-based Model for Sharing Data in Secure Trustworthy Digital Market Places". In ICAART (1) (pp. 290-296). © SCITEPRESS. DOI: 10.5220/0006661602900296

- Deljoo, A., Gommans, L., Van Engers, T., & de Laat, C. (2016, February). "An agent-based framework for multi-domain service networks". In Proceedings of the 8th International Conference on Agents and Artificial Intelligence (pp. 275-280). SCITEPRESS-Science and Technology Publications, Lda. © SCITEPRESS. DOI: 10.5220/0005821502750280
- Deljoo, A., van Engers, T., Gommans, L. & de Laat, C., (2017, March). "What is going on: Utility-based plan selection in bdi agents". In Workshops at the Thirty-First AAAI Conference on Artificial Intelligence.© AAAI
- Deljoo, A., Gommans, L., van Engers, T. & de Laat, C., (2016)."Regulating complex adaptive systems: Towards a computational model for simulating the effects of rules". In: Proceedings of International Conferen on Complex Systems 2016.
- Deljoo, A., Gommans, L., de Laat, C., & van Engers, T., (2016). "The service provider group framework". In Looking Beyond the Internet: Workshop on Software-defined Infrastructure and Software-defined Exchanges.
- Koning, R., **Deljoo, A.**, Trajanovski, S., de Graaff, B., Grosso, P., Gommans, L., van Engers, T., Fransen, F., Meijer, R., Wilson, R. and de Laat, C., (2017, March). "Enabling e-science applications with dynamic optical networks: Secure autonomous response networks". In Optical Fiber Communication Conference (pp. Tu3E-1). Optical Society of America. DOI:10.1364/OFC.2017.Tu3E.1.
- Koning, R., **Deljoo, A.**, Meijer, L., de Laat, C., & Grosso, P., (2019, October). Trust-based collaborative defences in multi network alliances. In 2019 3rd Cyber Security in Networking Conference (CSNet) (pp. 42-49). IEEE. © IEEE. DOI: 10.1109/CSNet47905.2019.9108945

CODE

The source code for the SARNET Alliance project is published at https://github.com/Adljoo. The source code is developed by Ameneh Deljoo. In the following, we provide the links to different repository.

- An agent-based model to simulate the collaborative network and evaluate the member's trustworthiness
  - https://github.com/Adeljoo/Collaborative-network

- An agent-based model to simulate the secure trustworthy digital data market places
  - https://github.com/Adeljoo/SDM-BDIAgent

- An agent-based model to simulate the stability of the alliances
  - https://github.com/Adeljoo/Alliance-Stability/blob/master/netlogo

- An agent-based model to simulate the Eduroam Case study
  - https://github.com/Adeljoo/Eduroam-credential

- The BDI agent's planner extension
  - https://github.com/Adeljoo/Planselector

- Iterated Prisoner Dilemma Game
  - https://github.com/Adeljoo/Prisoner-Dilemma

  - Normative BDI agent (N-BDI*)
    - https://github.com/Adeljoo/Normative-BDI
  - The thesis datasets
    - https://github.com/Adeljoo/Thesis-Deljoo

  It should be mentioned that the source codes can be changed or modified over time.

PROJECT ACKNOWLEDGMENTS

We would like to thank the following project partners and other organizations for making this thesis a reality:

- The Dutch Science Foundation NWO for funding the project (grant no: CYBSEC.14.003 / 618.001.016)
- COMMIT for funding part of the project (WP20.11).
- TNO for providing input and funding and project management.
- KLM for the interesting use-cases and input.

ACKNOWLEDGMENTS

If you can't fly then run, if you can't run then walk, if you can't walk then crawl, but whatever you do you have to keep moving forward

— Martin Luther King Jr.

I have received a great deal of support and assistance throughout the flying, running, walking, and crawling to have my name on this dissertation's front cover.

I express my gratitude towards my first promoter, Prof. Cees de Laat who has played a significant role in my academic development. Dear Cees, you have "critically" and encouragingly supported me during my PhD over the past four (+1) years, and I will not forget how you have helped me and guided me during my last year. Dear Cees, I am deeply indebted for all your help.

Dear Prof. Tom van Engers, considering being my second promoter, I am so grateful for your supervision, writing corrections, debates, jokes, and complaints. You played a significant role in my academic development. Tom, thank you for spending limitless hours brainstorming ideas with me, reading my unreadable manuscripts and clarifying my arguments throughout our sub-missions.

I wish to register my deepest gratitude for my daily supervisor Prof. Leon Gommans. Leon, you have been my mentor in my academic and personal life since I joined the SNE/ CCI Lab. During these four years, Leon supports and encourages me in one sort or another, which results from this Ph.D. thesis and shapes the person I am today. Leon, you help me to think, to be accurate, and defend as an independent researcher. Leon, thank you for your enthusiasm, positive attitude, guidance, your friendship through this journey.

I want to thank Dr. Arie Taal, life savor, who had provided invaluable input on my research when most needed. Thank you for all the insights on the notations and formula side of my research, and thank you for all the hours you have spent on providing many valuable suggestions that improved this manuscript.

Dear Committee members, thank you for taking the time to read my thesis and provide me your comments.

Ralph, Stojan, Ben and Gleb, I have learned so much from you guys. Ralph, you became my shoulder to cry and a buddy to gossip. Ralph, you helped and supported, especially in the first and fourth year of my PhD project a lot to get the ball of the PhD rolling. I am deeply grateful to my all officemate who have joined C3.203 gang, Ralph, Joe, Pieter, Lukaz, Souley, Tim, Ben, Milen, Xin Daniel, and Mary.

Many of my other colleagues have shared a lot with me, have helped me and inspired me in many ways that I can't possibly forget: Lukaz, Joe (Joseph), Reggie, Pieter, Spiros, Tim, Julius, Julio, Lu, Xin, Sara, Milen, Mostafa, Paola, Yuri, Adam, Zhiming and other SNE rocks stars. Dear Mary, I want to thank you separately for proofreading my manuscript, can not thank you enough. Rian, your one of the sweetest and caring person, I have ever seen; you care and help me in many ways, from translating the summary to what to wear in my defense. Thank you for your friendship and support.

I would like to thank the project partners Tako Huisman (KLM), Lydia Meijer (TNO), Frank Fransen (TNO), Rodney Wilson (Ciena), and Kees Nieuwenhuis (Thales) for providing invaluable outsiders perspectives on my research.

I also want to thank my colleague at KPN–DSH. Arend–Jan, thank you for being such a friendly boss, creating a friendly environment, and giving me the space to learn, experiment, and grow. Baha, Martin, Casper, Marco, Marko, Victoria, Twan bedankt voor alles.

Moving to another country makes it even harder, particularly because one has to build everything from scratch. I, however, have been fortunate for having friends who have supported me in many ways during this journey: Arash, Ali, Shima, Marzieh, Ali, Noopure, Abi, Sayeda, Zahid, Rian, Ahmadreza, Seyed Hossein, Fahimeh, Betty, Joke, Pieter. Thank you for supporting me in difficult and happy moments, listening to all of my frustrations, and inspiring me to be a better version of myself every day.

Arash, I am so grateful for your friendship. You are like a mentor and a brother to me and an amazing uncle to Sam. You always welcome us to your home, especially when I needed the most. You are a friend who can rely on for everything. I can not thank you enough and always jealous of Shima and Ali because they have you in their life.

Sami (khahar janam), I left you when you needed me the most, and it was the hardest thing I have ever done. I am so lucky to have you as my sister, role model, and an amazing aunt to Sam. No matter what life throws at you, you are always there for me. Sami janam, you devoted yourself to our parents all these years, and I can not imagine what life would be for them without you. You sacrifice for me and support me in many ways that I can not thank you enough. I love you my sisi. Mani jan, thank you for being such a cool brother–in–law and always keeps a smile on my sister's face. Thank you for taking care of my parents and being another son to them; it means a lot to me. Mohammad, Rahele and Rahil jan, I am trully thankful for your support during all these years that I have been away. Abolfazl and Fatemeh, thank you for your distant support over these years.

Maman and Baba, being your child and your daughter, is a blessing, you two are my life–coaches in many ways. I owe it to you; you made it possible for me to move out and follow my dreams. Baba, my hero, you inspire me with your hard work, passion for learning, and social skill (which I also inherited from you). Maman, Maman Goli, words can not express how much I love you; you taught me so many lessons since day one, but the best one is to be independent, self-determined, and mom. I walked on your path and never doubt that I will reach my destination if I follow the lights. Maman and Bana, I bend to my knee for you and thank you for all your scarifies that you make to give me/us the best of the best. Khili khili dosteton daram.

Hassan, Love of my life, this long journey would not have been possible without you. I know hon, it was not an easy path that we both took, but we have reached a significant milestone together. You stand by my side and hold me when I needed the most. Thank you for listening to all my frustrations and nagging during my Ph.D. time. Thank you for being such a great partner in crime in all my life and helping me to become a better version of myself. Thank you for knowing me, correcting my latex error, and even complaining to my supervisors when I was not in good shape. I love you so much and much, “yek mashine ghesti bardar namzadeto negahdar".

Sam my dearlings, I never could imagine what life would be without you. Sam, you bring so much joy and pride in my life, and you made me another person. You are the most fantastic teacher on the entire planet; for example, no one could force me to eat with my right hand, but you nailed it. I will never forget your smile when I come to daycare to pick you up or those days when you sat on my lap, giving me comfort and extra hands to type my manuscript. By all my heart, I know that you want to spend more time with me instead of Jullanda, but when you grow up, I want you to be proud as I am with my mom. I want you to know that I am always “standing on the side when you check it out", I’ll be right there no matter what happens. By this dissertation, I want to tell you that nothing can stop you from achieving your dreams. Thank you for being so patient with me and my Ph.D. journey. Mommy loves you to the moon and back.

SUMMARY

Simple attacks can be countered by simple technical measures, but a defence against organized attacks requires collaboration amongst service providers. In order to create effective defense strategies, sharing cyber intelligence amongst service providers is therefore becoming increasingly important. Additionally, networks have grown in scale, complexity, and degree of inter-connectedness, such that their protection can often only be guaranteed and financed as a shared effort. To create a collaborative network amongst different service providers to facilitate the sharing of information and cyber threat intelligence we need to organize, maintain and evaluate trust amongst the autonomous members who have their own desire and goals to achieve that may result in conflicting interests.

The series of studies reported upon in this thesis were conducted in the context of the SARNET project (Security Autonomous Response networks). The SARNET project aims to provide an ICT system that can defend itself autonomously. To design such a system, the SARNET project looks at three different layers, the Strategic, Tactical and Operational layer [81]. This thesis considers questions at the Strategic layer, where we studied what is needed to create and maintain a cyber security alliance. At this level, this research focuses on the question "What dynamic computational trust models enable cyber-intelligence sharing through partner selection for collaborative cyber defense operations?". In this research, we operationalized the trust concept and researched the impact of trustworthiness factors on the success of a collaboration. In order to answer the research questions, first, we identified the requirements to create alliances. These requirements are:

1. The creation and maintenance of trust amongst alliance members must be well organized.
2. The definition of common policies and standards through a democratic process that supports the federated way of working in such alliances.
3. The alliance must provide incentives that are based on common benefits that no single member can achieve on its own.

To fulfill above requirements, we used the Service Provider Group (SPG) [37] as a governance framework to define common policies and ways to organize its implementation.

To understand collaborative networks, we studied a number of cases including the Eduroam case, an example of a collaborative network

where members are bound by collaborative rules that describe collaborative actions [32]. We investigated the feasibility of an agent based modeling approach(ABM) to model such collaborative networks. The Belief-Desire and Intention (BDI) agent is used in this thesis. We implemented an ABM simulation of Eduroam [32] as our first collaborative network simulation and we extended the original BDI agent model. The extended BDI model called N-BDI* is able to reason about the norms and select a plan with the highest utility. We tested the N-BDI* framework with different scenarios such as secure trustworthy digital market place (STDMP), Eduroam and the Cyber threat Alliance [33].

The research resulted in a computational trust model. This model motivates the selection of the three main trustworthiness factors, benevolence, competence and integrity. We showed that each of these factors plays an important role in evaluating the trust of the trustee [35, 36]. In our computational trust model (SCTM), we employ both direct and indirect evidence on the trustee. Each collaboration comes with certain risks that need to be managed and minimized. Therefore, we identify two types of risk, (relational and performance risk), in the alliances and estimated them through the SCTM framework. We showed that our SCTM helps the members of the alliance to select the right partner to collaborate within the situation at hand, while keeping the interaction risk to a minimum. In the SARNET project, we recognized different situations requiring members to trust other parties. Based on these situations, we have developed different scenarios with different weights for the trustworthiness factors. We evaluated the SCTM model through a series of simulations.

SAMENVATTING

Eenvoudige aanvallen kunnen worden tegengegaan met eenvoudige maatregelen, maar een verdediging tegen georganiseerde aanvallen vereist samenwerking tussen dienstverlenende partijen. Om effectieve verdedigingsstrategieën te kunnen ontwikkelen, is het van een steeds groter belang dat cyber-intelligentie gedeeld wordt tussen deze dienstverleners. Tevens zijn netwerken groter en complexer geworden, en is er sprake van grotere onderlinge verbondenheid, zodat bescherming en beveiliging daarvan enkel gegarandeerd en betaald kan worden door samen te werken.

Om een gezamenlijk netwerk op te zetten tussen verschillende dienstverleners, zodat data over cyber-dreigingen gedeeld kan worden, moeten we vertrouwen organiseren, onderhouden en evalueren tussen de autonome leden van het samenwerkingsverband. Deze leden hebben hun eigen wensen en doelen, die voor conflicterende belangen kunnen zorgen.

De serie onderzoeken, gerapporteerd in deze thesis, zijn gedaan in de context van het SARNET (Security Autonomous Response networks)-project. Het SARNET-project heeft als doel om een ICT-systeem te bieden dat zichzelf autonoom kan verdedigen. Om een dergelijk systeem te ontwerpen, kijkt het SARNET-project op drie niveaus: Strategisch, Tactisch, en Operationeel [81].

Deze thesis gaat over het Strategisch niveau, waar we onderzochten wat er nodig is om een cyber-veiligheid samenwerkingsverband op te zetten. Op het genoemde niveau ligt de focus op de vraag: "Welke dynamische computationele modellen maken het delen van cyber-intelligentie via keuze van partners mogelijk, om verdedigingsoperaties in samenwerking uit te voeren?"

We hebben in dit werk het concept vertrouwen geoperationaliseerd en we hebben de effecten van betrouwbaarheid op het succes van het samenwerkingsverband onderzocht. Om de onderzoeksvragen te kunnen beantwoorden, hebben we eerst de voorwaarden geïdentificeerd, die nodig zijn om een samenwerking op te zetten. Deze voorwaarden zijn:

1. Het creëren en onderhouden van vertrouwen tussen de leden van het samenwerkingsverband moet goed georganiseerd zijn.

2. De definitie van beleid, procedures, en standaarden moet tot stand komen via een democratisch proces, dat de federatieve manier van werken ondersteunt.

3. De alliantie moet stimulans aan leden bieden, gebaseerd op gedeelde voordelen, die de leden alleen niet kunnen bereiken.

Om aan de bovengenoemde voorwaarden te kunnen voldoen, hebben we de Service Provider Group (SPG) [37] gebruikt als raamwerk om de gedeelde procedures, het gezamenlijke beleid en de overeengekomen standaarden te kunnen definiëren, en om dezen te kunnen implementeren.

Om gezamenlijke netwerken te begrijpen, hebben we onderzoek gedaan naar een aantal bestaande netwerken, waaronder het voorbeeld van Eduroam, waar leden aan onderlinge regels gebonden zijn, die de gezamenlijke acties beschrijven [32]. We bekeken de haalbaarheid van een op agenten gebaseerde aanpak om te modelleren.

We implementeerden een ABM simulatie van Eduroam [32] als eerste simulatie van een gezamenlijk netwerk en breidden het originele BDI agenten model uit. Het uitgebreide BDI model, N-BDI genoemd*, is in staat om de normen te beredeneren en het meest effectieve plan te kiezen. We hebben het N-BDI* raamwerk getest in verschillende scenario's, zoals secure trustworthy digital marketplace (STDMP), Eduroam en de Cyber threat Alliance [33].

Het onderzoek resulteerde in een computationeel model van vertrouwen (computational trust model). Dit model onderbouwt de keuze voor de drie belangrijkste factoren van vertrouwen: welwillendheid, competentie, en integriteit. We hebben laten zien dat elk van deze factoren een belangrijke rol speelt bij de evaluatie van vertrouwen bij de vertrouwde partij [35, 36].

In ons computationele model van vertrouwen (SCTM) gebruiken we zowel direct als indirect geleverd bewijs over de vertrouwde partij. Elk samenwerkingsverband heeft zekere risico's, die beheerst en geminimaliseerd moeten worden. Om die reden identificeren we twee soorten risico (relationeel en op prestatie) in de alliantie, en schatten we de risico's met het SCTM raamwerk.

We hebben gedemonstreerd dat het SCTM raamwerk de leden helpt om de juiste samenwerkingspartner te kiezen voor de situatie waar het genoemde lid zich op dat moment in bevindt en tegelijkertijd het interactie-risico tot een minimum te beperken. In het SARNET-project hebben we verschillende situaties herkend die vertrouwen vereisen richting andere partijen. Gebaseerd op deze situaties hebben we verschillende scenario's ontwikkeld, die gebruik maken van weging van de factoren van betrouwbaarheid. We hebben het SCTM model geëvalueerd in series van simulaties.

ABOUT THE AUTHOR

Ameneh Deljoo was born in Shiraz, Iran. She obtained a BSC, software engineering from Zanjan University in 2010, and MSC in Information Technology from Shiraz University in 2012.

In 2012, Ameneh moved to the Netherlands to perform research on collaborative networks in the Technology, Policy and Management (TBM) faculty at the Delft University of Technology. She worked on several EU projects during her time in TUDelft.

In June 2015, Ameneh started as a PhD candidate in the Systems and Networking Laboratory (SNE) Informatics Institute, Faculty of Science, University of Amsterdam. She worked for the SARNET project funded by the Dutch Science Foundation project SARNET (grant no: CYB-SEC.14.003/618.001.016) and by the Dutch project COMMIT (WP20.11).

During her PhD project, she co-authored a number of papers (see the publication list for the complete list). She supervised several master students working on topics related to her PhD project. Ameneh was also involved with teaching master's courses such as Policy Making and Rule Governance, complex network, Introduction into AI (seminar) at the University of Amsterdam.

Currently, Ameneh is a solution architecture at Data services Hub, KPN. Overall, Ameneh identifies herself as a multi-disciplinary researcher who is genuinely interested in exploring the information-sharing infrastructure.