YOU WIN SOME YOU RANSOM

Reconstructing the ransomware ecosystem using ground truth communication data of the Conti ransomware gang

V.F.F. Stolk

Delft University of Technology

TU Delft

YOU WIN SOME YOU RANSOM

Reconstructing the ransomware ecosystem using ground truth communication data of the Conti ransomware gang

by

V.F.F. Stolk

A thesis submitted to the Delft University of Technology in partial fulfillment
of the requirements for the degree of
Master of Science
in **Complex, Systems, Engineering and Management**
to be defended publicly on Thursday August 25, 2022 at 2:00 PM.

Student number:	4365711	
Thesis committee:	Chairperson	: Prof.dr. M.E. Warnier
	First Supervisor	: Dr. R.S. van Wegberg
	Second Supervisor	: Prof.dr. M.E. Warnier

TU Delft

ABSTRACT

Ransomware has evolved over the years, shifting from widespread attacks targeting individuals to focused attacks on businesses and agencies. These attacks are performed by ransomware gangs while establishing interaction within the ransomware ecosystem. In this thesis, the ransomware ecosystem is posited as being constructed of three separate sub-ecosystems: the attacker sub-system, the defender sub-system, and the governance sub-system. Since ransomware gangs put in an effort to hide their internal communication and operation from the outer world, difficulties arise in correctly understanding the ransomware ecosystem and a ransomware gang's establishment of interactions within this ecosystem. As a result, current interventions are ineffective.

While earlier research has been conducted on ransomware, we observe two knowledge gaps: 1) there is a lack of understanding of how ransomware gangs establish interactions with actors in the ransomware ecosystem, and 2) There has been a lack of research that uses ground truth data due to ransomware gangs keeping their internal communication and operations hidden. This thesis uses the leaked internal communication data of the Conti ransomware gang to fill these knowledge gaps and answer the research question: *"To which extent can the ransomware ecosystem be reconstructed using ground truth communication data of the Conti ransomware gang?"*.

To answer this question, a novel methodology is proposed that uses Latent Dirichlet Allocation (LDA) topic modeling to empirically determine overarching topics in Conti's internal communication. It is then researched how these overarching topics map to Conti's tactics, techniques, and procedures (TTP) which is a commonly used methodology to better understand how ransomware gangs operate. Subsequently, these TTP are leveraged to reconstruct the ransomware ecosystem while taking the perspective of how the Conti ransomware gang establishes interactions within the ransomware ecosystem.

The findings of this thesis indicate that Conti is a large and professional organization that incorporates and adjusts services of service-providing cybercriminals in the attacker ecosystem rather than developing their ransomware themselves using scarce IT talent. In addition, reconnaissance is one of the most critical activities that ransomware gangs perform to get to a successful ransomware attack. While researching Conti's TTP, this thesis identifies novel TTP of ransomware gangs, such as Conti's attack chain, reconnaissance procedure, and money laundering procedure.

We conclude that the ransomware ecosystem can be reconstructed from the attacker ecosystem, the defender ecosystem, and the governance ecosystem, in which ransomware gangs establish interactions within each sub-ecosystem while operating from the attacker ecosystem. In the attacker ecosystem, ransomware gangs establish interactions with service-providing cybercriminals to outsource sub-commodities of their ransomware value chain. This allows them to strengthen their attack vectors by relying on the expertise of others and have a more varied set of attacks. The defender ecosystem is comprised of defenders that defend themselves against ransomware. Ransomware gangs establish interactions by performing extensive reconnaissance on defender territories and valuable information and open-source tools that strengthen their attack vectors. The governance ecosystem comprises governance actors that create and maintain the governance framework that influences the attacker ecosystem and defender ecosystem. Ransomware gangs establish interactions with actors in the governance ecosystem to observe the regulatory frameworks in place and adjust their TTP based on the involved risks of getting caught.

ACKNOWLEDGEMENTS

Dear reader,

This thesis concludes my enrollment in the master of Complex Systems, Engineering, and Management (CoSEM) at the Delft University of Technology. During the last two years of being a CoSEM student, there have been many professors, lecturers, students, entrepreneurs, and friends that triggered my eagerness to learn and showed me new perspectives within the wonderful world of complex systems, ransomware, and technology. I am truly grateful for these inspirations. Since thanking each individual could be a thesis in itself, I would like to use this section to express my gratitude to a few who have been part of my thesis.

Firstly, I would like to express my gratitude to my first supervisor, Dr. Rolf van Wegberg. This work quite literally originated from one of our first whiteboard drawing/discussion sessions in which you came up with the idea of seeing the ransomware ecosystem as being constructed of three separate sub-systems. I greatly appreciate your willingness to discuss my thesis project frequently, and without these discussions, I would not have been able to produce this thesis in its current form.

Secondly, I thank my chair and second supervisor, Prof.Dr. Martijn Warnier. It is almost a year ago since we had our first meeting to discuss possible thesis subjects. Instead of picking an "off-the-shelf" project, you triggered me to come up with my own ideas, and from the 12 initial ideas, you helped me structure the process leading to this final result. Furthermore, I would like to thank you for constructive feedback and your willingness to have short discussions to increase the quality of the work in this thesis.

Third, I would like to express my gratitude to Fieke Miedema, Tristan de Wildt, Eric Galinkin, and the attendees of the FIOD expert session for helping by discussing my progress and results during the writing of this thesis. Lastly, I would like to thank my family and friends for their support in writing this thesis, especially Tom Bastiaans, since our Pomodoro sessions were an important factor in this resulting thesis.

To introduce the rest of this thesis, I would like to present a quote from Sun Tzu, a Chinese military general that lived in 544-496 BC. I came across this quote in a paper while reviewing ransomware ecosystem literature. This quote reflects the importance of researching the ransomware ecosystem since we should not only research our enemies but take a step back and research ransomware within its full context. I feel like this quote reflects my point of view, and this point of view has inspired me to write this thesis and made me study ransomware with full commitment during the past half of a year.

> *"If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle."*
>
> — Sun Tzu, *The Art of War*

V.F.F. Stolk
The Hague, August 2022

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 BACKGROUND

1.1.1 Ransomware Ecosystem

Ransomware has evolved over the years, shifting from widespread attacks targeting individuals to focused attacks on businesses and agencies (Beaman et al., 2021). These attacks typically target machines containing highly sensitive files such as important financial data, business records, databases, hospital patient records, and government documents, effectively bringing business to a standstill (Li and Liao, 2022; Popli and Girdhar, 2019). Furthermore, many ransomware gangs have been incorporating additional pressure mechanisms by threatening to publish confidential data, which illustrates how these ransomware gangs evolved (Beaman et al., 2021). The cost of recovering from ransomware attacks can be enormous due to inoperativeness and cascading effects in a victim's supply chain, while the incentives for cybercriminals are huge (Fang et al., 2020; Galinkin, 2021; Pal et al., 2021). Therefore, the European Union Agency for Cybersecurity (ENISA) assessed ransomware as the prime threat in the cybersecurity threat landscape for 2021 (ENISA, 2021). However, to effectively intervene with ransomware, ransomware gangs and their interactions within their playing field should be correctly understood.

A problem that arises when finding effective interventions is that the playing field of ransomware gangs and its context, which we define in this thesis as the ransomware ecosystem, is more complex than most studies currently assume (Galinkin, 2021; Laszka et al., 2017; Raheem et al., 2021). Most studies currently define the ransomware ecosystem as a collection of ransomware gangs or ransomware gangs and their victims in which the victims and ransomware gangs mutually interact. However, other scholars have been addressing the existence of more complex relations and configurations of the ransomware ecosystem, e.g., see Kenneally (2021); Meland et al. (2020); Pal et al. (2021); Van Wegberg et al. (2017). For instance, ransomware gangs that are not "tech-savvy" can still extort businesses through services that allow them to outsource parts of their ransomware value chain. Furthermore, Pal et al. (2021) address that ransomware victims could harm other businesses in their supply chains through cascading effects, for example, because of their inoperability or found exploits. This illustrates how interactions within the ransomware ecosystem go beyond the mutual interaction between ransomware gangs and their victims as often assumed.

Other actors that are believed to play a role in the ransomware ecosystem are cyber-insurers (Fang et al., 2020; McDonald et al., 2022; Pal et al., 2021). Their exact role is part of an ongoing discussion among scholars since it is twofold. On the one hand, cyber-insurers can create financial incentives for adopting better cyber hygiene among victims (Kenneally, 2021). On the other hand, ransomware coverage in insurance policies is encouraging ransomware gangs and may be fueling the entire ransomware economy (MacColl et al., 2021). Therefore, cyber-insurers may have an important role in the ransomware ecosystem, although the exact role is yet to be determined (Laszka et al., 2017). Regulating actors such as national governments confine the rights and obligations of cyber-insurers and victims through their framework of rules and regulations. Furthermore, this framework of rules and regulations influences the tactics, techniques, and procedures that ransomware gangs use to prevent being shut down by law enforcement agencies. However, due to the complex nature of the ransomware ecosystem,

it is difficult to determine how these interventions could effectively influence actors to show the desired behavior. To further investigate the ransomware ecosystem, we first turn to how we can observe ransomware gangs and other cybercriminals in the ransomware ecosystem.

1.1.2 Cybercriminals in the ransomware ecosystem

With the evolution of ransomware, ransomware gangs went from performing attacks solely by themselves to attacking from an interconnected ecosystem of cybercriminals. In this ecosystem, ransomware gangs exchange resources with other cybercriminals, making it easier to perform ransomware attacks. While the ransomware value chain activities were first completely performed by a ransomware gang, different cybercriminals now provide services that allow the parts of the value chain to be outsourced (Meland et al., 2020; Van Wegberg et al., 2017). The introduction of these services allows ransomware gangs to mitigate some of their risks. Moreover, it allows ransomware gangs with lesser technical knowledge to perform successful ransomware attacks while the cybercriminals providing these services create additional revenue streams. Besides cybercriminals that provide services to outsource parts of the ransomware value chain, other cybercriminals are present that support ransomware attacks with illicit services of which bulletproof hosting providers are an example. Bulletproof hosting providers are hosting providers that allow illegal activities on their servers and are resilient to complaints of illicit activities (Goncharov, 2015). The existence of different relations among cybercriminals in this ecosystem has been widely discussed in scientific literature (Bayoumy et al., 2018; Cartwright et al., 2019b; McDonald et al., 2022; Meland et al., 2020). To further illustrate these relations, we will use the ransomware value chain and show how the different parts can be outsourced.

The ransomware value chain consists of four phases: development, distribution, account takeover, and cash-out, as shown in figure 1.1 (Van Wegberg et al., 2017). Each of these different phases in the value chain can be outsourced. Ransomware-as-a-Service (RaaS) is a widely discussed business model in which ransomware gangs create additional revenue streams by letting other ransomware gangs use their ransomware using an affiliate model. This allows ransomware gangs without technical knowledge to use others' ransomware in exchange for a percentage of the profits (Meland et al., 2020). Other cybercriminals in the ransomware ecosystem provide services to distribute the ransomware, for example, through sending emails with malicious attachments (Phishing-as-a-Service) or by selling found exploits (Exploit-as-a-Service) (Bayoumy et al., 2018; Keshavarzi and Ghaffary, 2020; Lee et al., 2019).

Figure 1.1: Value chain of a ransomware attack

During the account takeover, ransomware gangs infiltrate the victim's network, encrypt important files and look for additional exploits to see if they can access other victims' networks. When files are encrypted, and the victim has been presented a ransom note, ransomware gangs have "customer service" to help them with the ransom payment through cryptocurrencies, decryption of their files, and give them security advice to prevent future attacks (Cartwright et al., 2019b; Keshavarzi and Ghaffary, 2020; Van Wegberg et al., 2017). Again, this part of the value chain can be outsourced through other cybercriminals in the ransomware ecosystem. Finally, in order to cash-out their earnings, it needs to go through multiple stages of money laundering. The risk of getting exposed is higher during the cash-out phase of the ransomware value chain because transactions of cryptocurrencies provide a level of pseudo-anonymity and are not fully anonymous. Therefore, the cash-out phase is often outsourced to cybercriminals providing money laundry services (Huang et al., 2018). Cryptocurrencies, for example, are often sent through a 'mixer', making the cryptocurrency very difficult to trace (Huang et al., 2018; Raheem et al., 2021). Money laundry services are another example of outsourcing parts of the value chain,

illustrating how ransomware gangs interact with other cybercriminals in the attacker ecosystem. In this thesis, the attacker ecosystem is viewed as a sub-system of the ransomware ecosystem consisting of ransomware gangs and other cybercriminals that interact and share resources to help perform successful ransomware attacks.

1.1.3 Victims in the ransomware ecosystem

Potential victims of ransomware defend themselves against attacks and are similar to the ransomware gangs interconnected. For the sake of simplicity, we call these defenders in the rest of this thesis. These defenders often damage other companies in their supply chains when attacked by ransomware because the attack has cascading effects on other defenders in their supply chain (Pal et al., 2021). In addition, ransomware gangs can find confidential files of companies in the defender's supply chain or find additional exploits which encourage to attack of other related defenders. An illustrative example is a ransomware attack on a Managed Service Provider (MSP). These MSPs offer IT services to their clients, which they often do remotely. This remote connection makes these MSPs an important target for ransomware attacks because through this remote connection, other companies can be attacked, increasing the potential ransom to be gained (Kshetri and Voas, 2022). On the other hand, defenders can share information among other defenders on cybersecurity best practices to help other defenders increase their defense. However, multiple authors have claimed that defenders may show strategic behavior since the spending on one's defense may positively affect another's defense (Cartwright et al., 2019b; Chen et al., 2021; Laszka et al., 2017). The discussed examples illustrate how the defenders may influence other defenders in the potential of being successfully attacked. In this thesis, the defender ecosystem is defined as a sub-system of the ransomware ecosystem in which the defenders influence each other directly or indirectly on the potential of being successfully attacked.

1.1.4 Governance actors in the ransomware ecosystem

The regulatory and governance framework for ransomware is rather loose and non-existent (Kenneally, 2021). There is no framework for ransom payment transparency, and although governments are promoting not to pay ransoms, some governments sponsor ransomware attacks for political means (Lee et al., 2019; McDonald et al., 2022). Furthermore, in the US, ransom payments can be written off as ordinary, necessary, and reasonable tax expenses (Galinkin, 2021). However, governance can influence both the attacker ecosystem and the defender ecosystem to steer the problem of ransomware. National governments and international governmental organizations, for example, form the framework in which law enforcement agencies can work to try to seize and shut down different ransomware actors (Meland et al., 2020). Furthermore, national governments can stimulate defenders to adopt better cyber hygiene or discourage criminals by setting higher punishments (Galinkin, 2021). However, many of the involved actors and the interactions among actors in the governance ecosystem are still unclear. In this thesis, the governance ecosystem is defined as a sub-system of the ransomware ecosystem in which a collaboration of actors develops and maintains the governance framework that influences both the defender- and attacker ecosystem.

1.1.5 An integrated view of the ransomware ecosystem

Some actors that comprise the ransomware ecosystem do not fit in the sub-systems as discussed in sections 1.1.2, 1.1.3, and 1.1.4. Cyber-insurers, for example, are not easily categorized into one of the three sub-systems. However, many scholars have been discussing their importance in the ransomware ecosystem (Cartwright et al., 2019a; Kenneally, 2021; Laszka et al., 2017; McDonald et al., 2022). Similarly, cybersecurity advisors such as Fox-IT and Northwave that advise companies in cybersecurity and mediate in case of ransomware attacks cannot easily be placed in one of the sub-systems yet have an important role in the success of ransomware attacks.

These actors and each of the subsystems are influencing each other, eventually affecting the success of ransomware attacks. As discussed in section 1.1.4, a national government can influence ransomware gangs operating from that country, for example, by setting higher sanctions or stimulating ransomware gangs through state-sponsored attacks. In addition, governments can create financial incentives for companies to adopt better cyber hygiene or incentivize choosing to pay the ransom as in the case of tax redactable expenses in the US. Furthermore, there are examples of dissatisfied employees attacking their own company with ransomware using RaaS or helping ransomware groups to infiltrate the IT systems (McDonald et al., 2022). These are just a few examples illustrating how these sub-systems influence each other when we observe the collection of these sub-systems and the additional actors from an integrated view of the ransomware ecosystem. Although some relations between sub-systems and between actors within sub-systems have been studied, many roles of actors and their interactions with relation to ransomware are currently unclear.

Ransomware gangs have a central role in the ransomware ecosystem. In this thesis, it is posited that ransomware gangs intentionally make decisions on establishing interactions with actors in the ransomware ecosystem. These decisions to establish interactions relate to a ransomware gang's strategic business decisions, techniques, and procedures. Tactics, techniques, and procedures is a concept that is commonly used to describe the behavior or modes of operations of threat actors using these three categories (Egloff and Smeets, 2021; Yeboah-Ofori and Islam, 2019). Since the most dominant ransomware gangs are large well-structured organizations and their success in ransomware attacks relies on well-executed communication, it is logical to assume that they internally discuss topics entailing these TTP. It is therefore worthwhile to empirical research the internal discussions on strategic decisions related to establishing interactions since this would form a basis for a better understanding of the ransomware ecosystem.

However, ransomware gangs rely on multiple anonymous communication protocols to keep their internal communication and operation hidden from the outer world (Bayoumy et al., 2018; Keshavarzi and Ghaffary, 2020). Therefore, it has been difficult to observe the internal TTP of ransomware gangs empirically. This complicates finding effective interventions since the ransomware ecosystem is not yet correctly understood, starting from how ransomware gangs establish these interactions within the ransomware ecosystem. Taking this unclarity as its main focus, this thesis aims to provide insights into how ransomware gangs establish interactions with actors in the ransomware ecosystem and reconstruct it accordingly. To this end, internal chat log data of the Conti ransomware gang is used. Conti has been the most dominant ransomware gang in the ransomware ecosystem since the shutdown of REvil (ENISA, 2021). The Conti chatlog data used in this thesis originates from an anonymous researcher leaking Conti's internal messaging and files during the 2022 Russian invasion of Ukraine. The Northwave security group has translated these chatlogs, and the translated chatlog data is used in this thesis (Northwave Security, 2022).

This thesis is the first academic effort that empirically explores a ransomware gang's internal establishment of interactions within the ransomware ecosystem. Furthermore, it is the first to empirically study the internal tactics, techniques, and procedures of a ransomware gang based on ground truth data. In the work presented, topics in Conti's internal conversation data are empirically determined. Using these topics, the tactics, techniques, and procedures of the Conti ransomware gang are mapped. It is then researched how Conti internally establishes interactions with actors in the ransomware ecosystem based on their TTP.

1.2 CURRENT GAPS IN THE RANSOMWARE ECOSYSTEM LITERATURE

Research on ransomware has mainly focused on modus operandi and is mostly taken from a technical or descriptive perspective (Chen et al., 2021). Mei et al. (2021) studied the modus

operandi of different ransomware families using a System Provenance Graph, whereas Keshavarzi and Ghaffary (2020) studied the modus operandi by creating a separated attack chain for ransomware attacks. Focusing on the cash-out phase of the ransomware value chain, Raheem et al. (2021) studied the modus operandi of different ransomware families in relation to Bitcoin ransom payments. Various other studies have researched the modus operandi of ransomware actors by looking into specific parts of the ransomware value chain (Huang et al., 2018; Lee et al., 2019; McDonald et al., 2022). These types of studies, therefore, only investigate parts of the ransomware value chain and often observe ransomware in isolation, leaving out its context. Furthermore, these studies observe ransomware from an external perspective. Hence, we lack the understanding of how ransomware gangs strategically operate and internally establish interactions with their environment.

To create a more comprehensive understanding of how ransomware gangs internally operate and how these intentionally establish interactions in the ransomware ecosystem, we need to research ransomware from an internal perspective. In other words, we need to shift focus from observing ransomware gangs from an isolated external perspective to observing how ransomware groups interact with their complex environment from an internal perspective. This is where we use TTP to describe how ransomware groups internally operate and use these to reconstruct the ransomware ecosystem. This eventually would form the basis for a better understanding of how ransomware gangs establish interactions within the ransomware ecosystem. This understanding helps policymakers more effectively design interventions targeting critical actors to decrease the likelihood of a successful ransomware attack. Understanding how ransomware gangs establish interactions with actors in the ransomware ecosystem supports the creation of new interventions that could help companies create a better defense against ransomware. In addition, it could help target critical actors in the attacker sub-system to disrupt the value chain of ransomware gangs.

Numerous studies have studied the ransomware ecosystem, yet these studies often simplify the ransomware ecosystem by leaving out actors and interactions, and an actual definition of the ecosystem is never presented. Laszka et al. (2017) observed the ransomware ecosystem as a relation between groups of defenders attacked by ransomware and ransomware gangs and simplifies their interactions using a game-theoretic approach. They highlight that in future research, the role of cyber-insurance in the context of ransomware should be considered and that it should be researched how their policies would have effective outcomes. Next, Huang et al. (2018) observed the ransomware ecosystem as a collection of defenders, ransomware gangs, and money laundry service providers. In this study Huang et al. (2018) mention that an improved understanding of the ransomware ecosystem is a key first step to identifying new and potentially more effective intervention strategies, which fits within the general notion that the ransomware ecosystem is not yet correctly understood. Other authors have used the concept of ransomware ecosystem to describe a collective of ransomware programs and how these interact with their victims (Mei et al., 2021; Raheem et al., 2021). However, to our knowledge, no study researches the ransomware ecosystem from a ransomware gang's internal perspective, resulting in a lack of understanding of the ransomware ecosystem.

In sum, we observe two knowledge gaps: a) there is a lack of understanding of how ransomware gangs establish interactions with actors in the ransomware ecosystem, and b) There has been a lack of research that uses ground truth data due to ransomware gangs keeping their internal communication and operations hidden. Ground truth data may be a means to a better understanding of a ransomware gang's establishment of interactions. To fill these knowledge gaps, four research activities will be employed. First, we will review existing literature that studied the different actors involved in each of the different sub-systems of the ransomware ecosystem to lay the groundwork for the ransomware ecosystem. Furthermore, literature is reviewed on tactics, techniques, and procedures (TTP) and the known TTP used by ransomware gangs. Second, we will take an internal perspective using the leaked chatlog data of the Conti ransomware gang and empirically determine topics in their internal communication using Latent Dirichlet Allocation (LDA) algorithms. Third, these topics are mapped to Conti's TTP based on ground

truth communication data. Finally, these TTP are leveraged to reconstruct the ransomware ecosystem, creating a better understanding of how ransomware groups establish interactions within the ransomware ecosystem.

1.3 RESEARCH QUESTIONS

Concluding, the involved actors in the ransomware ecosystem and their interdependencies and how these construct the ransomware ecosystem is a blind spot in the current scientific literature. Through the following research question, the identified knowledge gaps that lead up to this blind spot can be filled:

> *To which extent can the ransomware ecosystem be reconstructed using ground truth communication data of the Conti ransomware gang?*

The focus of this thesis is to understand how ransomware gangs establish interactions within the ransomware ecosystem and therefore create a better understanding of the ransomware ecosystem. This focus is situated on a macro level, while the analysis of Conti's communication data is situated on a meso level. This can be explained since Conti has a representative role in the ransomware ecosystem because it is currently the most successful and largest ransomware gang active (ENISA, 2021). Hence, other ransomware gangs will likely base their tactics, techniques, and procedures on Conti to increase its success in attacking victims. In addition, Bátrla and Harašta (2022) argue that Conti DarkSide/Blackmatter, and Revil are highly representative for the current ransomware ecosystem. Therefore, Conti is a representative case study for studying other ransomware gangs' TTP.

Ransomware gangs such as Conti are often large well-structured organizations with hundreds of employees and rely on well-executed communication for their success. From Conti's leaked internal communication data, topics are distilled and subsequently mapped to TTP, which is a methodology that is commonly used in scientific literature to observe how ransomware gangs operate. Since it is stipulated that ransomware gangs intentionally establish interactions with actors in the ransomware ecosystem, it is likely that the tactics, techniques, and procedures distilled from Conti's communication cover the establishment of interactions within the ransomware ecosystem. Furthermore, since the Conti ransomware gang has a representative role in the ransomware ecosystem, their tactics, techniques, and procedures may be leveraged to reconstruct the ransomware ecosystem. Hence, the approach used in this thesis, in which we distill TTP from Conti's internal communication from a meso-level analysis to reconstruct the ransomware ecosystem from a macro-level analysis, is a valid approach.

Four sub-questions have been formulated to pinpoint the research activities required to answer the main research question:

SQ1: *Of which actors does the ransomware ecosystem comprise and how do they interact?*

SQ2: *What are overarching topics in Conti's communication data?*

SQ3: *How do these overarching topics map to Conti's tactics, techniques, and procedures?*

SQ4: *How can we leverage identified tactics, techniques, and procedures to reconstruct the ransomware ecosystem?*

1.4 THESIS STRUCTURE

Chapter 2 aims to answer the first sub-question by reviewing existing literature on the actors involved in the different sub-systems of the ransomware ecosystem. Additionally, literature is

reviewed on TTP and known TTP of ransomware gangs. Next, Chapter 3 discusses the methodology used in this thesis to answer sub-questions two, three, and four. Chapter 4 presents a high-level exploration of Conti's chatlog data which is used to generate topics from. Subsequently, Chapter 5 answers the second sub-question by discussing the topics that come forward from the LDA models. Section 6.1 aims to answer the third sub-question by mapping Conti's TTP based on the topics coming forward. Section 6.2 presents the leveraging of identified TTP to reconstruct the ransomware ecosystem, which allows us to answer sub-question four. Finally, the discussion, recommendations, and conclusions are presented.

2 RANSOMWARE ECOSYSTEM

This chapter reviews previous academic work on actors in the ransomware ecosystem and their interactions in four sections. The first section is tailored to discuss relevant actors and their interactions within the attacker sub-system of the ransomware ecosystem. The second section focuses on relevant actors and their interactions within the defender sub-system. The third section focuses on the actors and their interactions within the governance sub-system. The fourth section discusses tactics, techniques, and procedures and how these have been used within ransomware gangs. This chapter concludes with a short conclusion of the reviewed literature.

2.1 ACTORS AND THEIR INTERACTIONS WITHIN THE ATTACKER ECOSYSTEM

Ransomware literature is mainly focusing on research from a descriptive or technical perspective. Examples of such research can be found in Huang et al. (2018), Lee et al. (2019), McDonald et al. (2022), Mei et al. (2021), Raheem et al. (2021) and Richardson and North (2017). Richardson and North (2017) use a descriptive study to present a timeline of historical events to illustrate how ransomware has evolved. In this timeline, they illustrate the first examples of RaaS and illustrate how ransomware gangs evolved their tactics. In their research, they showed how tactics evolved to dynamic pricing being used by different ransomware programs based on the IP address of the infected computer. This illustrates how ransomware gangs are aware of the willingness to pay and adjust their tactics to increase the chance of the ransom being paid. However, Richardson and North (2017) argue that ransoms should not be paid since criminals talk to each other. Therefore a victim paying the ransom may become a target for other ransomware groups. Other scholars such as Beaman et al. (2021) and McDonald et al. (2022) used the work of Richardson and North (2017) to further describe more recent advances of ransomware. Beaman et al. (2021) highlight how ransomware gangs make use of remotely working connections in their tactics and discuss the known preventive measures: backing up, enforcing strict access control, and user awareness. McDonald et al. (2022) argue that ransomware gangs will continue to evolve their tactics, such as targeting disgruntled employees for inside distribution using RaaS or incorporating stealing sensitive files adding an additional layer of extortion, encouraging payment.

The preventative measures, as mentioned by Beaman et al. (2021) are often used in a game theoretic setting to research decreasing the economic incentive of defenders to pay ransoms and therefore decreasing the incentive for ransomware gangs to attack. Examples include the studies of Cartwright et al. (2019b), Fang et al. (2020), Galinkin (2021) and Laszka et al. (2017). Laszka et al. (2017) studied the interaction between organizations and ransomware attackers with a focus on modeling of security investment decisions for mitigation, i.e., level of backup effort as well as the strategic decision to pay or not pay a ransom. Fang et al. (2020) introduced the existence of fake ransomware gangs that do not care about benefitting from the ransom and showed that these can, in principle, make more money. However, the expected payoffs of both genuine and fake ransomware gangs increase with the chance of encountering genuine ransomware gangs. This implies that in most cases, ransomware gangs will try to decrypt the data when the ransom is paid, stimulating other defenders to pay the ransom.

Cartwright et al. (2019b) agree with the notion that, in general, ransomware gangs will put in the effort to decrypt the data. In their research, they mention that ransomware is rare in being a cybercrime that positively benefits from publicity and greater knowledge. The more individuals and organizations recognize that ransomware is a genuine extortion scenario in which access to files can only be regained through paying the ransom, the more willing defenders might be to engage with the ransomware gangs. Hence, fake ransomware gangs may be competing with genuine ransomware gangs since they decrease the general recognition of ransomware being a genuine extortion scenario in which files can be recovered. Cartwright et al. (2019b) argue that there may also be competition between the different genuine ransomware gangs, which contradicts the findings of Richardson and North (2017) saying ransomware gangs share vulnerability information. If ransomware gangs compete, they may not be willing to share information about attacked victims since this could lose a competitive advantage.

In their study, Cartwright et al. (2019b) mention that ransomware actors are aware of the state of backups of the organizations they attack, contributing to the general thought of ransomware gangs precisely planning their attacks regarding whom to attack and when. Galinkin (2021) builds on the work of Laszka et al. (2017) but rather sees the interaction of ransomware gangs and defenders as playing in a lottery than a strategic game between the two. Furthermore, he argues that there are three variables to influence if one tries to disrupt the ransomware economy: the value of payments, the cost of operating ransomware, and decreasing the probability of payment. The value of payments is hard to influence looking at the only increasing value of cryptocurrencies such as Bitcoin, Ethereum, and Monero. Even if we could influence the value of payments of these cryptocurrencies, it would not stop ransomware gangs from shifting to other payment methods. Furthermore, Galinkin (2021) explains how RaaS has a great effect on the costs of operating ransomware since it requires minimal financial investment to become an affiliate for other ransomware gangs.

Other scholars have also touched upon the concept of RaaS in their work. Bayoumy et al. (2018) has laid the basis of RaaS distribution in dark web markets by studying the dark web ecosystem for ransomware. In their research, they argue that the development and distribution of ransomware are stimulated by social networks on the dark web, and that dark web meeting places and marketplaces are a key environment for cybercriminals, allowing access to the skills and expertise of other cybercriminals in the ransomware ecosystem. In the attacker ecosystem, they identified three RaaS stakeholders: RaaS authors, RaaS vendors, and RaaS distributors. In addition, Bayoumy et al. (2018) argue that it may be possible that the different dark web markets are connected. These findings imply that the ransomware gangs may interact with RaaS stakeholders through dark web meeting places and marketplaces to stimulate their affiliate program.

Meland et al. (2020) continued this work by studying RaaS within the dark web and describe that RaaS is a way of democratizing crime, giving ordinary people and smaller players an easier way into the criminal market while reducing the risk of exposure for the one on top of the value chain. Furthermore, a dissatisfied employee might decide to partner up with a RaaS developer, which helps ransomware gangs effectively infect an organization from the inside and split the resulting profit. In their work, Meland et al. (2020) illustrate how dark web markets have a role in the distribution of RaaS and, therefore, in the ransomware ecosystem. The active vendors on these dark web markets are shown to be resilient from being shut down by law enforcement and quickly switch to new dark web markets. Important factors for the success of vendors and therefore for their resilience are trust and reputation. If a vendor has a high reputation, it's more likely to be successful in new dark web markets, especially since most of the time, these reputations are transferable when markets are shut down. This was when Alphabay and Hansa were shut down, and vendors shifted to Empire.

Although many scholars mention that RaaS allows anyone with a computer to attack defenders with ransomware, Mei et al. (2021) and Van Wegberg et al. (2018) show that easily accessible dark web markets do not play as big of a role as often assumed. They argue that most RaaS

items sold on the dark web markets are believed to be frauds and that there is a lack of strong growth for commoditization in these markets. This is illustrated by a moderator of a dark web forum:

"The public space is supposed to be filled with scams and stupid products because you don't have to prove your worth to get into the public sphere. The only way to experience the inner workings is to be able to convince others that you should be allowed into invite-only spheres as mentioned"

Gaining access to these invite-only spheres can be challenging, and therefore RaaS may only have a modest effect. Moreover, there are few items for sales, and the successful sales do not indicate a large economy of RaaS in dark web markets. These findings illustrate that it is likely that ransomware gangs use invite-only spheres for their affiliate programs and that large-scale RaaS programs may not be distributed over easily accessible dark web meeting places and marketplaces. Finally, Meland et al. (2020) created an overview of the value chain for the RaaS economy, as shown in figure 2.1.

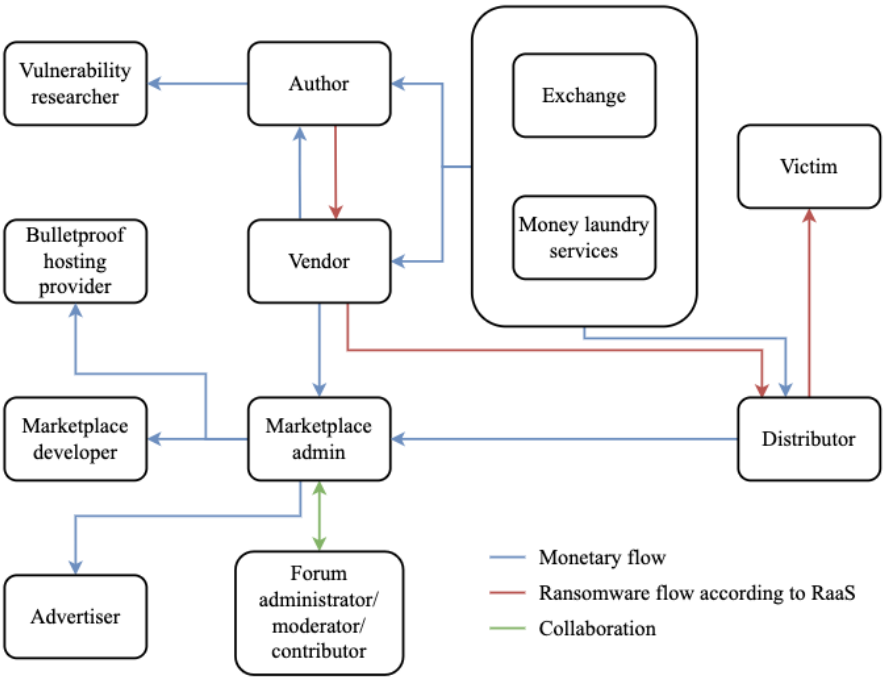

Figure 2.1: Value chain for the RaaS economy adopted from Meland et al. (2020)

In the value chain for the RaaS economy, a vulnerability researcher discovers and sells zero-day vulnerabilities to RaaS authors. These vulnerability researchers have high expertise in hardware and software, and Meland et al. (2020) argue that many vulnerability researchers have system administrator roles in respected companies. Authors are professional developers that create ransomware programs that take advantage of vulnerabilities, which can be purchased from vulnerability researchers. Ransomware gangs can be observed as authors but may also take the role of vendors. Meland et al. (2020) argue that there can be fierce competition between different malware authors. Vendors take a role in the RaaS value chain by marketing and selling RaaS on dark web marketplaces or their own private websites. Vendors may also be authors, as in the case of ransomware gangs, but Meland et al. (2020) argue that most dark web vendors have little programming experience and sell a wide range of goods from drugs to guns and ransomware. Therefore, it is likely that RaaS vendors active in dark web markets and meeting places are not the actual developers of the ransomware themselves.

The distributors buy RaaS through a vendor and can be observed on the dark web. They share their experiences and feedback on ransomware purchases using reviews and forums. In addition, some distributors use these forums to seek partnerships involving ransomware developers

or offer vulnerability information for sale. Defenders are attacked with ransomware and may lose their data as a result. To prevent data loss, they may need an exchange to obtain the ransom amount in cryptocurrency. The marketplace admins provide a platform that vendors and distributors use to trade RaaS. Meland et al. (2020) argue that these marketplace admins should be a trusted third party that governs the monetary transaction. However, many examples are known in which marketplace admins run off with the money (exit scams) (Meland et al., 2020). Law enforcement agencies put a lot of effort into shutting down these markets, and competing marketplaces may try to eliminate competition.

Meland et al. (2020) contribute to the role of Bulletproof Hosting Providers, money laundry services, and exchanges in the RaaS value chain. Bulletproof Hosting Providers (BHP) host the website services of the dark web markets and often host the command control servers of ransomware vendors. By using these BHPs, the resistance against being shut down for dark web markets and ransomware vendors increases. Money laundry services like mixers and money mules are used to launder the earnings from ransomware operations. However, Monero is often used, which is said not needed to be mixed due to its anonymous character (Meland et al., 2020). Exchanges have an important role in the RaaS value chain for cybercriminals since they offer currency exchange services, which allows them to spend their profits. The role of exchanges and money laundry services was previously highlighted by Huang et al. (2018), showing that unique clusters of ransomware programs and their trail to money laundering can be identified. Furthermore, Huang et al. (2018) argue that these exchanges not only have an important role but should also be seen as victims since they too can be victims of ransomware attacks.

Keshavarzi and Ghaffary (2020) use a descriptive perspective to present a chain of attack events in ransomware and illustrate this attack chain according to two ransomware programs. In this research, they address the importance of botnets in ransomware, which is often used for distributing ransomware through phishing. Given that phishing is one of the major infection vectors, the issue of spam-sending botnets must be considered as one of the key actors involved in distributing ransomware. A botnet is a network of a large number of computers that have been compromised, mostly unnoticed by the user. It is under the control of a botherder that is able to let its botnet perform tasks such as sending phishing emails or crypto jacking (Keshavarzi and Ghaffary, 2020).

Combining the insights from previous scholars in their ransomware research, it becomes evident that ransomware gangs interact with different actors to construct the RaaS value chain. Exchanges, money laundry service providers, BHPs, vulnerability researchers, dark web market- and meeting places, dark web vendors, and botnets are all actors that comprise the attacker ecosystem besides ransomware gangs. These actors interact with each other by providing financially motivated illegal services, as illustrated in the ransomware value chain. Different ransomware gangs may be competing and, therefore, not always willing to cooperate with other ransomware gangs, although scholars present signs of collaboration and competition. Innersphere dark web markets and forums may have a great impact on RaaS by bringing together the different actors to construct the RaaS value chain, while publicly available markets and forums are shown to have a limited effect on the distribution of RaaS. In the next section, the literature on the defender site of the ransomware ecosystem is reviewed.

2.2 ACTORS AND THEIR INTERACTIONS WITHIN THE DEFENDER ECOSYSTEM

In contradiction to the interactions of actors within the attacker ecosystem, little is known about the interactions and actors within the defender ecosystem. As argued by Chen et al. (2021), most ransomware studies take a technical and descriptive perspective, illustrating the techniques employed in such attacks, developing detection approaches, and suggesting best practices for defenders. Prior literature has barely explored how defenders make the decisions of security

investment and ransom payment. However, Chen et al. (2021) argue that independent decisions about their security investments often result in incentive misalignment issues where system optimality in a defender's investment level is not achieved. This can be observed as a type of free rider problem.

Moreover, Chen et al. (2021) argue that the ransomware gangs respond to measures taken by defenders by lowering ransoms and lowering the attack rate slightly. This would imply that ransomware gangs are aware of measures taken by defenders. Laszka et al. (2017) agree with the general notion that ransomware gangs are aware of pervasive investments in backup technologies and that they focus on victims that do not back up. Another example of free riding that Cartwright et al. (2019b) address is smaller firms benefiting from larger firms spending more on cybersecurity, meaning they may spend less on the deterrence of ransomware attacks. However, Cartwright et al. (2019b) find it unlikely that victims would consider this when determining the cybersecurity budget.

Cartwright et al. (2019a) argue that individual firms are relatively risk-seeking and that a large portion of the subjects in their study is therefore investing more in recovery than in prevention. This is arguably a reflection of the reality in the ransomware ecosystem. Since most defenders are relatively risk-seeking, a dominant strategy would be to take insurance instead of investing in preventative measures. However, as Kenneally (2021) and Galinkin (2021) argue, cyber-insurers are fueling the ransomware economy since the ransomware gangs still get paid, and insurance does nothing to lower the actual problem. Another important interaction among defenders is the possibility for cascading effects in the supply chain because of cascading service disruption as identified by Pal et al. (2021). An example of cascading effects is when important infrastructure such as the electricity grid is attacked by ransomware. The disruption of the electricity grid then harms other defenders dependent on the electricity grid. In addition, customers and partners in the value chain of an attacked defender can be harmed by being extorted with their confidential files found in the defender's network, such as customer blueprints (ENISA, 2021).

2.3 ACTORS AND THEIR INTERACTIONS WITHIN THE GOVERNANCE ECOSYSTEM

Like the defender ecosystem, the governance ecosystem is barely covered in scientific literature. This lack of coverage is mainly due to the governance- and legal framework of governing ransomware being loose. As Kenneally (2021) argues, there is a loose legal framework for payment transparency, meaning that defenders do not have any consequences when they are not transparent about paying the ransom. To govern the ransomware ecosystem, it is important to know when a ransom is paid so that we can have full information on attacks and subsequently learn from them. Because of this loose legal framework, many defenders hide that they paid the ransom to protect from reputational damage, which is often followed by a significant drop in stock price. In addition, Kenneally (2021) argues that there have been no civil penalties against defenders, insurers, or response firms for paying the ransom. This illustrates the lack of an adequate governance framework and includes the lack of policies and processes for cyber-insurers to bring about sufficient risk management coordination or implementation incentives (Kenneally, 2021). This is illustrated by cyber-insurers failing to demand back-ups at defenders, resulting in them reimbursing the costs for ransoms which is in the short term cheaper but in the long term fuels the ransomware economy, leading to additional attacks.

McDonald et al. (2022) add that cyber-insurance covering ransoms is the most influential incentive to pay ransoms to ransomware gangs. They argue that of all the organizations that paid the ransom, 94% were reimbursed through their insurance, showing that the cyber-insurers may indeed be a driver of the ransomware economy. Bateman (2020), Fang et al. (2020), Kenneally (2021) and Laszka et al. (2017) all argue that the role of cyber-insurance should be better re-

searched because they may create a stable market of financial incentives for organizations to better adapt cyber hygiene. In addition, cyber-insurers could push the use of certain cybersecurity vendors by having creative policy premiums. However, Richardson and North (2017) believe that governments and regulatory organizations should create incentives to adopt preventative measures such as regularly creating back-ups.

As Chen et al. (2021) argue, promoting and advocating for defenders to refuse to pay ransoms may only be helpful if it is advertised what the benefits are of such a strategy. The defenders are most likely to adopt mitigation strategies if they can recognize that the anticipated benefits outweigh the anticipated costs in the long run. That is, the defenders will only refuse to pay ransoms if they understand that their long-term costs because of additional attacks are higher than their short-term gain of recovering their files quickly. Although governmental and regulatory organizations could help disrupt the ransomware ecosystem by influencing both defender and attacker actors, it could also stimulate ransomware attacks through state-sponsored attacks (Lee et al., 2019; Keshavarzi and Ghaffary, 2020). WannaCry, for example, is known to be used by North Korean hackers for cyber-military operations. In addition, political organizations may use ransomware to let certain victims post a political statement before decrypting their files.

2.4 TACTICS, TECHNIQUES, AND PROCEDURES (TTP)

Tactics, Techniques, and Procedures (TTP) is a concept that is frequently used in the cybersecurity field to describe the behavior or modes of operations of cybercriminals (Egloff and Smeets, 2021; Yeboah-Ofori and Islam, 2019). TTP leverage specific capabilities, behaviors, and exploits that ransomware gangs can use on defenders. These TTP are relevant to identifying attack patterns and resources deployed to better understand the ransomware gang's operations in the ransomware ecosystem. In addition, TTP can give insights into the motives of ransomware gangs, the intended effects, and the eventual impact on defenders. Yeboah-Ofori and Islam (2019) argue that tactics describe how ransomware gangs operate during the ransomware campaign. They add that this includes how ransomware gangs conduct reconnaissance for initial intelligence gathering, how the information is gathered, and how the compromises were conducted. Sending a spear-phishing email to a specific group of employees within a defender's firm is an example of a tactic. Spear-phishing is a more focused and personalized form of phishing in contradiction to the well-known mass-mail phishing campaigns (Parmar, 2012).

Yeboah-Ofori and Islam (2019) describe techniques as the strategies that are being used to facilitate the compromise of a defender's files. These include tools, skills, and the capabilities deployed. Furthermore, how ransomware gangs establish control, maneuver and obfuscate through a defender's IT infrastructure, and steal data are part of the techniques used. An example of a technique could be the ransomware gang concealing the malicious contents of a spear phishing email, making it harder to detect. Procedures are about how ransomware gangs implement the techniques and tactics to achieve an objective, generally performing a successful ransomware attack but may be a sub-objective (Egloff and Smeets, 2021; Yeboah-Ofori and Islam, 2019). Many examples of TTP of ransomware gangs can be found in scientific literature. However, since discussing all of these can be a study on its own, we only discuss a handful of examples while looking at the evolution of ransomware.

In a recent study Oosthoek et al. (2022) categorize ransomware into two categories: commodity ransomware and ransomware as a service (RaaS). Commodity ransomware is often seen in the early years of ransomware. It is characterized by widespread targeting, fixed ransom demands, and tech-savvy ransomware gangs focusing on single devices. These ransomware gangs are responsible for both the development of the ransomware and its spreading. The actual development of commodity ransomware itself is often based on preexisting work by building on leaked or shared source code. The modus operandi of commodity ransomware was rather mass exploitation than targeting specific defenders. This is illustrated by the delivery vectors being mainly phishing and exploiting vulnerabilities in text editors and spreadsheet software or

through malicious executables. Well-known examples of these types of ransomware are WannaCry and NotPetya. However, NotPetya is often claimed to be 'fake ransomware' that is rather being used as a political weapon than financially motivated (Kaminska et al., 2021; Mos and Chowdhury, 2020; Ren et al., 2020). Creating proper backups was the generally advised mitigation strategy. Still, as Oosthoek et al. (2022) argue, it is evident that commodity ransomware is just a proving ground for a higher-impact utilization of ransomware. From this can be concluded that tactics used in commodity ransomware are widespread targeting and using fixed ransom demands, while a technique that is used is the exploiting of vulnerabilities in text editors and spreadsheet software.

Ransomware as a service (RaaS) has been previously discussed as a group of ransomware developers that license their ransomware on an affiliate basis. As Oosthoek et al. (2022) argue, they often provide a payment portal which is typically provided over the Onion Router (Tor), an anonymous web protocol. This portal allows for negotiating with victims and a dynamic generation of payment addresses (typically Bitcoin or Monero) (ENISA, 2021). These RaaS actors often use double extortion schemes, meaning that they not only encrypt a victim's data but also threaten to publish their confidential data publicly. The evolution of ransomware to the RaaS model allowed existing ransomware gangs to shift to a more lucrative business model where lower-skilled affiliates gain access to exploits and techniques previously reserved for the tech-savvy ransomware gangs (Oosthoek et al., 2022). This was illustrated by the Conti playbook leaked in August 2020 and represents a manual for Conti's affiliates on how to compromise business networks. However, the affiliates of RaaS ransomware groups can have different approaches.

Affiliates may scan the entire internet and try to compromise any possible victim and engage in price discrimination after identifying the victim. The price is often determined based on the victim company's size and can be justified by compromised financial documents (Oosthoek et al., 2022). A different strategy is known as big game hunting, in which affiliates target large firms that can afford to pay high ransoms. In contradiction to commodity ransomware, RaaS ransomware gangs often rely on spear-phishing as a delivery vector. Furthermore, they exploit recently disclosed vulnerabilities which makes remote desk protocols (RDPs) a vulnerability. As multiple authors argue, cryptocurrencies enable ransomware groups to effectively monetize the large-scale compromission of victims' IT systems (McDonald et al., 2022; Oosthoek et al., 2022; Raheem et al., 2021; Richardson and North, 2017). From these findings can be concluded that in the evolution to RaaS ransomware groups have used different TTP. An example of a tactic and technique that come forward is using double extortion schemes and using Tor for setting up payment portals.

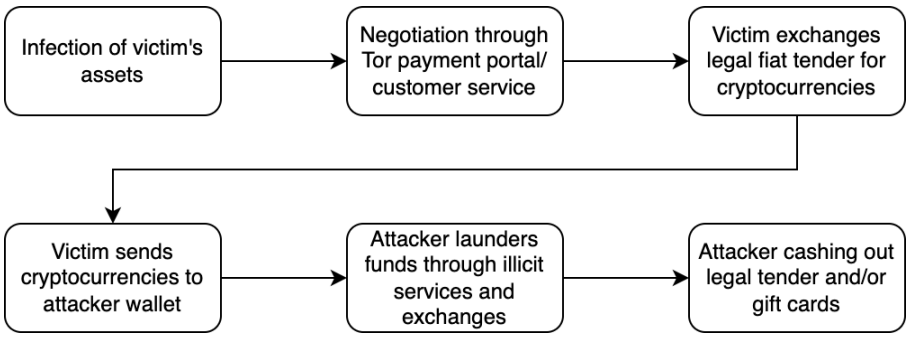

Figure 2.2: Procedure of ransom payment and its laundering (Oosthoek et al., 2022)

Additionally, Oosthoek et al. (2022) identified the procedure of how ransom payments are executed and laundered, which is visualized in figure 2.2. First, the defender's assets are infected, and a ransom notice is displayed, followed by negotiation through the payment portal. Next, the defender exchanges legal fiat tender for cryptocurrencies such as Bitcoin, Ethereum, or Monero and sends the amount to the ransomware gang's wallet. From here, the obtained cryptocurrencies are often routed through various illicit services such as mixers to obfuscate ownership and

reduce the risk of being traced. Finally, the ransomware gang cashes out the cryptocurrencies in legal tenders and/or gift cards.

To conclude, TTP is a concept often used to describe the behavior or modes of operations of ransomware gangs. Tactics describe how ransomware gangs operate during the ransomware campaign, techniques can be seen as the strategies that are being used to facilitate the compromise of a defender's files, and procedures describe how ransomware gangs implement the technique and tactics to achieve an objective, which may be a sub-objective to performing a successful ransomware attack. Tactics that come forward from literature are widespread targeting, using fixed ransom demands, and big game hunting. In addition, a technique that is known to be used is the exploiting of vulnerabilities in text editors and spreadsheet software.

2.5 CONCLUSION

This chapter aims to determine which actors comprise the ransomware ecosystem and how they interact by reviewing previous academic work. Actors that comprise the attacker ecosystem besides ransomware gangs are exchanges, money laundry service providers, BHPs, vulnerability researchers, dark web market- and meeting places, dark web vendors, and botnets. It becomes evident that these actors interact with ransomware gangs and other actors to construct the RaaS value chain by providing illegal services. The RaaS value chain by Meland et al. (2020) can be used as a guideline to see how actors in the attacker's sub-system interact from an economic lens. These actors are most likely interacting through inner-sphere dark web markets and forums, as it is shown that publicly available markets and forums have a limited effect on the distribution of RaaS. The different ransomware gangs may be competing, although scholars have also presented arguments for ransomware gangs collaborating in sharing vulnerabilities. Ransomware gangs interact with defenders by performing reconnaissance and ransomware attacks since it is argued that ransomware gangs are aware of measures taken by defenders and that they adjust their TTP based on these.

Actors that comprise the ransomware ecosystem from a defender ecosystem perspective are defenders and cyber-insurers. Defenders may show free-riding behavior by leaning on the cybersecurity spending of larger firms, although it is unlikely that they take this into account when budgeting. Furthermore, defenders may interact with other defenders by disrupting their businesses through cascading effects in their supply chain or by them being extorted through comprised customer files such as blueprints. Defenders are generally choosing to pay for the ransom and getting insurance that covers ransoms over prevention. National and international governmental organizations that create and maintain the governance framework and the other actors in the ransomware ecosystem comprise the ransomware ecosystem from a governance ecosystem perspective. These governance actors interact with defenders and cyber-insurers, for example, by setting rules and regulations for payment transparency or by creating incentives to adopt preventative measures. Governance actors can interact with ransomware gangs by stimulating ransomware attacks through state-sponsored attacks. Arguably, cyber-insurers have an important role in the ransomware ecosystem. However, their exact role is still unknown since they may fuel the ransomware economy and could create a stable financial market for adopting better cyber hygiene among defenders. The insight from the state-of-the-art literature provides a basis for the ransomware ecosystem, which can be complemented using insights from Conti's internal communication.

3 METHODOLOGY

In section 2.1, 2.2, and 2.3 the basis for the ransomware ecosystem is laid by reviewing scientific literature on what actors comprise the ransomware ecosystem and how these actors interact. In section 2.4 the concept of tactics, techniques, and procedures (TTP) is explained, and examples of known TTP for ransomware gangs are presented.

This chapter lays out the methodology used to further research how ransomware gangs establish interactions with these actors in the ransomware ecosystem, taking the Conti ransomware gang as a case study. Subsequently, these findings are used to reconstruct the ransomware ecosystem. This chapter is structured as follows. Section 3.1 lays out the methodology used to empirically determine overarching topics in Conti's internal communication data using LDA topic modeling. Section 3.2 discusses how these topics are used to map Conti's tactics, techniques, and procedures (TTP) and how the ransomware ecosystem can be reconstructed based on these. This chapter concludes with a short conclusion of the methodology used in this thesis to answer the research questions and fill the knowledge gaps.

3.1 DETERMINING OVERARCHING TOPICS IN CONTI'S INTERNAL COMMUNICATION DATA

Section 1.2 illustrated the lack of understanding of how ransomware gangs establish interactions with actors in the ransomware ecosystem and how scientific research on ransomware has been lacking the use of ground truth data. To fill these knowledge gaps, this thesis uses Conti's leaked internal communication data to empirically determine overarching topics in Conti's communication. Cybersecurity journalists and cybersecurity companies have been publishing articles on Conti's leaked communication data based on them reading the messages (Checkpoint, 2022; Krebs, 2022a,b). However, these articles lack a structured and empirical methodology, and therefore their findings are primarily based on their biases as researchers regarding what they find most interesting. The goal of determining overarching topics in Conti's internal communication is to use a structured methodology to empirically determine what the main topics are that are internally discussed within Conti. The primary advantage of using such an approach is that the resulting findings are not primarily biased towards the researcher's interpretation and allow for empirically researching a large dataset of messages on topics discussed. These resulting topics can subsequently be used to map Conti's TTP to empirically observe how ransomware gangs internally operate since it is logical to assume that topics regarding Conti's TTP are internally discussed. The reason for this is Conti is a large well-structured organization that relies on well-executed communication, as previously argued in section 1.3.

To empirically determine topics using a structured methodology, this research utilizes Latent Dirichlet Allocation (LDA) topic modeling. LDA topic modeling is chosen since it is simple and has been used in various sciences for topic modeling of text corpora, which can be used as a type of text summarization of large sets of documents (Porter, 2018). Furthermore, previous studies have used unsupervised topic modeling using LDA on other large cybercriminal datasets such as dark web markets' subreddits and carding forums (Kigerl, 2018; Porter, 2018). Hence, topic modeling using LDA can give insights into larger datasets such as Conti's leaked internal communication data. In the rest of this section, it is discussed how LDA topic modeling

is utilized following a structured methodology to empirically determine overarching topics in Conti's leaked internal communication data.

Section 3.1.1 discusses the LDA algorithm and its parameters on a high level and how it generally functions to model topics of text corpora. Section 3.1.2 briefly discusses Conti's leaked communication data, and section 3.1.3 presents the methodology used for further high-level exploration of this data. Following the high-level analysis of Conti's communication data, the methodology used for determining overarching topics consists of five steps. A graphical summary of the methodology is presented in figure 3.1. Section 3.1.4 discusses the first three steps in which the chatlog data is first split into three datasets to allow clustering messages to increase the performance of the LDA algorithm. Secondly, in two of the three datasets, the messages are clustered, and one of the three datasets is not since this is not meaningful for analysis. The reasoning for splitting the data into three different datasets and clustering two of these is further elaborated on in section 3.1.4. Next, each dataset is pre-processed using the same methodology. Section 3.1.5 discusses steps four and five in which first the LDA model parameters are set, and LDA topic modeling is performed for each dataset. In the final step, the resulting topics are labeled using pyLDAvis and four different relevancy metrics settings.

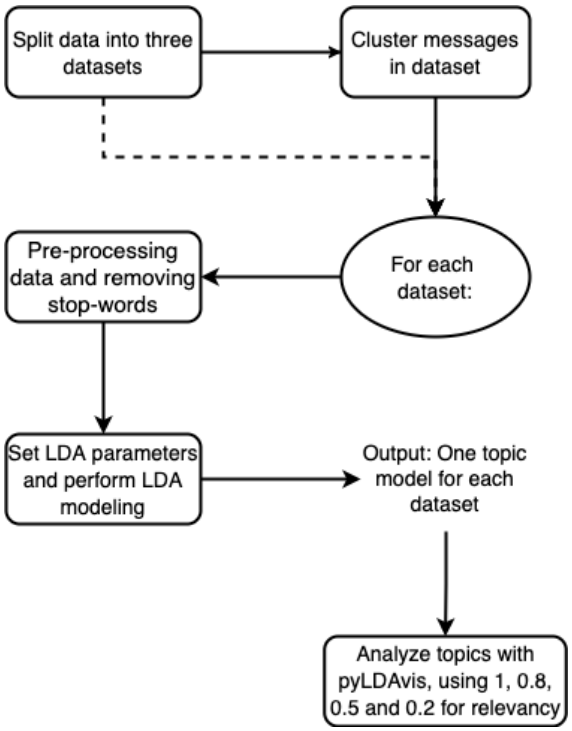

Figure 3.1: Summary of the methodology used for determining overarching topics in Conti's communication

3.1.1 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is an algorithm that can be used to perform unsupervised topic modeling and is frequently used in various research (Blei et al., 2003; Kigerl, 2018; Porter, 2018; Maier et al., 2018). LDA is an unsupervised technique, meaning it is unknown beforehand what the correct cluster categories are (Kigerl, 2018). LDA produces a model of a corpus of documents in which documents in this thesis refer to a message or a set of messages sent by members of the Conti ransomware gang. The composition of these documents will be further highlighted in section 3.1.4. The LDA model assumes that each document in the corpus is derived from a generative process where each document consists of a distribution of a finite set of topics (Porter, 2018). Each topic is a multinomial distribution of the vocabulary of words in the corpus while each word of the document is drawn from a topic in the generative process.

In other words, the LDA model identifies topics in a corpus of documents and represents each document as a distribution of these topics, while a topic is a distribution of words.

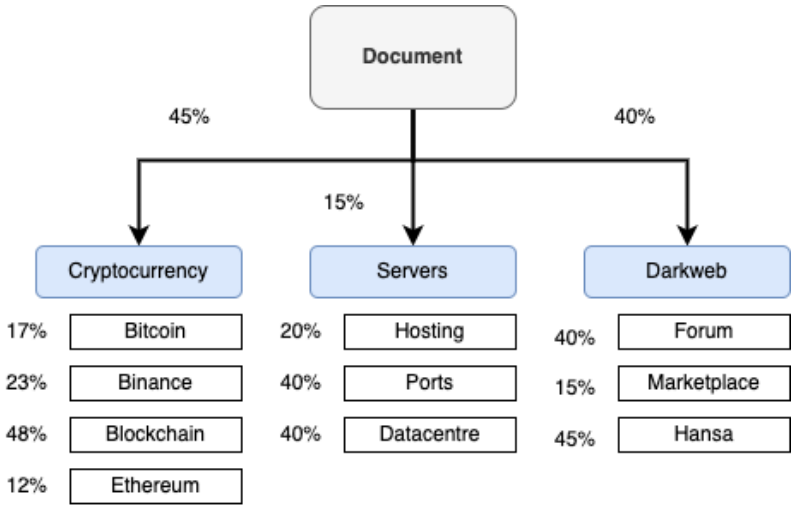

Figure 3.2: Schematic overview of the LDA document generation assumption

To further illustrate how the LDA algorithm assumes the generation of documents, figure 3.2 shows an example of a document with a vocabulary of ten words. In this example, the document is assumed to be a distribution of the three topics cryptocurrency, servers, and dark web, and the topics consist of a distribution of words. By observing actual words in each document, the LDA algorithm estimates the topic distribution per document and the word distribution per topic. This is done based on hyper-parameters α, β and k. The values α and β respectively represent the distribution of words over topics and the distribution of topics over words, and k represents the number of topics. A high value for α would indicate that each document has a relatively even distribution of the topics, while a low value would indicate this distribution is rather sparse. Similarly, a high value of β indicates that topics are an even distribution of the vocabulary of words, while a low value for β represents a sparse distribution of topics per document (Blei et al., 2003; Porter, 2018; Kigerl, 2018). The default values for α and β of 1 divided by the number of topics are used in this methodology as this is done in a similar study by Porter (2018). The number of topics k is an input parameter of the LDA model. The optimal number of topics in which topics are most coherent can be determined using a methodology that will be discussed in section 3.1.5.

Consequently, the LDA algorithm outputs the distributions of topics per document and the distribution of words per topic. Examples of these LDA outputs are presented in table 3.1 and table 3.2. From the distribution of words per topic, topics in a text corpus can be identified. The most relevant terms associated with each topic can be determined based on their distribution score. Using the distribution of topics per document, documents can be assigned to topics based on their distribution score. This results in a set of topics associated with a list of the most relevant terms. Furthermore, each document can be assigned to one or more topics based on their distribution scores. For a more extensive elaboration on the LDA algorithm see the work of Blei et al. (2003), Kigerl (2018), Kim et al. (2019) and Porter (2018). In the next sections, we discuss Conti's leaked communication data and how we can apply LDA to determine topics in this data empirically.

3.1.2 Conti's Jabber- and Rocket communication data

Conti is currently one of the most dominant ransomware gangs (ENISA, 2021; Oosthoek et al., 2022). Some tactics, techniques, and procedures (TTP) came to light when the "Conti playbook" was leaked in August 2021 (Cisco Talos, 2021). Consequently, it is known that Conti uses the RaaS model for additional revenue streams and uses a double extortion scheme in their ran-

Table 3.1: Example LDA output of per-topic word distributions (Kim et al., 2019)

	Topic 1	**Topic 2**	**Topic 3**	**...**	**Topic K**
Word 1	0,10	0,50	0,20	...	0,10
Word 2	0,40	0,04	0,01	...	0,30
Word 3	0,05	0,22	0,52	...	0,05
...	...	...	...	...	...
Word N	0,11	0,22	0,36	...	0,1

Table 3.2: Example LDA output of per-document topic proportions (Kim et al., 2019)

	Topic 1	**Topic 2**	**Topic 3**	**...**	**Topic K**
Doc 1	0,01	0,05	0,02	...	0,10
Doc 2	0,02	0,04	0,01	...	0,03
Doc 3	0,04	0,11	0,09	...	0,03
...	...	...	...	...	...
Doc N	0,06	0,01	0,02	...	0,05

somware attacks. Hence, Conti does not only encrypt the data but also steals data and threatens to leak the stolen files publicly. For both decrypting the data and a guarantee of not leaking the data, different ransoms are asked (Oosthoek et al., 2022; Tuttle, 2021). In February 2022, chatlog files containing over 160.000 messages send among Conti members between June 2020, and March 2022 were leaked. These messages were sent using two chat services: a Jabber server and a Rocket server.

Jabber is an open-source instant messenger based on the XMPP protocol with thousands of independent servers. Furthermore, it is known to be used by cybercriminals because of its support for strong encryption (Keshavarzi and Ghaffary, 2020). A well-known Jabber server is the Exploit.im server run by the community at Exploit.in, a Russian cybercrime forum of which joining requires a certain level of vetting or payment. Besides the Jabber server, Conti also uses the Rocket.Chat platform for communication. Rocket.Chat is a free chat service that allows communication in different channels, similar to Slack but offers an on-premises solution that allows it to be run on private servers (Rocket.Chat, 2022). The on-premises solution provides a high level of privacy which is crucial for the continuity of a ransomware gang. In addition, Rocket.Chat can incorporate customer service communication which may indicate that the Rocket server is used for communication with affiliates or defenders (Rocket.Chat, 2022).

Each message in the chatlog files has the sender, the receiver, the original message, the translated message (to English), and the original language of the message. In addition, each message has a Unix timestamp associated with it which allows us to analyze the activity of the different chat services over time. The messages in the Rocket.Chat chatlog file have additional information such as if the message was pinned, the attachments in the message, URLs sent in the message, and replies to the message. In this thesis, a translated version of the chatlog files is used for analysis. Northwave Security translated the original chatlogs, and these translated chatlog files are publicly available on GitHub (Northwave Security, 2022). The Northwave translation of the chatlogs consists of three separate CSV files: one file for the logged communication in the jabber server in 2020, one for the logged communication in the jabber server in 2021 and 2022, and one file for the logged communication in the Rocket server.

3.1.3 High-level exploration of the chatlog data

To better understand Conti's organization and operation and to understand how the chat services are used differently, a high-level exploration of the chatlog data is performed. This understanding helps us to interpret the topics and messages better, resulting in a higher quality of topics and TTP. First, we present descriptive variables for each of the three chatlog files. Sec-

ondly, we observe how the activity of Conti within both chat services is laid over time. This is done by plotting the total number of messages sent per month, weekday, and hour.

Next, to determine how the two chat services are being used in Conti's operations, the active users are sorted on the number of messages they have sent and received. Subsequently, the conversations between the most frequent senders for each chat service are visualized as a network. In these networks, the nodes represent users, and edges represent the existence of a conversation between users in the chatlogs. A directed edge from node A to node B exists if the user representing node A sent a message to the user representing node B. This high-level analysis shows that the two different chat services have been used for different purposes. The Jabber chat service is mainly used for peer-to-peer communication, and the Rocket chat service for group communication in multiple sub-channels and a general channel. This implies that these chat services should be analyzed separately, and since it is assumed that the sub-channels may be used for different purposes, we analyze these separately from the general Rocket channel. Hence, three different chatlog datasets are used in this thesis. Chapter 4 of this thesis provides a more detailed overview of the chatlog data and discusses the results of this exploration.

3.1.4 Pre-processing the chatlog data

The first step in pre-processing the chatlog data is to transform the messages in the Jabber chatlog data into a collection of threads between the possible combinations of two users so that the conversations between a pair of users form the documents in which topics are determined. Clustering the chatlog data decreases the model runtime and improves the performance of the LDA algorithm determining topics since it is known that the LDA algorithm has more difficulty with the data sparsity in short texts (Yan et al., 2013). We used this specific method of clustering since we believe that it is logical to assume that certain pairs of users discuss similar topics. This methodology of clustering messages into larger texts allows pairs of users that discuss similar topics to be clustered based on the distribution of topic scores. However, we leave this out of the scope of this thesis and leave this for future research.

The high-level exploration of the chatlog data (see chapter 4) indicated how the two chat services are used differently. Since the Rocket chat service is used for group communication in different channels, the Rocket chatlog data is split into two segments of which, one containing the messages sent to different sub-channels and one containing the messages sent to the general channel. Similar to the Jabber chatlog data, messages sent to sub-channels are clustered into larger texts to improve the performance of the LDA algorithm (Yan et al., 2013). Messages sent to a sub-channel are clustered so that the collection of messages to each sub-channel forms a document. Using this approach, the LDA algorithm determines topics in the messages sent to each channel. We hypothesize that sub-channels may be used for different purposes. By clustering messages to the sub-channel they are sent in, these purposes can be identified using LDA to determine topics in each sub-channel.

The collection of messages that were sent to the general channel was not clustered into larger texts. The reason for not clustering these messages is that the conversations take place between a large group of users in the general channel, and it is assumed that the topics coming forward from this channel vary widely. Therefore, in contradiction to the other two datasets clustering the messages does not lead to a logical level of analysis. Consequently, for the messages sent to the general channel, each message forms a document from which topics are determined. This provides us with three different datasets from which overarching topics can be determined using LDA. Each dataset is first pre-processed using the same methodology, inspired by similar studies in which LDA is used to determine overarching topics (Kigerl, 2018; Maier et al., 2018; Porter, 2018; Waal et al., 2008). This led to the following method of pre-processing:

- Replace hypertext markup language (HTML) code, special characters, unencrypted messages, domain names and URLs, Bitcoin addresses, and emails with a single space

- Remove words with more than 20 characters
- Expand English contractions
- Replace characters not in A-Za-z0-9 with a single space
- Lowercase text
- Replace trailing whitespaces with a single space
- Replace line breaks in the documents for a single space
- Tokenize the data by spaces, and remove common stop-words (using the NLTK Python library) and tokens only occurring once
- Perform lemmatization on the tokenized data

As the language used in the chatlogs is rather informal and contains a lot of slang, the standard set of stop-words needed to be extended as done in similar studies, e.g., by Kigerl (2018) and Porter (2018). Stop-words are words that occur frequently in texts but add little semantic value, such as "a," "the," and "do." Each document is essentially either a conversation among two Conti members using colloquial language, a collection of messages sent to one of the subchannels using similar colloquial language, or a message sent to the general Rocket channel. Therefore, the initial results from inputting the pre-processed chatlogs before extending stop-words led to topics focusing on these colloquial terms. The generated topics revolved around slang, swearing words, cryptic terms, or common words such as "something," "bro," "chvv," "like or "dude." Identifying these relatively meaningless terms is performed by running the pre-processed chatlog datasets as input of the LDA algorithm and observing which twenty words are identified as most relevant for a given number of topics. Meaningless words were removed by continuously extending the stop-words until the LDA algorithm did not identify any meaningless words as the twenty most relevant keywords in each topic. As a result, the coherence of topics became much clearer, and the topics were easier to interpret. The list of words that were used to extend the NLTK's standard stop-words can be found in appendix A.

Once the stop-words are removed from the data, the collection of tokenized documents is passed to a function that identifies commonly occurring bigrams and trigrams, transforming the tokenized documents into a collection of unigrams, bigrams, and trigrams. An N-gram is a contiguous sequence of n items from a given text sample. These n-grams were extracted from the pre-processed data so that two or three words are grouped from single words into a more meaningful collection of words. Examples of bigrams that occur in the collection of words after generating n-grams are *team_vacation*, *account_setting* and *crypto_chain*. We note that stemming and lemmatization are not always beneficial for the ability to understand topics, as mentioned by Waal et al. (2008). However, in this thesis, we observed that the use of lemmatization improves the quality of topic generation and still allows the topics to be understandable. When stemming was applied to the collection of words, the topics became harder to interpret, and therefore stemming is left out of the methodology of this thesis.

3.1.5 Setting LDA parameters and determining overarching topics

Topic modeling using LDA is a text mining procedure that analyzes keyword frequencies in the pre-processed data inputted to the algorithm (Kigerl, 2018). Therefore, each word that appears in the pre-processed collections of messages becomes a variable, representing the frequency count of the number of times the word is used in a document. This methodology is often considered a bag-of-words method, in which textual data sources are converted in a term frequency matrix, where a row represents a document, and a column represents the frequency with which each word is used in a document, with one column per word (Kigerl, 2018).

LDA performs soft clustering, meaning the identified clusters are not mutually exclusive. This implies that the different messages or threads can be assigned to more than one category. The assignment of categories by LDA is stochastic in which each document is assigned to each of the k topic clusters so that the sum of probabilities for each document is equal to 1.0 (Kigerl, 2018). LDA is titled latent Dirichlet Allocation because it estimates latent constructs (topics) while assuming that the category probabilities follow a Dirichlet distribution (Kigerl, 2018). A Dirichlet distribution is a distribution over a distribution which in this case represents documents distributed over topics with topics distributed over words. The model initially assigns the different documents from the corpus to a specified number of topics k. This initial step is based on a random guess. Subsequently, the probabilities for each word belonging to a given topic are calculated based on the frequency of a word appearing in one topic and the frequency of a word not appearing in other competing topics. The LDA algorithm then uses these probabilities to update the probability of documents belonging to topics in every following iteration so that each iteration is based on the previous probabilities, learning from previous allocations. This process is called the Bayesian inference (Kigerl, 2018). The analysis of the chatlog data in this thesis using the LDA algorithm was performed in Python using the Gensim library (Rehurek and Sojka, 2011).

The LDA algorithm uses three input parameters to generate topics from the input data: the number of topics k, the distribution of words over topics α, and the distribution of topics over words β. The LDA algorithm, therefore, does not determine the optimal number of topics in which the topics have the highest coherence. To explore the optimal number of topics as input parameter, different LDA models were run for a value for k from 5 to 40 with an increment of 5 using two measures of fit. For each value of k, the different measures of model fit are calculated, which can be used to select the number of topics or an interval for which the topic coherence is best. Based on the selected number of topics or interval for k, a similar model fit testing is performed while using a smaller range of values for k and an increment of 2 to get better estimations of the optimal number of topics. Finally, these approximations are combined with pyLDAvis to determine the number of k in which we can best interpret the topics. We found that this is generally the case when there is the least overlap between topics and topics have the most meaningful set of keywords.

Two coherence measures are used for determining the optimal value of k based on the work of Röder et al. (2015) and the available measures of fit within the Gensim library. The two metrics U_{mass} and C_v are used since Röder et al. (2015) argue that these are the fastest and most accurate metrics. While U_{mass} is a minimization metric, meaning that a lower score gives better coherent topics, C_v is a maximization metric in which more coherent topics are indicated by a higher score. Both measures of model fit were calculated using the built-in functions of the Coherence model and LDA model of Gensim. The values for α and β were set to their default value of 1 divided by the number of topics. We note that we did not try to optimize the performance of the LDA model by finding optimal values for α and β as Porter (2018) used a similar approach in a study for identifying topics from dark web markets' subreddit posts. Furthermore, the aim of this study is not to build the best performing LDA model but rather to use LDA topic modeling to empirically determine topics in Conti's communication data that are easily interpretable. We believe that the methodology used supports the aim of this thesis and serves the goal of finding the best interpretable topics from the internal communication data of Conti.

A common issue regarding LDA topic modeling is that the generated topics are difficult to interpret or not coherent to humans (Chang et al., 2009). The relevancy metric introduced by Sievert and Shirley (2014) is used to increase the certainty of being able to interpret the generated topics. Topics output a ranked list of the most probable terms, but this can be problematic since the most common terms generally have a high ranking. This makes it hard to distinguish the differences in topics (Porter, 2018). The relevancy metric introduced by Sievert and Shirley (2014) has the following equation:

$$rel(term\ w \mid topic\ t) = \lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w) \tag{3.1}$$

Inspired by the work of Porter (2018) we can adjust the weight λ after the generation of topics to influence the term ranking associated with the topic. When $\lambda = 1$, the standard ranking is returned. In addition, as λ approaches 0, the ratio of the word-topic probability to the overall word probability increases, meaning that words with high probability $p(w)$ are ranked lower. The analysis of topics was assisted by using the pyLDAvis Python library that allows interactive topic model visualization and is based on the work of Sievert and Shirley (2014). This tool helps to determine the overlap between topics based on keywords and identifies the twenty most relevant terms for each topic. Furthermore, using this tool, we can modify the most relevant term distributions per topic. To obtain a clearer understanding of the generated topics, different settings for λ were used. We examined the term distribution when setting λ to 1, 0.8, 0.5 and 0.2, similarly to Porter (2018). The different rankings were recorded if the new terms provided valuable information or helped to clarify the topic. However, if a value for λ lower than 1 gave a more meaningful collection of terms than the original ranking, we replaced the setting instead of representing both settings. Furthermore, similarly to the methodology of Porter (2018) some of the repetitive terms such as "server" are omitted in the listings to show more valuable terms in other topics.

Each topic is assigned a label that relates to the semantic meaning of the topic. The topic labels are determined based on the most relevant terms associated with the topic, while these relevant terms are observed in their original context. Labeling these topics contributes to the primary goal of empirically determining overarching topics in Conti's leaked internal communication and using these empirical findings to map Conti's TTP. In other words, the scientific contribution in this thesis is focused on the empirical determination of the topics in Conti's internal communication. While observing the relevant terms in their original context, messages allocated to the topic are analyzed to put these labeled topics in the context of Conti's ransomware operations. This is done by observing the messages in which the most relevant terms occur, and therefore is based on the empirical methodology. This contributes to the goal of empirically determining topics in Conti's communication and mapping these to their TTP in two ways. First, by observing the messages related to topics, we can validate their labeling and come to more accurately labeled topics. Second, the discussion of the context of these topics helps to interpret their labels better and to form a better basis for mapping these empirically found topics to Conti's TTP.

Question marks are used to show that we are uncertain about the topic label, in which a single question mark indicates we are uncertain that this is truly the topic, and three question marks by themselves indicate that we are uncertain what the topic is in general. To increase the validity of the topic labels, the topic labels are validated during an expert session with members of the cybercrime team of the Fiscal Information and Investigation Service (FIOD). These members have experience investigating cybercrime and ransomware organizations from a financial perspective and therefore helped to give better labels to the topics. A summary of the validation sessions is presented in Appendix B.

3.2 RECONSTRUCTING THE RANSOMWARE ECOSYSTEM USING CONTI'S TTP

Section 3.1 laid out the methodology for empirically determining overarching topics in Conti's internal communication. These topics are mapped to Conti's tactics, techniques, and procedures to describe how the Conti ransomware gang internally operates. While mapping these topics to Conti's TTP, the discussions of messages allocated to topics are used to put the mapping of the topics to TTP in the context of Conti's ransomware operations. That is, the empirical findings

are positioned centrally in the mapping of Conti's TTP to leverage the use of ground truth communication data and contribute to the identified knowledge gaps. In contrast, the allocated messages are used to discuss the context of these findings. Focusing on these empirical findings, this research differentiates from research by cybersecurity companies and -journalists in which messages are read, and the findings have a bias toward the role of the researcher, as previously explained.

A focus is laid on topics that regard Conti's TTP, meaning that all of these topics are used in the mapping of Conti's TTP. Since Conti likely uses multiple tactics, techniques, and procedures, we focus on discussing the most relevant TTP. That is, we prioritize discussing novel TTP with respect to what has already been observed in scientific literature and TTP that helps to determine how Conti establishes interactions with actors in the ransomware ecosystem. In addition, we focus on finding the representative and characteristic TTP for the Conti ransomware gang and put this in contrast to the work of previous scholars as reviewed in chapter 2.

Next, Conti's TTP are leveraged to reconstruct the ransomware ecosystem in which we used ground truth communication data to empirically come to these TTP. Based on these TTP it is discussed how Conti establishes interactions with actors within the ransomware ecosystem. Since these are derived from their internal communication data, the established interactions with actors in the ransomware ecosystem are discussed from an internal perspective of the Conti ransomware gang. First, we discuss the establishment of interactions for actors in each of the three sub-systems of the ransomware ecosystem. Finally, we synthesize these findings for each of the three sub-systems with the actors and their interactions as discussed in chapter 2 to come to a reconstruction of the ransomware ecosystem. In this synthesis, we discuss how ransomware gangs establish interactions with other actors in the ransomware ecosystem and bring this in contrast with previous scientific work as reviewed in chapter 2.

3.3 CONCLUSION

This chapter presented the methodology used for answering the research questions as formulated in section 1.3. First, a methodology is presented for a high-level exploration of Conti's leaked communication data. Next, the methodology for applying LDA topic modeling to these communication data is presented. In this methodology, the communication data is first being split into three datasets to allow for clustering messages into larger texts to improve the performance of LDA. Subsequently, each dataset is pre-processed, and the optimal number of topics for each LDA model is determined using two measures of model fit and pyLDAvis. Third, the topics that come forward from the LDA models are labeled and discussed while keeping the empirically identified topics central in this research and using allocated messages to bring these topics into their context. Finally, these topics are mapped to Conti's TTP, and based on these TTP the ransomware ecosystem is reconstructed. Taking this approach, this research differentiates from the research of journalists and cybersecurity companies on Conti's leaked communication data while contributing to the knowledge gaps as identified in section 1.2.

4 HIGH-LEVEL EXPLORATION OF CONTI'S INTERNAL COMMUNICATION DATA

This chapter presents a high-level exploration of the Conti chatlog data used in this thesis. This chapter aims to come to a better understanding of the organization and operation of the Conti ransomware gang. In addition, this chapter aims to come to an understanding of how the two chat services are used differently. These understandings help to interpret the topics and their context, which results in a higher quality of the empirically determined topics. Section 4.1 discusses Conti's organization and operation by researching Conti's activity over time. Section 4.2 further researches Conti's organization and discusses the differences in use-cases for the two chat services by researching the most- and least active Conti members in both chat services and by visualizing the chat services as networks. Finally, section 4.3 presents a brief conclusion of the findings in this chapter.

4.1 CONTI'S CHAT ACTIVITY OVER TIME

To gain a better understanding of Conti's organization and operation and to understand the differences between the two chat services, a high-level exploration is performed. The chatlog data used in this thesis covers Conti's communication in these two chat services from June 2020 to March 2022. Furthermore, the data used consists of three separate files, of which one contains the messages sent on Conti's Rocket.Chat server and the other two respectively contain the messages sent on the Jabber server in 2020, and in 2021 and 2022. The number of messages, the number of encrypted messages, and the number of senders and receivers are computed for each file. An overview of these variables is presented in table 4.1. The senders and receivers computed represent users that have sent or received one or more messages in that chatlog file. It is therefore likely that there is at least overlap between the two Jabber files and that there may be overlap in senders and receivers between the Rocket and Jabber files. The number of encrypted messages relates to the messages that are translated to English but were encrypted when the user received the message and, therefore, the receiver was not able to read the message. This may indicate the existence of internal problems with using the chat service. Furthermore, there are messages present in the chatlogs that were not decrypted and therefore were also not translated, but these are not considered in the computing of the descriptive variables. The main reason for not observing these is it being rather difficult to observe those because of their format and since passwords and hashes that are sent are similarly formatted.

Table 4.1: Overview of descriptive variables per chatlog file

	Jabber 2020	**Jabber 2021-2022**	**Rocket**	**All chatlog files**
Senders	205	274	248	592
Receivers	302	341	99	562
No. messages	107.967	60.773	88.116	256.856
No. encrypted messages	1	15.470	0	15.471

From table 4.1 can be concluded that most messages are sent using the Jabber server. However, the activity on the Jabber server significantly decreases from 2020 to 2021 and 2022. This could have multiple reasons, but since about 25% of the messages sent in the Jabber chatlogs between 2021 and 2022 are received encrypted, it seems logical to assume that Conti switched

to a different Jabber server for communication because of these encryption problems. In addition, from table 4.1 it becomes evident that each Chatlog file does not have the same number of senders as receivers, which may indicate that Conti is likely using these chat services to establish interactions with actors outside of their organization. This confirms the initial thought of Conti rather than establishing interactions from an interconnected ecosystem of criminals than operating solely by themselves. Finally, the number of receivers observed in the Rocket chatlog file is significantly less than the receivers, which indicates that it is likely to be used for group communication, in which the receivers may be channels. To further explore the operation and activity in the chatlogs, the total number of messages over different units of time is analyzed.

Figure 4.1 presents the total number of messages sent per month for both chat services. From this representation can be observed that there is a decrease in activity for both chat services from 2020 to 2021. As previously argued, this can likely be explained by Conti switching from using a few chat service providers to using multiple chat service providers to increase their resilience in case of problems with a chat service. This also implies that the chatlog data used in this thesis does not cover all communication within Conti. In addition, from figure 4.1 can be observed that there is a decrease in activity from August 2021 to September 2021, which may be related to the internal leaking of the Conti playbook (Cisco Talos, 2021).

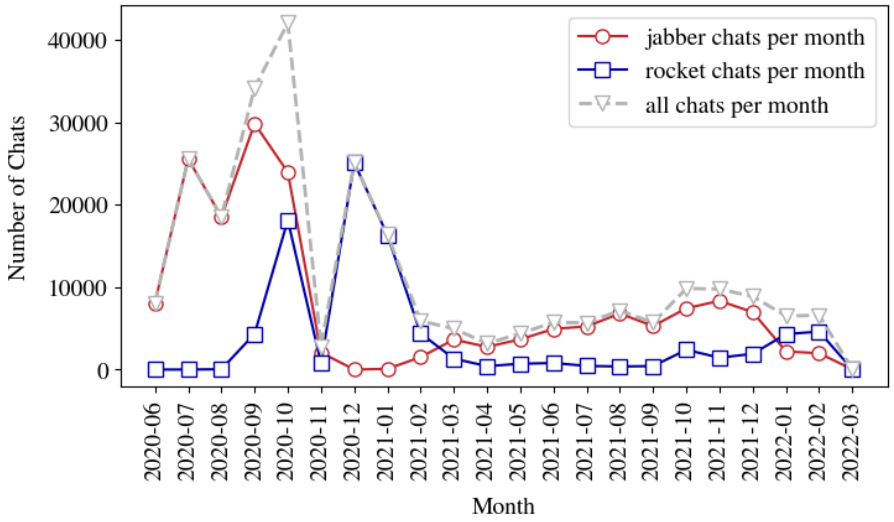

Figure 4.1: Total number of messages sent per month for each chat service

Figure 4.2 presents the number of messages sent per weekday for each chat service. When observing figure 4.2, it becomes evident that Conti is barely active on the weekend and mainly works during weekdays. Furthermore, when looking at the messages send per hour in figure 4.3, we can conclude that Conti generally follows an ordinary working routine since most activity happens between 08:00 and 21:00. However, figure 4.3 illustrates that between 21:00 and 03:00, there is some activity which may indicate that Conti operates from different time zones. Since the activity between 21:00 and 03:00 in the Rocket.Chat server is far greater than the Jabber server it is likely that different groups of users are active on the Rocket.Chat server than on the Jabber server. For example, the Rocket.Chat server could be used more by Conti's affiliates or by separate teams such as the "customer service" team that focuses on the ransom negotiations with victims and helps them decrypt their files.

4.2 COMPARING USE CASES OF THE JABBER AND ROCKET SERVER

As discussed in section 4.1, the differences in activity and descriptive variables indicate that the Jabber Server and the Rocket Server may have different use cases. To further explore the differences between the two chat services and to gain more insights into Conti's operation and

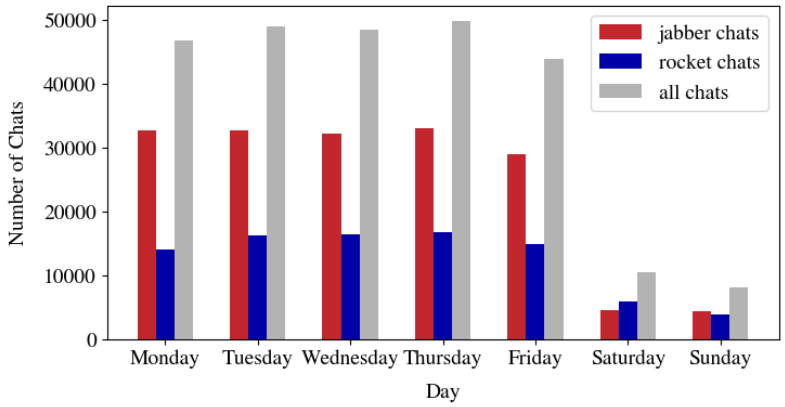

Figure 4.2: Total number of messages sent per weekday for each chat service

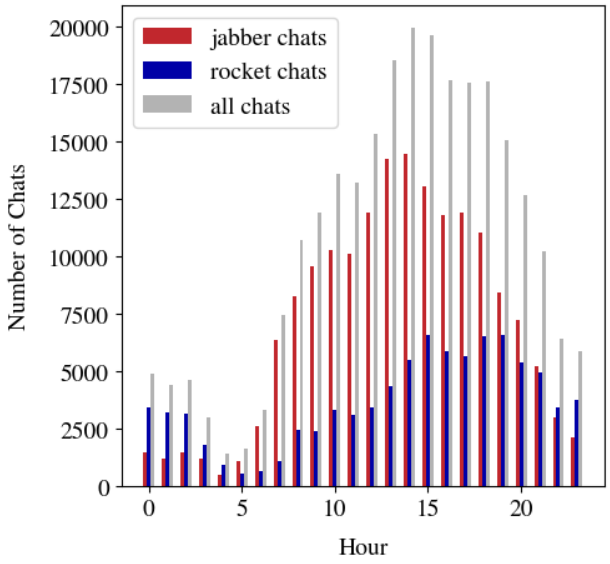

Figure 4.3: Total number of messages sent per hour for each chat service

organization, the users that are most frequently sending and receiving messages are determined for both chat services as presented in tables 4.2, 4.3, 4.4, and 4.5. From table 4.2 can be observed that each of the most frequent senders and receivers in the Jabber chatlogs use emails with aliases and .onion extension, meaning that they use an encrypted mail server that is run over the Tor network and is, therefore, more secure. Furthermore, it becomes evident that the activity in the sense of messages sent and received is greatest for *target*, *bentley*, *stern* and *defender*. This indicates that these users are likely to be important managers in Conti's organization and give commands to multiple teams and team leads since managing multiple teams are likely to need more communication. Accordingly, it seems logical that the usernames that follow in the ranking after defender are team leads since their role relies on communicating with other Conti members to manage their team within Conti's organization correctly.

Table 4.3 shows that multiple users exist that have only sent a single message in the Jabber chatlogs. All of these users also received at least one message, which may indicate that these messages are, for example, used for testing or setting up a connection to Conti's Jabber server. Furthermore, among those users, four usernames do not contain a *.onion* extension in their email. Admin@expiro-team.biz refers to the malware family called Expiro, which infiltrates executable files on 32- and 64-bit Windows OS versions (Lin, 2017). Expiro can be used to install malicious browser extensions, lower browser security settings, and steal credentials. This implies that Conti is likely to collaborate with other larger cybercriminal organizations such

Table 4.2: Users that are most frequently sending and receiving messages in Jabber chatlogs

Rank	Username	Sent	Received
1	target@q3mcco35auwcstmt.onion	26.770	9.878
2	bentley@q3mcco35auwcstmt.onion	17.441	19.024
3	stern@q3mcco35auwcstmt.onion	11.947	16.634
4	defender@q3mcco35auwcstmt.onion	9.667	10.712
5	hof@q3mcco35auwcstmt.onion	5.041	6.123
6	mango@q3mcco35auwcstmt.onion	4.118	3.439
7	driver@q3mcco35auwcstmt.onion	4.038	3.854
8	deploy@q3mcco35auwcstmt.onion	3.780	5.175
9	mushroom@q3mcco35auwcstmt.onion	3.690	3.472
10	bio@q3mcco35auwcstmt.onion	3.196	2.171
11	professor@q3mcco35auwcstmt.onion	2.251	4.314
12	troy@q3mcco35auwcstmt.onion	1.546.	3.641

as Expiro to bundle forces besides collaborating with their affiliates following the RaaS value chain. Exploit.im and chatterboxtown.us are both Jabber chat service providers, which confirms the thought that Conti does not only use one Jabber server but rather uses multiple Jabber servers that may be interconnected for communication. This is further illustrated by messages sent in June 2020 to all of Conti's organizations to register on external Jabber accounts in case a Jabber server is compromised. Furthermore, *def* and *mashroom* seem to be other versions of the aliases *defender* and *mushroom*, which may likely be registered at external Jabber servers or are set up in case something happens with their primary account.

Table 4.3: Users that are least frequently sending and receiving messages in Jabber chatlogs

Rank	Username	Sent	Received
1	admin@expiro-team.biz	1	1
2	pin2@q3mcco35auwcstmt.onion	1	1
3	max17@q3mcco35auwcstmt.onion	1	1
4	dantis@q3mcco35auwcstmt.onion	1	1
5	larry@q3mcco35auwcstmt.onion	1	1
6	def@q3mcco35auwcstmt.onion	1	1
7	mashroom@q3mcco35auwcstmt.onion	1	1
8	billgeizh@q3mcco35auwcstmt.onion	1	1
9	odw5mdwotufuxxrgw3[. . .].onion	1	1
10	beny@q3mcco35auwcstmt.onion	1	1
11	rozetka	1	1
12	exploit.im	1	1
13	mavelak@q3mcco35auwcstmt.onion	1	1
14	verchunls@chatterboxtown.us	1	1
15	good_place@conference.q3mcco35auwcstmt.onion	1	3
16	redroom@q3mcco35auwcstmt.onion	1	2

Similar to the Jabber chatlog data, the users that are most frequently sending and receiving messages in the Rocket chatlog data are presented. Table 4.4 presents the users that most frequently send and receive messages, while table 4.5 presents the users that have sent and received the least messages in the chatlog data. From these frequencies can be observed that a sender or receiver only sends or receives messages, confirming the notion of the Rocket server only being used for communication through different channels. Furthermore, it can be observed that the general channel significantly received the most messages, and it is, therefore, useful to further explore how the communication in the general channel differs from the other sub-channels. Moreover, since topics in the general channel may vary widely and the sub-channels may have more focus, it is worthwhile to analyze the general channel and the sub-channels separately when determining topics using LDA.

In addition, it is notable that the aliases used in the usernames and their format do not match the aliases and format used in the Jabber chat. Therefore, it seems logical to assume that generally, the Rocket server is used by different users and for different purposes. However, the user rozetka is also occurring as one of the users from the Jabber chat services, although only a single message is sent and received by this user. This indicates that some users may be present in both chat services, forming a link between the users active on the Jabber server and those active on the Rocket server. For instance, this may relate to helping with questions of affiliates, communicating with a team of developers, or negotiating with victim organizations. However, it may also be the case that different aliases are used in different chat services and that, therefore, the Rocket server and Jabber server are less differently used than the high-level exploration indicates. Arguably, this is not very likely since Conti is a rather large organization, as can be observed from the 592 usernames active in the chat services. Furthermore, as previously explained, cybercriminals devote themselves to working anonymously and keeping their identities away from their illegal activities, meaning that Conti members may not know the real identities of other members. Using multiple aliases in such a large organization is expected to lead to confusion and inefficiencies, and it is therefore not likely that this is applied on a large scale.

Table 4.4: Most messages sent and received per username/channel in Rocket chatlogs

	Most message received		***Most messages send***	
Rank	**Channel**	**Received**	**Username**	**send**
1	GENERAL	39.966	tl1	28.162
2	pcAjgzgZ5CvxFqGTv	4.378	user8	11.892
3	CD3unmS5YbWcpczbh	3.042	tl2	5.761
4	cMs2nDpvjqoP42TMf	3.034	user9	5.648
5	pQT2ur5KsovPfq7dN	2.896	user4	4.968
6	bcfjvf652di6wjZHA	2.390	user7	4.580
7	v3tBoYNZMCHwesdqJ	2.188	user3	3.522
8	fpRNTcoCaeBefKPD4	2.134	angelo	2.909
9	LnpEcH4KA3qcTk2Pc	1.846	kermit	1.396
10	Ny9GRiwt6QBXPgF5u	1.826	homer	1.307
11	mYvb3eKbqQhMmfxD7	1.712	rozetka	1.173

Table 4.5: Least messages sent and receiver per username/channel in Rocket chatlogs

	Most messages received		***Most messages send***	
Rank	**Channel**	**Received**	**Username**	**Send**
1	jPx6TKsX4jD6YiKnToepsydPpqCisSxhcr	1	hewsi	1
2	z9vn8MvcY5bMazYNN	2	secret	1
3	Rmne8eAkiu37dhm5zaLgWcQx7CGaqXfqkN	2	Lincoln	1
4	oepsydPpqCisSxhcrwLxG2ENqMEjmgKccu	4	greco	1
5	MXpJAXxAMdre5zHBe	5	freter	1
6	5AvY88GigdgbSKr9AoepsydPpqCisSxhcr	5	stomp	1
7	MtKNTY3DtoTqRm2E3Rmne8eAkiu37dhm5z	5	shulman	1
8	7JTqsEQgD5iRCZkZzjT6adLNqjY4RoZWbc	6	black	1
9	GGx6RdTts8Fvu8begoepsydPpqCisSxhcr	8	ithan	1
10	5JgpXux9tPavnZ97YDfYf7ePYsHgW9zD5Z	12	scott	1
11	mYvb3eKbqQhMmfxD7	1.712	rozetka	1.173

To further illustrate the differences between the two chat services and gain more insights into Conti's organization and operation, we visualized the conversation among members as a graph in which a directed edge between node a and node b indicates that node a has sent at least one message to node b, with nodes a and b being representations of users or channels.

The communication of the 20 most active members within the Jabber chatlogs is visualized in figure 4.4 whereas the communication of the 100 most active users among the 15 most used

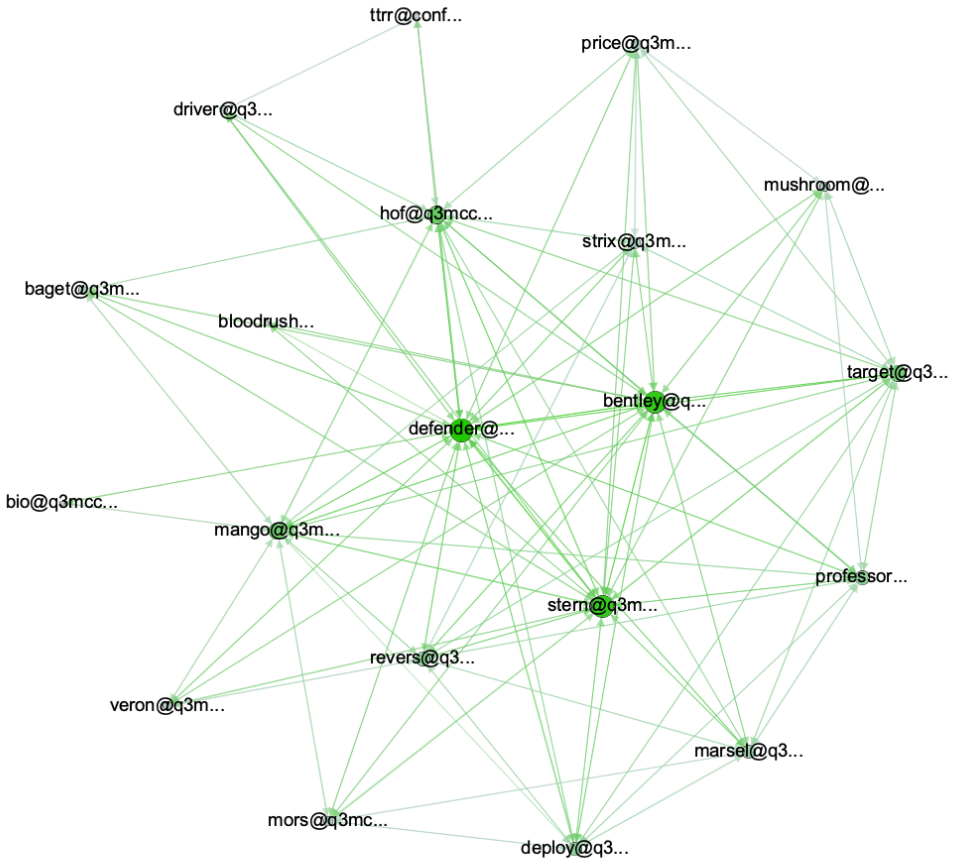

Figure 4.4: Network of the 20 most active users in the Jabber chatlogs

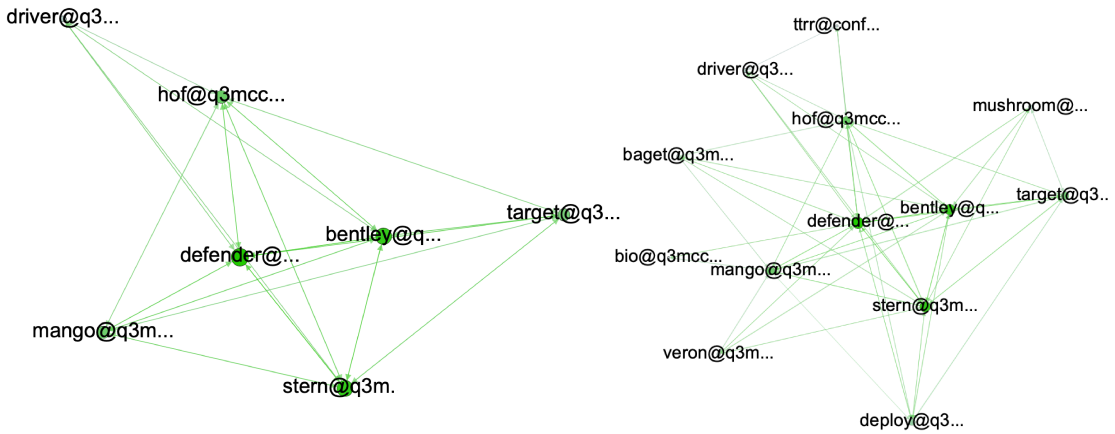

Figure 4.5: Deconstruction of the Jabber communication network

channels within the Rocket server is visualized in figure 4.6. Larger versions of these networks can be found in appendix C. From figure 4.4 can be concluded that all communication among the 20 most active users is bilateral. Furthermore, from figure 4.4 can be observed that defender, bentley, stern, target, and mango are most likely the top-level managers of Conti. The reason for this is that these members are all communicating with each other and have the most communicational relations in the communication network, indicating that they are managing and overseeing the work of many Conti members. Furthermore, when the communication is deconstructed based on the number of communicational relations as shown in figure 4.5, it becomes evident that users such as hof, deploy, and mushroom are most likely team leads or senior team members since these are all connected with the top-level managers. Users such as bio and baget are only connected with parts of the top-level management, meaning that they are likely to be lower in the hierarchy of Conti and are likely to communicate more with their team members. Moreover, this indicates that the top-level managers have different roles in overseeing the work of different teams, in which Stern is often seen as the "big boss" of Conti (Checkpoint, 2022; Krebs, 2022b).

If the Jabber communication network is compared to the Rocket communication network, it becomes evident that the Jabber server is solely used for peer-to-peer communication and that the communication of Rocket reforms around channels. Furthermore, there is no sign of users communicating directly with other users on the Rocket chat server. In addition, it can be observed that generally, three communication clusters are formed of which one forms around the general channel, one forms around the channel that is ranked eighth in the most message received, and one forms around the remaining 13 sub-channels. When observing the communication network with the 15 most used channels, these sub-communities do not share any channels other than the general channel. This indicates that these sub-communities work in parallel and are, for example, separated by case or team.

It can be observed that for the third cluster, which is indicated in green in figure 4.6, the usernames have different formats from the other two clusters. As can be observed from table 4.4, these are the users that are most frequently sending and receiving messages on the Rocket server. This cluster may be a collection of channels that are used for Conti's team leads since team lead could be abbreviated to tl as in usernames tl1 and *tl2*. Furthermore, similarly to the users in the Jabber network, the users in the third cluster are interconnected with more channels. The existence of these three sub-communities and indicate there being differences in use-cases between the Rocket server and Jabber server. However, the Rocket communication network, as presented in figure 4.6 only shows the communication in the 15 most used channels. Therefore, based on the high-level exploration, the exact reason for the existence of these sub-communities cannot be determined. The high-level exploration provides background on the use-cases of both chat services, creating a better understanding of Conti's organization and operations. To gain further insights into Conti's internal operations and the establishment of interactions with actors in the ransomware ecosystem, we turn to empirically determining topics in their internal communication data in the following chapter.

4.3 CONCLUSION

This chapter aims to come to a better understanding of the organization and operation of the Conti ransomware gang. In addition, it aims to gain insights into how the two chat services are used differently within Conti. From the high-level exploration, it became evident that Conti may have switched from using a few chat service providers to using multiple chat service providers to increase its resilience. This implies that the leaked communication data researched in this thesis may not cover all communication within Conti. While Conti generally follows ordinary working days, the differences in daily activity between the Jabber server and Rocket server indicate that it is likely that different users are active in each chat service. From the most- and least active users in the Jabber server can be concluded that Conti is collaborating

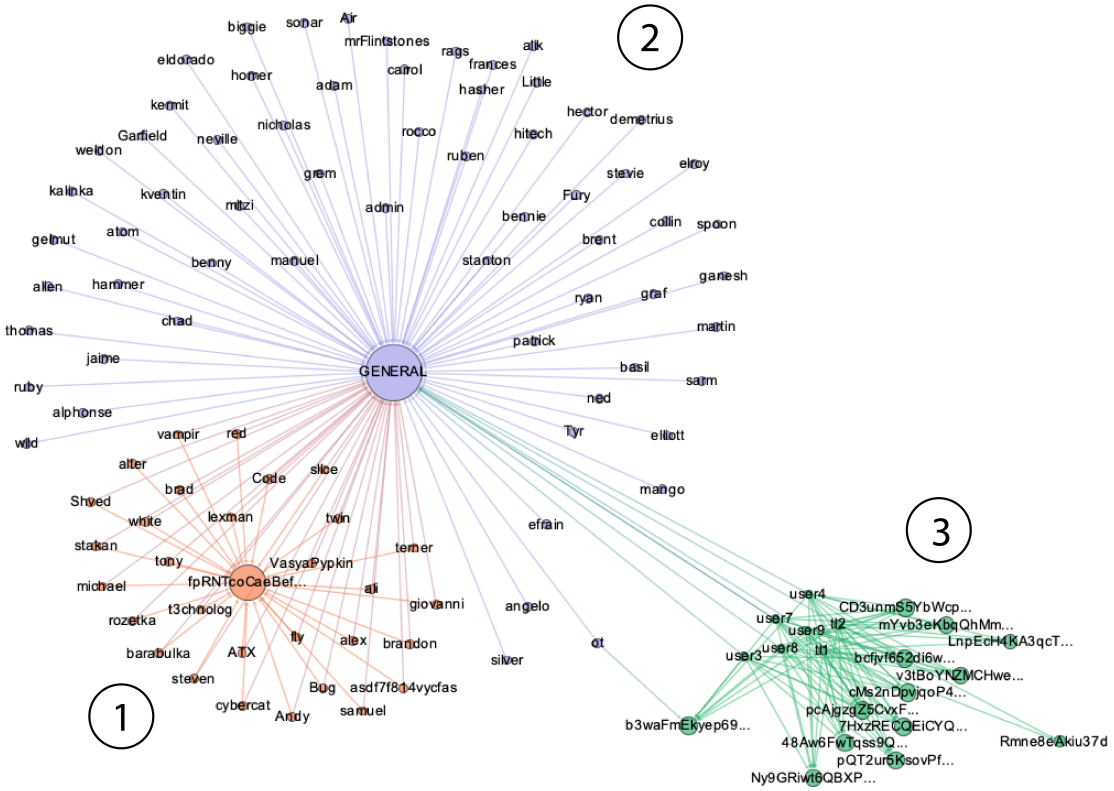

Figure 4.6: Communication network for the 100 most active users in the Rocket chatlogs

with other cybercriminal organizations such as Expiro to strengthen their ransomware attacks. Furthermore, it becomes evident that Conti has multiple top-level managers and team leads in their hierarchy to manage their operations and successfully perform ransomware attacks.

It becomes evident that the Jabber server is used for peer-to-peer communication and the Rocket server for group communication through channels. Furthermore, there is little overlap in the users active in the two chat services, which imply that different parts of Conti's organization use these chat services. The Rocket server has a general channel and multiple sub-channels, and it is worthwhile separating the general channel from the sub-channels when determining topics. Finally, in the Rocket server, three sub-communities can be identified, although their function cannot be determined from the high-level exploration. Further research in the following chapters could determine the function of these sub-communities.

5 OVERARCHING TOPICS IN CONTI'S INTERNAL COMMUNICATION

This chapter presents the results of the empirical determination of topics in the internal communication of the Conti ransomware gang by applying LDA. Section 5.1 discusses the parameter settings for the optimal number of topics k for each of the three LDA models in which topics are most coherent. Section 5.2 discusses the topics coming forward from each of the three LDA models by first discussing the empirical findings from applying LDA topic modeling on the chatlog data and subsequently discussing the context of these topics by discussing allocated messages to these topics. Section 5.3 concludes this chapter by providing a brief conclusion of the main findings in this chapter.

5.1 LDA MODEL PARAMETER SETTINGS

As argued in section 3.1 and 4.2, the chatlog data is split into three text corpora so that these can be analyzed separately. Therefore, three LDA models are created in which each may have different parameter settings. We refer to these models as the Jabber LDA model, the Rocket general LDA model, and the Rocket sub-channel model. The LDA algorithm uses a specified number of topics k to cluster the documents in the chatlog corpora into topics that need to be determined before running the model. The model fit testing is performed using seven different model specifications, starting with an initial five topics and monotonically increasing to 40 topics with an increment of five. The results of the fit metrics for the three different models are presented in figure 5.1 and figure 5.2. Both a minimization (U_{mass}) and a maximization (C_v) metric are used to determine the model fit. Accordingly, a lower score for U_{mass} implies a better fit, while a higher score for C_v indicates a better fit (Röder et al., 2015).

From figure 5.1 and figure 5.2 can be observed that the topics coming forward from the Jabber LDA model are most coherent when using approximately 10 topics since this provides the lowest U_{mass} score and the highest C_v score. In addition, the topic coherence is optimal for the Rocket general model when choosing k between 10 and 15 topics. However, from figure 5.1 and 5.2 can be observed that the optimal number of topics for the Rocket sub-channel model cannot easily be determined since the coherence of topics decreases with an increase in the number of topics. This may indicate that the topics revolve around similar subjects, meaning it is more difficult to create coherent clusters. However, since it can be observed from figure 5.1 that the U_{mass} score is at its lowest around 10 topics, the Rocket sub-channel model likely has its best fit when clustering the documents into approximately 10 topics.

To further explore the values of k for which the three LDA models have the most coherent topics and therefore are likely to output the best interpretable topics, another model fit testing is performed while using a smaller range of values for k. This allows for a more specific determination of the optimal value for k. In the model fit test, five different model specifications are used in which we let k alternate between seven and 17 topics, using an increment of two. The results of this model fit test are respectively presented in figure 5.3 and figure 5.4.

From figure 5.3 and figure 5.4 can be observed that the Jabber LDA model clusters the documents in the most coherent topics when using approximately 11 topics. In addition, it can be observed that the Rocket general model has the highest fit when setting k approximately be-

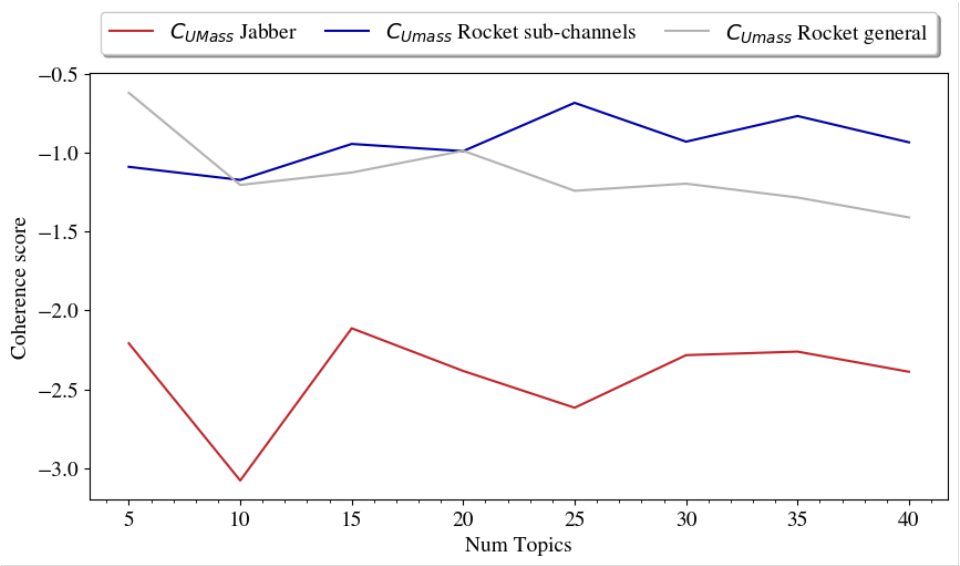

Figure 5.1: Model fit scores using the U_{mass} coherence metric

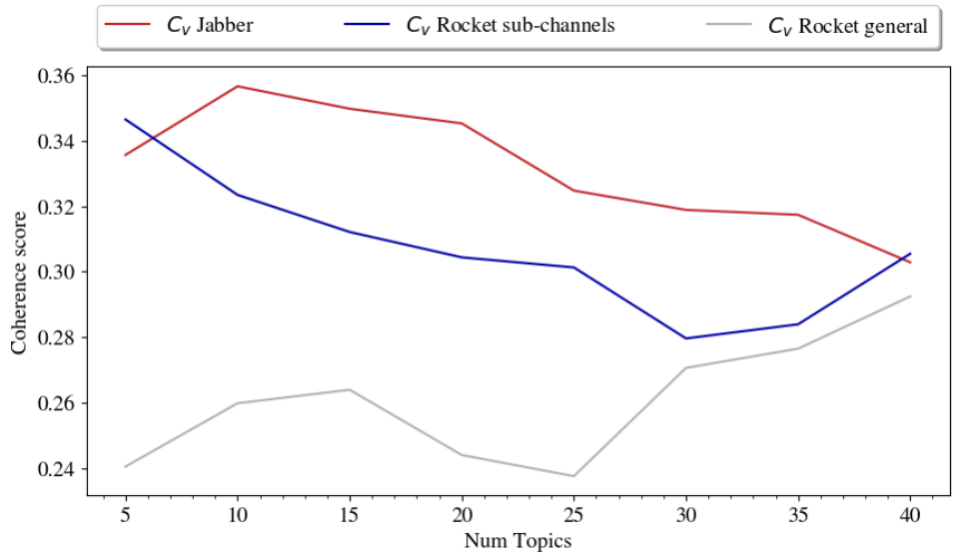

Figure 5.2: Model fit scores using the C_v coherence metric

tween nine and 13 topics which are respectively indicated by its highest Cv score and its lowest Umass score. Although it cannot be easily determined from its Cv scores, the Umass scores for the Rocket sub-channel model show that it is likely to produce the most coherent topics when using approximately 13 topics.

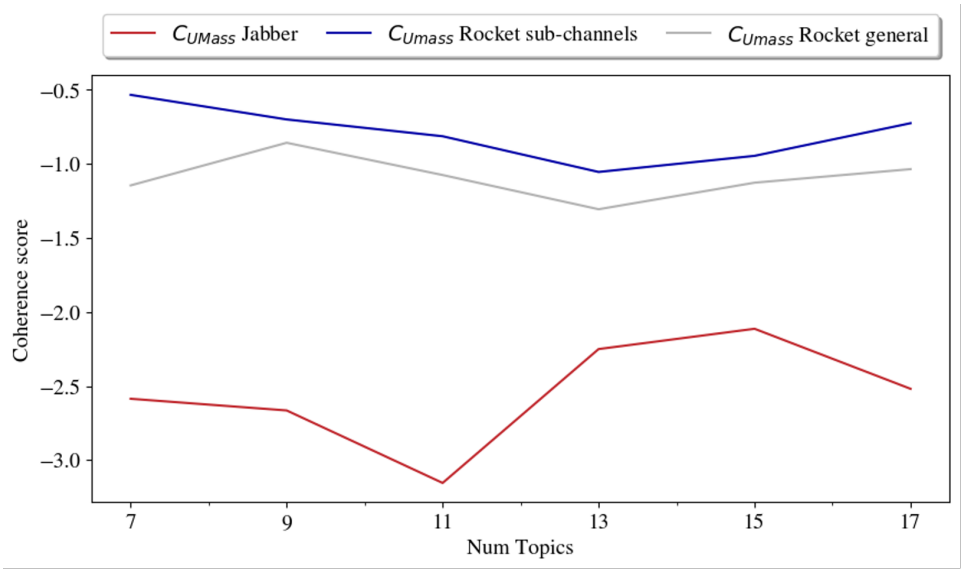

Figure 5.3: Detailed model fit scores using the U_{mass} coherence metric

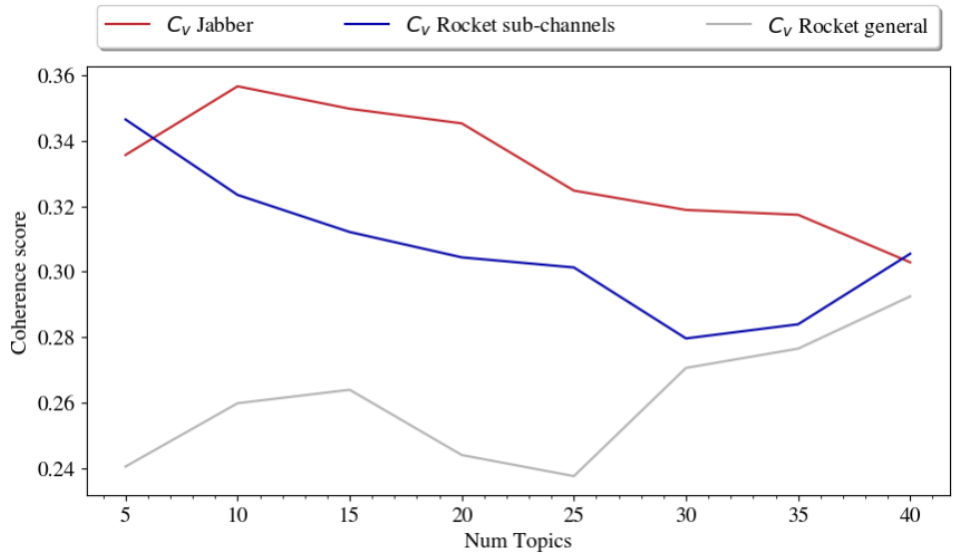

Figure 5.4: Model fit scores using the C_v coherence metric

As argued in the methodology in chapter 3, the primary goal of applying LDA modeling is to empirically determine topics in Conti's internal communication. This relies on topics being best interpretable. Generally, topics are found to be the best interpretable if there is little overlap between topics and if each topic contains the most meaningful set of most relative terms. Inspired by the study of Porter (2018), Intertopic Distance Maps (IDMs) are plotted using pyLDAvis to visualize the overlap between topics and present the most relevant terms for each topic. The approximations of the most optimal values or ranges for k have been used as starting references for observing the overlap and most relevant terms of topics. For the Rocket General model, 13 topics are used as a reference since it is assumed that the general channel is used for rather informal topics. Therefore, more detail is likely to give better interpretable topics. Through iteratively increasing and decreasing the number of topics and observing the overlap between topics and their associated terms, the final number of topics for the three LDA models is de-

termined. This results in the optimal value of k being 12 topics for the Jabber LDA model, 11 topics for the Rocket General LDA model, and 14 topics for the Rocket sub-channel model. In the next section, we discuss the generated topics and their labels for each LDA model.

5.2 TOPICS IN CONTI'S INTERNAL COMMUNICATION

This section discusses the topics as empirically identified from Conti's internal communication using LDA topic modeling. For each LDA model, we listed the 20 most relevant terms for each identified topic using the most relevant settings for λ. The topics are sorted from most prevalent to least prevalent, similarly to the study of Porter (2018), in which the prevalence of a topic defines the percentage of the corpus that the topic is comprised of. Furthermore, similarly to the methodology of Porter (2018) some of the repetitive terms such as "server" are omitted in the listings to show more valuable terms in other topics. Each topic is assigned a label that relates to the semantic meaning of the topic. Question marks are used to show that we are uncertain about the topic label, in which a single question mark indicates we are uncertain that this is truly the topic. Three question marks by themselves indicate that we are uncertain what the topic is in general.

These topic labels are determined based on the most relevant terms associated with the topics in their original context. The labeling of these topics is validated during an expert session with members of the cybercrime team of the Fiscal Information and Investigation Service (FIOD). These members have experience in investigating cybercrime and ransomware organizations from a financial perspective and therefore helped to better interpret the most relevant terms and context and label topics. A summary of the validation session is presented in appendix B. As argued in section 3.1.5 this chapter is focused on the empirical determination of topics so that these can subsequently be mapped to Conti's TTP. While discussing the identified topics, examples from allocated messages are used to illustrate the context of the topic in Conti's ransomware operation. As previously argued, putting the empirical findings in their context results in a higher quality of mapped TTP and, in turn, leads to a more valid reconstruction of the ransomware ecosystem.

5.2.1 Topics in the Jabber chatlog data

Figure 5.5 present the Intertopic Distance Map (IDM) for the topics generated from the Jabber LDA model. This IDM helps to interpret the prevalence and overlap of topics that come forward in the chat logs. From figure 5.5 can be observed that topics 1, 2, and 3 have the highest prevalence among all 12 topics. Furthermore, it becomes evident that topics 1, 2, and 3 overlap, indicating that these resolve around similar topics. Topics 4 to 8 have a medium prevalence in the Jabber chatlog data, and topics 9 to 12 have a low prevalence in the chatlog data. Table 5.1 presents the results of the Jabber chatlogs, being the different topics with their corresponding labels, prevalence ranking, and its 20 most relevant terms.

The 12 topics that come forward from the Jabber LDA model are general conversation on Conti operations, general conversation regarding business unit, reconnaissance, development, payment infrastructure, account takeover, attack vectors, IT helpdesk, PGP messaging setup, infrastructure configuration, recruitment and HR, and onboarding new members. However, the topic onboarding new members is not labeled with full certainty and may therefore have a different meaning. These topics indicate that the Jabber server is predominantly used for communication on Conti's operations since it includes topics that cover the different aspects of Conti's business operations.

Topics 1 and 2 are constructed of general conversation among Conti members as indicated by terms such as "question", "problem" and "issue." As topic 1 focuses more on the general conversations on operations, such as questions, commands, and updates between the different

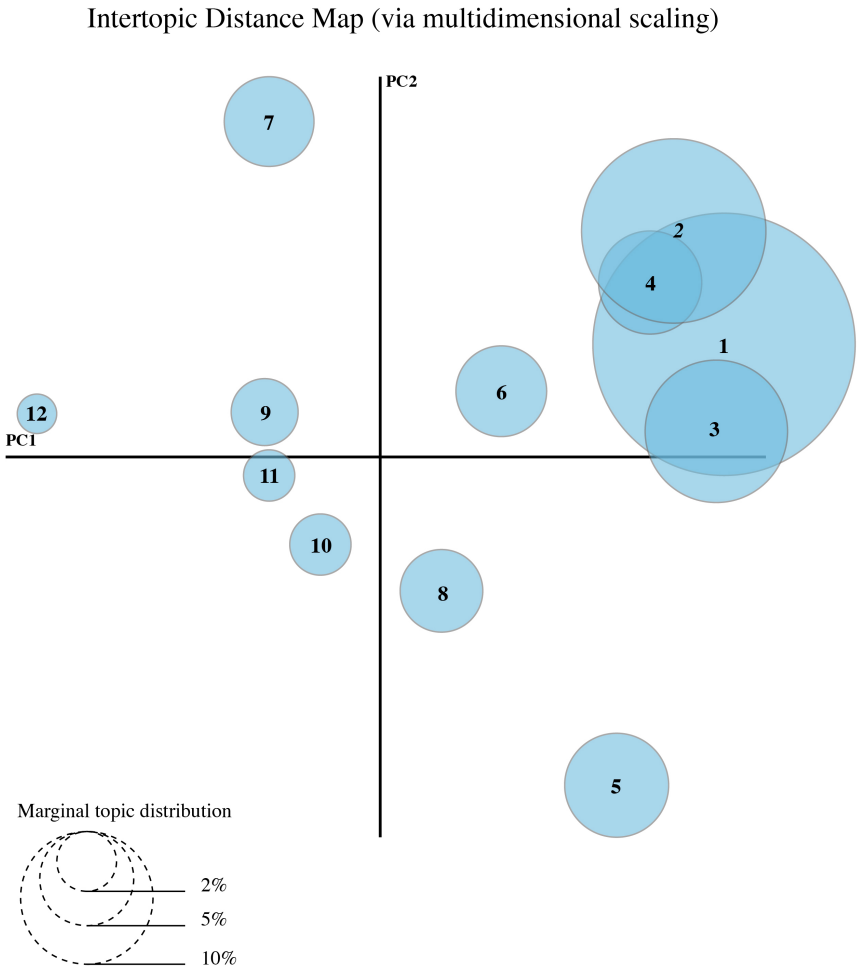

Figure 5.5: Intertopic Distance map Jabber LDA model

managers and team leads, topic 2 is constructed of general conversations with a more focused context. That is, the conversations in topic two are conversations of general context such as questions or commands similar to topic 1 but their subjects reform around different business units such as the development unit. This can be observed from associated terms such as "flood," "autotest," and "update," and the overlap between topics 2 and 4 as visualized in figure 5.5.

Topic 3 is constructed of conversations regarding Conti's reconnaissance activities in which they focus on collecting as much information as possible about their (potential) targets to devise a successful ransomware attack (Dargahi et al., 2019). This is indicated by the terms group, server, bot, log, version, pass, password, and update. In addition, it can be observed from the associated terms that in their reconnaissance, Conti collects information about victims' servers, passwords, and vulnerabilities. From figure 5.5 can be observed that reconnaissance has a high prevalence in the corpus. Therefore it is evident that a large share of Conti's operation is designated to gaining information on victim territories. Furthermore, it can be observed that there is an overlap between topic 1 and topic 3, which may indicate that reconnaissance is also often discussed in general conversations on Conti operation. This may be explained because reconnaissance is highly prevalent in the Jabber chatlog data.

Table 5.1: Topics in the Jabber chatlogs

#	Topic	Relevancy λ and terms
1	General conversations on Conti operations	λ = 1: question, problem, server, pay, contact, message, us, bot, issue, request, project, download, case, system, code, launch, help, number, link, finish λ = 0.5: question, problem, pull, fact, plan, professional, study, request, communicate, function, hand, finish, web, case, pay, number, person, contact, issue, point λ = 0.2: professional, colleague, involve, communicate, implement, implementation, master, position, question, web, pull, tk, language, bypass, regard, function, period, teach, method, fact
2	General conversations regarding business units	λ = 1: launch, knock, burn, download, crypt, fire, clean, update, watch, upload, remove, delete, pass, roll, tap, issue, open, pay, problem, prepare λ = 0.5: tap, crypted, sign, autotest, open, keywords, severity, pancake_private, knock, static, flood, folder, crypt, detective, update, remove, prepare, collect, pass, crypted λ = 0.2: tap, keywords, pancake_private, severity, approx, unwanted, rename, eng, cryptocurrency, dell, skippy, static, autotest, tmp, extraction, severe, refer, hlor, snapshot, fastpath
3	Reconnaissance	λ = 1: delete, pay, group, download, panel, launch, server, bot, accept, log, version, loader, pass, password, dll, clean, update, order, us, knock
4	Development	λ = 1: bot, request, line, value, client, server, field, crash, service, post, gasket, version, code, source, download, cut, parameter, module, push, parse
5	Payment infrastructures	λ = 0.8: project, coin, us, code, dex, system, trend, nft, others, scanner, sonicwall, success, defi, etherium, online, defi_amon, figure, busy, team, blockchain
6	Account Takeover	λ = 1: crypt, dll, pay, bot, stern, exe, people, report, link, clean, issue, mango, point, dock, money, online, contact, download, target
7	(Purchasing) Attack vectors	λ = 1: bugtracker_web, program_period, lock, net, offer, fly, certificate, transfer, steal, hand, exploit, charge, bank, hack, buck, site, link, buy, title, grid, fuss
8	IT Helpdesk	λ = 1: message, contact, server, account, register, ff, icon, important, encrypt, manually, service, cover, case, tab, decrypt, domain, error_acc, encrypt_thrown, connection
9	PGP messaging setup	λ = 1: touch, block, key, message, pgp, page, otr, interview, mapsio, public, encrypt, apro, apr, link, rocket, dump, pas, setting, download, healthy
10	Infrastructure configuration	λ = 1: server, domain, register, specify, message, standard, account, port_important, setting, port, remains_tab, icon_behind_box, us, download, mango, help, upload, pas, tab, delete
11	Recruitment and HR	λ = 0.8: skill, authorization, server, colleague_web, blockchain, short, access, otr, pas_data, case_tdu, text_output, silver_password, purple, developer, bigu, team_extraction, rocketchat, criterion, algorithm_point, password
12	Onboarding new members?	λ = 1: congratulation, face, railway, mapsio_apro, apr, page, net, sps, people, dock, us, repeat, au, gm, macros, opayt, cyber, hors, pay

Topic 4 includes conversations dealing with the actual development of ransomware and its attack vectors. Therefore, most conversations on this topic are of technical nature, and therefore multiple technical terms such as request, line, module, and parse come forward in the 20 most relevant terms. It becomes evident when observing the associated terms that this topic includes conversations referring to the different components that make the Conti attacking infrastructure. This is indicated by the terms "bot," "gasket," and "post," which respectively relate to botnets, backdoors, and servers. The term gasket refers to a backdoor tool called Gasket (Falcone et al., 2021). A backdoor is a class of malware that offers supplementary access to a victim's IT system, often used besides other forms of access such as RDP (Severi et al., 2020). Furthermore, a backdoor can be used to set up a secure and anonymous connection to the ransomware gang's command control (C&C). The terms "post" and "request" refer to the POST request method, which is used to transfer and store data to servers. In practice, this is known to be used for communication of a victim's device with the CC, for example, to generate encryption keys or to extract data (Kharraz et al., 2015; Lemmou and Souidi, 2017).

Topic 5 consists of conversations regarding Conti's payment infrastructures indicated by the terms coin, dex, nft, defi, etherium, defi_amon, and blockchain. Similarly, these terms relate to cryptocurrencies and other blockchain-related payment infrastructure. Dex refers to decentral exchanges, which are peer-to-peer markets in which transactions are held directly between crypto traders. This indicates Conti may use these decentral exchanges for money laundry practices or business transactions (Coinbase, 2022). Amon is a digital wallet that allows cryptocurrencies to be spent in fiat using their card (Amon, 2022). This may indicate Conti using these services to spend their earnings in cryptocurrencies using Amon's card. Furthermore, it can be observed that Conti's payment infrastructure is focused on cryptocurrencies since no terms can be identified that relate to cash or other payment methods. Finally, Conti's payment infrastructure is ranked fifth in prevalence, meaning that it is relatively often discussed in the Jabber server.

Topic 6 includes conversations on account takeover, meaning that it entails the actual encrypting of victims' IT systems, allowing Conti and its affiliates to extort them. "DLL" refers to Dynamic-link library (DLL), which is a collection of small programs that larger programs can load when needed to complete specific tasks (Subedi et al., 2018). Ransomware gangs are known to use a technique called DLL injection to infect systems and subsequently distribute and run the ransomware (Conti et al., 2018; Mekdad et al., 2021; Dargahi et al., 2019). In addition, different executables, indicated by the term exe, are used to take control over the victim's files and system. Botnets and endpoints (referred to as "point") are used to infiltrate systems to take over the account. Botnets are networks of computer devices that are infected with malware and are (often without the user's knowledge) controlled by a so-called bot herder (Wang et al., 2011). These botnets may be used for multiple purposes in Conti's operation, such as reconnaissance, transferring of files and data, and installing and running applications. As can be observed from the most relevant terms, botnets clearly have an important role in the account takeover. Finally, the terms mango and stern refer to members that are involved in Conti's account takeover and, therefore, can be observed in the most relevant terms.

Topic 7 deals with Conti's attack vectors and the purchasing of those. "Lock" refers to the ransomware, which is also known as "locker," which encrypts a victim's files. "Steal" refers to the stealing of credentials which is, for example, done by injecting malicious software into browsers. "Offer," "buck," and "charge" refer to prices that are handled for illicit services and products that are used for attack vectors. For example, it can be observed from the most relevant terms that exploits, crypts, loaders, certificates, hacking services, and network access are likely to be purchased through external actors.

Topic 8 entails discussions on Conti's IT helpdesk for affiliates. Since Conti embedded the RaaS business model as previously explained, they benefit from affiliates being able to perform successful attacks even if they have lesser technical knowledge. Therefore, Conti has a helpdesk for affiliates having questions or issues related to using Conti's ransomware to attack victims. The

terms "message," "contact," and "register" refer to helpdesk conversations on problems with Jabber services and the referral to different contacts for different questions. Furthermore, the term "ff" refers to Fast Flux, which is an evasion technique that is used to increase the resilience of Conti's infrastructure. Fast Flux allows the botnet server, which is inter alia responsible for phishing and malware delivery, to be hidden behind an ever-changing network of IP addresses (Surjanto and Lim, 2020). The term "ff" occurring in the most relevant terms indicates that Fast Flux is one of the subjects often discussed within the helpdesk conversations. Moreover, the terms "case" and "decrypt" refer to attack cases in which victims' files need to be decrypted, indicating that the encryption of files is another example in which affiliates request technical assistance.

Anonymous communication is a crucial part of a ransomware gang's infrastructure since it is essential for its continuity to hide its illegal activities and its member's identities. Pretty Good Privacy (PGP) is an encryption program that provides this privacy for data communication and is often used by criminals (Broadhurst et al., 2018). Topic 9 entails conversations discussing the PGP messaging setup within Conti's organization, indicated by terms such as "touch," "block," "key," "pgp," "setting," and "otr." Off-the-record messaging (OTR) is an alternative to PGP that is assumed to have better functionalities for anonymous communication (Borisov et al., 2004). Since anonymous communication has such a crucial function within a ransomware gang's organization, it may be expected that PGP messaging setup comes forward in the topics. However, topic 9 has a low prevalence score meaning that PGP messaging setup is not often discussed in the Jabber chatlogs.

Conversation allocated to topic 10 entails Conti's infrastructure configuration. This entails configuring and registering domains, specifying internal functions and parameters, and locating files as indicated by the most relevant terms. Conti's infrastructure is one of the topics coming forward since a stable infrastructure is an important factor for the continuance of Conti. If for some reason, Conti's infrastructure fails or is sabotaged, e.g., by law enforcement, Conti is unable to continue its business of extorting companies by encrypting their files. Therefore, it can be argued that Conti's infrastructure is configurated with a focus on resilience.

Topic 11 contains conversations that deal with the recruitment and HR within Conti, which is indicated by the terms skill, colleague_web, developer, and criterion. Other terms such as blockchain and server may relate to the skills needed for vacancies. Furthermore, case_tdu may relate to practice cases in the hiring procedures. Finally, topic 12 contains conversations that we think may deal with the onboarding of new members since "congratulations" relate to new members being added to the Jabber server. Other relevant terms such as "pay," "gm," and "people" may relate to members being onboarded about salaries being paid, the general managers within Conti, and other important people.

Context of topics in the Jabber chatlog data

In section 5.2.1 we presented the empirical findings, being the topics coming forward from the Jabber chatlogs by applying the methodology. To further illustrate the context in which these topics can be observed, we discuss some examples using messages allocated to the topics.

From messages allocated to the reconnaissance topic can be observed that Conti's reconnaissance involves testing different attack vectors in possible target environments such as different versions of Windows, antivirus software, firewalls, browsers, and endpoint protection software. In addition, the allocated messages illustrate that the reconnaissance entails observing potential victims' revenue based on publicly accessible sources such as published quarterly reports and sites such as ZoomInfo. Generally, it can be observed that Conti focuses on the larger companies with revenue starting from tens of millions. Conversations on this topic also deal with different configurations of admin panels and using bots to observe victims' servers and networks. Furthermore, from observing allocated messages, it becomes evident that extensive reconnaissance

has a high prevalence since it supports the strategic organizing and planning of ransomware attacks. Consequently, this increases their chance of success.

When observing conversations allocated to the development topic, it becomes evident that Conti develops and adjusts multiple attack vectors. These developments are not solely performed for Conti's own improvement but often originate from development requests from affiliates. However, it is worth noting that Conti's development is predominantly based on the incorporating and adjusting of external toolkits and services. This is in line with the findings of Van Wegberg et al. (2017) that cybercrime is increasingly utilizing commoditization. From conversations allocated to the payment infrastructure can be observed that Conti discusses the development of its own blockchain project, which allows for more control over its payment infrastructure. That is, by developing their own blockchain or DAG Conti could develop a protocol in which their transactions among their affiliates and defenders are validated by themselves, which provides full anonymity.

Furthermore, from messages regarding the payment infrastructure, it can be observed that plans exist to develop a cybercriminal social network on this blockchain. In addition, Conti plans this blockchain to be a centerpiece in the attacker ecosystem, allowing other criminal gangs to build their own project on Conti's blockchain. Subsequently, this blockchain can then be used to decentrally store and sell compromised data. Other conversations in the payment infrastructures topic discuss crypto coins that are currently being used within Conti. For example, Conti uses the Siacoin blockchain to store data and programs safely. Moreover, Conti uses Emercoin and its EmerDNS protocol to host domain names anonymously and safely, protecting them from law enforcement. Finally, it can be noted from conversations that Conti uses the pump and dump technique to increase the value of its profits. Pumping and dumping is a manipulation scheme in which the price of a cryptocurrency's coin is inflated, after which the assets are sold for a higher price.

Messages revolving around the account takeover validate that botnets are used for all the mentioned purposes. Furthermore, it can be observed that links are often used in combination with a victim's domain name and an executable. For example, when a member asks for links, it can be answered by *www.[victim's company name].com/[exe name].exe*. For these executables, attractive names such as "Preview_document" are uploaded to these domains. These websites look like being owned by licit companies, and therefore, it is likely that these are victims of Conti. Conti may either use these links to infect the website or hijacks these websites to distribute the ransomware from these domains. Due to our limited extent of technical knowledge, we cannot fully interpret the exact use cases.

When observing messages related to the (purchasing) attack vectors topic, it becomes evident that social engineering is utilized in the attack vectors of the Conti ransomware gang. For instance, Conti has realistically looking pages of which an example can be observed at contirecovery.info, which promotes installing a new add-blocker plugin into one's browser. Subsequently, the user is referred to an authentic-looking Google plugin page from which malicious code is installed. Using these injections, Conti secretly observes defenders' browser behavior to observe cookies and passwords being saved in the browser to retrieve further access within a defender's system. This is one of many examples that illustrates how Conti uses social engineering to progress through victims' networks. For example, from observing allocated messages, it can be validated that exploits, crypts, loaders, certificates, hacking services, and network access are commonly purchased externally. Furthermore, examples can be observed in which services are bought from spammers that distribute the ransomware using spear-phishing based on the reconnaissance of defender territories. Most of these external actors are often found through acquaintances of Conti members or through dark web forums, where users present themselves with their services and prices. The prices paid for externally bought attack vectors are often based on the revenue of the target company, as can be observed from examples in which network access is purchased.

From conversations regarding anonymous communication can be observed that both OTR and PGP are used in parallel, and separate use cases for both cannot be clearly distinguished. Within this topic, conversations generally deal with setting up and registering OTR or PGP and helping users with questions and problems with their anonymous messaging protocols. Similarly, messages regarding Conti's infrastructure configuration show examples of how infrastructure components are often used in parallel to increase its resilience, validating the thought of resilience being a focus in the infrastructure configuration.

Since Conti is a relatively large organization with over 350 employees, messages regarding recruitment and HR show how recruitment is professionally embedded within Conti's operation. For example, Conti has a dedicated team of recruitment officers. These recruitment officers discuss with team leads about necessary skills and present resumes from potential candidates that match the vacancies. These resumes are typically retrieved from Russian-speaking headhunting services such as headhunter.ru. However, Conti does not allow traces of Conti's job openings on recruitment websites. Therefore Conti bypasses the recruitment websites and directly access the resume pool, and contact candidates by email. Access to these resume pools is gained through purchasing the software used for maintaining the resume pools (Checkpoint, 2022). When candidates are contacted, they are often lied to, leaving out the actual industry of Conti to protect its anonymity and have more success in hiring candidates.

Skill is central in the search for new candidates, and often job descriptions are made for specific vacancies so that these can be shared by employees with acquaintances. For a customer service function, necessary skills are a good knowledge of spoken English and being between the age of 18 and 25. Subsequently, a salary of 450 to 500 dollars is offered while working remotely five days a week from 18:00 to 02:00 in the Moscow time zone. In contradiction, several developers initially start working for free to gain development experience and hope for a quicker promotion in Conti's hierarchy, which indicates that Conti is also likely to hire interns. Recruitment officers highlight that all Conti employees work remotely and that this is a strict criterion.

From the conversations regarding recruitment and HR can be noted that Conti is specifically looking for admins of dark web forums and markets. One of the reasons that come forward for targeting admins is to incorporate these in the development of Conti's social network project, which contributes to the ability to sell data among other cybercriminals through this platform, using the networks and experience of forum admins. Similarly, recruiters are looking for developers with blockchain experience to develop Conti's blockchain project, as previously illustrated. Finally, recruitment and HR do not only cover recruiting new Conti members externally. For example, it can be observed that an overview of Conti's internal programming skills is created by asking each developer to report their skills.

5.2.2 Topics in the general Rocket channel

This section discusses the topics coming forward from the Rocket general LDA model. Figure 5.6 presents the Intertopic Distance Map (IDM) of the topics generated from the Rocket general LDA model, which shows the prevalence and overlap. From figure 5.6 can be observed that topics 5 and 7 overlap and that the prevalence of topics is more evenly distributed compared to the IDM of the Jabber chat logs as presented in section 5.2.1, indicating that topics are more evenly discussed. However, the prevalence of topics 12, 13, and 14 is lower than the other topics, indicating that these makeup less of the conversations in the general channel of the Rocket server. Table 5.2 presents the results for the Rocket general LDA model, being the different topics with their corresponding labels, prevalence ranking, and its 20 most relevant terms.

From the 14 topics that come forward from the LDA model, the topics that are labeled with a high certainty are malware hosting, credential collection, general conversations on attack operations, cash-out, attacking Windows, Conti's cloud, two topics that cover general conversations on attack operation and two topics on general conversations among Conti members. The topics target selection, communication infrastructure, browser injection, and grabbers could not be la-

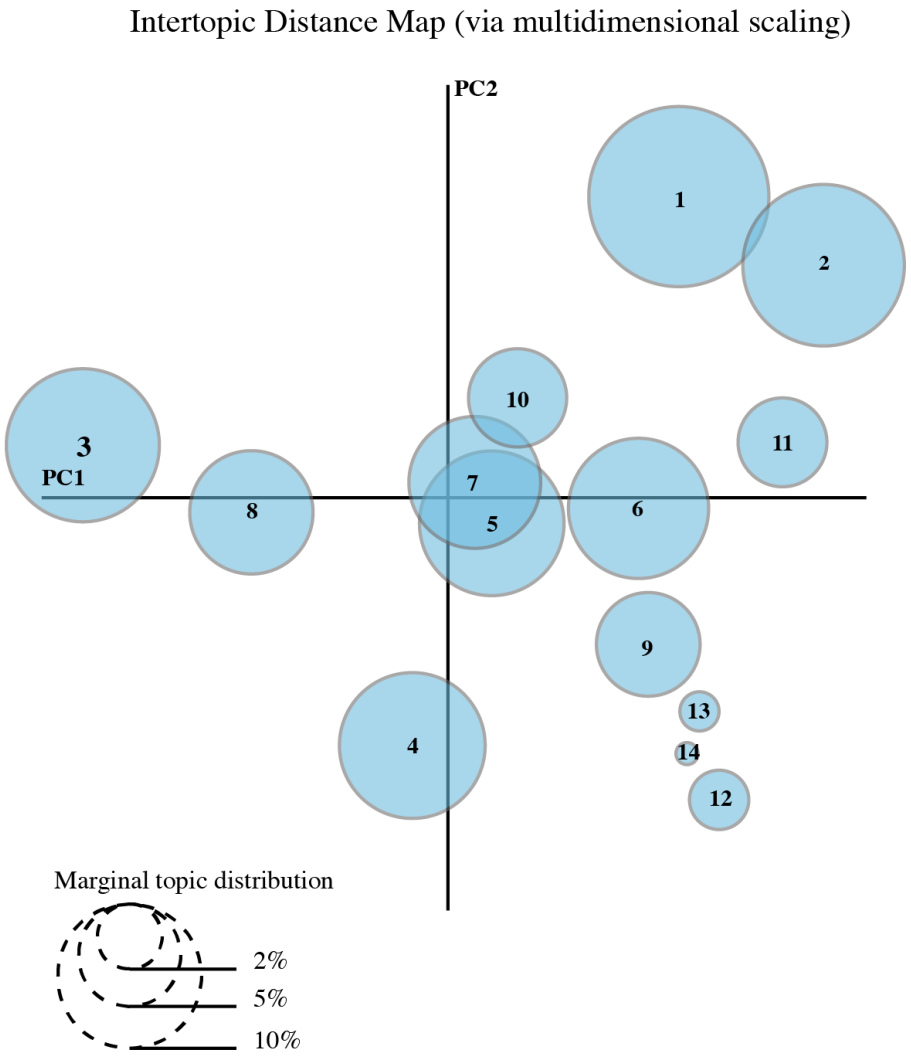

Figure 5.6: Intertopic Distance map Rocket general LDA model

beled with full certainty but may likely be labeled as we currently did. For topic 9, we could not identify a meaningful label based on the most relevant terms in their context. From these topics can be observed that the general channel of the Rocket server is generally used differently from the Jabber server. This is evident since the topics coming forward in the general Rocket channel do not focus on the operation of Conti as a business but rather revolve around ransomware attacks.

Topics 1 and 5 entail general conversations on attack operations which can be observed from terms such as "fly," "burn," "leak," "kill," "launch," "catch," "unload," and "hacker" which point to attack operations. Fly is a term that is generally used in the context of initializing an attack. In addition, unload and launch refer to distributing and launching the ransomware. Since topic 1 has the highest prevalence and the fact that there are two topics for general conversations on attack operations illustrates how the general channel of the Rocket server is more focused on ransomware attack operations. Similar to topics 1 and 5, topic 7 consists of messages dealing with general attack operations, which explains the overlap with topic 5, as shown in the IDM in figure 5.6. However, it is different from topics 1 and 5 in that these general conversations are more focused on attacks in which Microsoft Windows is used as the operating system. This is illustrated by the term "window," which refers to Windows but is changed to window because of applying lemmatization in the pre-processing of the chatlog data.

Table 5.2: Topics in the Rocket general chatlogs

#	**Topic**	**Relevancy λ and terms**
1	General conversation on attack operations	λ = 1: fly, buy, burn, service, price, link, game, word, update, leak, message, site, pass, hacker, catch, exchange, lie, ball, plow, watch
2	Target selection?	λ = 1: block, drive, watch, buy, russian, rise, fly, sell, link, clean, force, report, stick, post, switch, door, window, friend, video, guard
3	Malware hosting	λ = 0.8: host, server, exe, download, session, log, user, password, manual, script, delete, connect, domain, instal, clean, dump, execute, archive, version, credit,
4	Credential collection	λ = 1: sort, pull, collect, help, system, copy, computer, credit, log, download, host, team, server, cut, administrator, ip, session, user, message, provide
5	General conversations on attack operations	λ = 1: save, norm, roll, hard, log, lie, kill, map, launch, dump, solve, note, help, open, strongly, catch, red, problem, rest, unload
6	Cash-out	λ = 1: fly, lead, block, pull, fire, plan, project, wallet, problem, kosh, help, analysis, attack, gun, payment, direction, post, hand, record, exchanger
7	Attacking Windows	λ = 1: server, user, window, scan, message, wheelbarrow, code, pass, ip, connect, download, upload, log, determine, access, point, type, module, restart, exchange
8	Conti's cloud	λ = 1: decrypt, server, hash, domain, option, specify, open, ball, connect, launch, admin, rclone, lock, service, session, console, user, password, download, filter
9	???	λ = 1: us, car, help, require, scan, colleague, port, head, sound, rocket, knock, request, code, git, lock, lie, picture, demand, disable, return
10	Communication infrastructure?	λ = 1: key, hand, server, log, pass, save, cut, program, example, service, sim, sell, card, lose, request, people, number, buy, russian, live
11	General conversations among Conti members	λ = 1: healthy, us, lose, join, log, pm, buy, watch, mark, friend, sleep, hide, trick, word, payment, nature, die, connect, announce, body λ = 0.5: healthy, honor, muscovite, unnecessary, sanction, energy, necessarily, surplus, decorate, celebrate, announcement, paperclip, admit, creature, struggle, surgery, locomotive, orthodix, trick, earn
12	Browser injection?	λ = 1: case, option, route, level, window, session, rocket, machine, lie, browser, request, poke, us, connect, pass, decrypt, sort, party, tor, display
13	General conversations among Conti members	λ = 1: Toad, block, russian, problem, ukraine, site, us, server, gorec, adam, menu, git, general, message, scatter, source, bot, issue, idea, launch
14	Grabbers?	λ = 1: administrator, admin, base, enter, password, install, setting, team, member, net, pass, key, user, specify, figure, group, grandfather, [victim surname], grabber, clipboard

Topic 2 may revolve around target selection. Terms such as "watch," "friend," and "Russian" may indicate that Conti members are observing networks of potential targets and determine whom to attack and whom not. Other research on Conti's chatlogs has shown that Conti is likely to have close ties with the Russian Government and, therefore, may like to exclude Russian-based companies (Krebs, 2022a). Topic 3 contains conversations regarding the hosting of malware. Malware hosting is essential for distributing ransomware over victims' IT systems. For example, Conti needs to host their Command Control servers (C&C) from which they operate the ransom. In addition, Conti needs to host their malicious domains from which they social engineer victims into installing ransomware, as illustrated in section 5.2.2. Terms such as "host," "server," "session," "log," and "domain" point to hosting, while terms such as "exe," "download," "execute," and "archive" point to malware strains.

Topic 4 contains messages that point to the collection of credentials indicated by the terms pull, collect, copy, credit, log, server, administrator, and user. Pull, collect, and copy refer to the collection and copying of credentials, while the terms server, administrator, and user refer to targeting specific users. It may therefore be argued that Conti focuses on specific users and administrators with specific rights in targeted servers. Topic 6 contains messages regarding the cash-out of Conti's earnings from extorting victims. The term "kosh" refers to either cash or laundered assets, being more difficult to trace by law enforcement. Conti's cash-out operations are crucial for converting the retrieved ransom into usable assets. Since Bitcoin transactions, which are stored in its blockchain, are publicly available, Bitcoin is a relatively traceable cryptocurrency (Paquet-Clouston et al., 2019). Therefore, multiple techniques are used to make Conti's profits more difficult to trace to its source of illegal activities. Wallet refers to digital wallets that can be used to spend their earnings, and exchanger refers to criminals providing services of exchanging cryptocurrencies for cash.

Topic 8 entails messages regarding Conti's cloud. Conti uses their cloud, for example, to store obtained files, malicious programs, and passwords indicated by the terms "decrypt," "hash," and "ball." Decrypt and hash refer to kerberoasting, which is a technique that allows ransomware gangs to collect encrypted server passwords and subsequently decrypt these using hash cracking programs (Badhwar, 2021). Ball refers to a victim's shares which is a technical term for resources such as files, folders, or printers that have been made available to other users within the network (Morato et al., 2018). These shareable resources may, for example, be used to transfer confidential files to Conti's cloud storage or to easily infect multiple users within the network. Rclone is an open-source program for managing cloud environments and may be used within Conti to transfer files from- and to different clouds such as Conti's own cloud, victims' clouds, and public cloud services such as Mega.

Topics 10, 12, and 14 were labeled with a certain level of uncertainty, meaning that these topics may correspond to their labels but may have a different meaning. Topic 10 contains messages about Conti's communication infrastructure, indicated by the terms sim, card, and number. Topic 12 is likely to reform around messages that regard collecting credentials and cookies using browser injections, a technique that has been extensively discussed. Topic 14 has a low prevalence, meaning that it is not well represented among the messages send in the general Rocket channel but may regard grabbers. Grabbers are a form of malware that collects credentials and information from online forms or clipboards. Finally, topics 11 and 13 regard messages reforming around general conversations among colleagues that cover topics such as politics, the Russian-Ukrainian war, football, and popular movies and series.

Context of topics in the Rocket general chatlog data

From messages revolving around the general conversations on attack operations topics can be observed that the Rocker sever has an overall commanding leader that is responsible for the progress and success. Furthermore, it can be observed that there are three teams active in the Rocket server that each focus on separate attack cases, which gives meaning to the three clusters identified in the networked representation of conversations in the Rocket chats in section 4.2.

Each team has one or two team leads that are responsible for issuing cases, overviewing the progress of tasks, dealing with atypical tasks, and teaching, advising, and instructing their team members. Furthermore, team leads are responsible for documenting help articles on the Q&A forum that helps Conti members with certain attack-related topics. These examples illustrate the level of professionality within Conti's organization. In addition, it becomes evident that each channel represents a case that is named after its victim's full domain, e.g., google.com. However, since the naming of these channels is not decrypted, we cannot go further into Conti's targets. Finally, it can be observed that Conti is likely to have a physical office located in Russia, from which the attacking team is working. This contradicts the finding of all Conti members working remotely, as observed in topic 11 in the previous section.

Topic 3 revolved around malware hosting. Ransomware gangs are known to host malware using various ways (Tajalizadehkhoob et al., 2017). For example, Conti may steal credentials of legitimate registered domains in order to create a large number of sub-domains that are mapped to their servers hosting malicious content (Dargahi et al., 2019). In addition, malware can be hosted directly from compromised websites which is previously explained in section 5.2.1 or can be hosted using public cloud services and be distributed using links to this cloud (Dargahi et al., 2019; Tajalizadehkhoob et al., 2017). Finally, bulletproof hosting providers are likely used to host some of Conti's CC servers (Tajalizadehkhoob et al., 2017). However, this does not become evident from the conversations regarding topic 3.

From observing messages allocated to the collection of credentials, it becomes evident that Conti actively collects usernames and passwords from their victims to gain further access within their networks. The credential collection is often first attempted by extracting cookies, saved logins, and histories from browsers using open-source tools such as SharpChromium, which can be retrieved from GitHub. Security researchers often publish these tools for educational purposes, but Conti is actively searching for these tools so that they can be used within their attack operations. These tools are often secretly inserted using forms of social engineering, as previously illustrated. In addition, Conti uses a technique called Kerberoasting, which is a Kerberos-based attack, to collect Active Directory (AD) service account credentials. Kerberoasting is used following the compromise of a domain user account. Kerberoasting allows Conti to crack the active directory service account's credentials using third-party software (Badhwar, 2021). When collecting credentials, Conti focuses explicitly on users with administrator roles within the network, generally targeting domain administrators and enterprise administrators. Finally, it becomes evident that as a last resort, Conti may create personas of administrators to try and discover passwords based on personal information.

The messages regarding Conti's cash-out illustrate that each of the three teams active in the Rocket server has their own dedicated mixer to obfuscate profits, indicating that Conti's obtained ransoms are directly mixed when obtained. Mixers are specialized intermediaries that break links between senders and receivers by mixing coins and transactions with those of other users, making it extremely difficult to trace their origin (Paquet-Clouston et al., 2019). From conversations regarding Conti's cash-out, several tactics and techniques can be observed for turning personal salaries into untraceable currencies. Often Conti members discuss the use of exchangers, which relate to intermediaries that provide services of exchanging cryptocurrencies for cash using physical transactions. From the expert session with the cybercrime team of the FIOD, it became evident that exchangers are known to be commonly used. That these exchangers often use physical transactions becomes evident from messages stating that COVID-19 hindered exchangers because of obligations of QR-codes in Russia and messages showing that no exchangers are nearby. Secondly, Conti members discuss exchanging their mixed Bitcoins for XRP or Monero through loosely regulated Russian exchanges such as bestchange.ru and audia6.best. Third, digital wallets and cards are used to spend salaries obtained from Conti's activities such as Advcash, Tinkoff, and Webmoney, in which providers without identity verification are preferred. If identification is required, members mention using fake passports for verification.

Furthermore, when observing conversations allocated to the cash-out topic, it becomes evident that the Conti members active in the general Rocket channel are aware of the risks involved in each method of money laundering. For instance, it is discussed that digital wallets that are bound to exchanges should be avoided and that it is best to have a new wallet address for each transaction. Furthermore, tools such as the Anti Money Laundry bot (AMLbot) are used to verify how successful the obtained cryptocurrencies are laundered and to what extent the earnings are still relating to illegal sources. Conti has a member responsible for overseeing that the earnings are laundered before these are used for Conti's financial transactions or for the pay-out of salaries. Moreover, it becomes evident that another member has the responsibility of obtaining these laundered earnings so that these can be accordingly distributed over Conti's organization. Finally, from the messages related to Conti's cloud can be observed that in many cases, public cloud services and file transfer services are used to infect a defender's IT system with the Conti ransomware.

5.2.3 Topics in the Rocket sub-channels

In this section, the topics coming forward from the Rocket sub-channel are presented and discussed. Figure 5.7 presents the IDM of the topics in the Rocket sub-channels. From figure 5.7 can be observed that, in contradiction to the models previously discussed, most topics overlap. Since it becomes evident from section 5.2.2 that the sub-channels in the Rocket server are used for individual attack cases, we can expect more overlap in topics due to similarities. However, topics 1, 3, and 8 do not show overlap. Since the other topics all largely overlap, labeling these topics are likely to result in uncertain results and conclusions due to them being hard to interpret. Therefore, we only labeled topics 1, 3, and 8 and will discuss some of the interesting terms coming forward from the other topics. The generated topics with their prevalence ranking and their 20 most relevant terms are presented in table 5.3.

Topic 1 contains messages regarding general conversations on attack operations since the most relevant terms refer to a mixture of parts of the attack operations, such as reconnaissance and credential collection. The term help may indicate help being asked of different aspects of the attack operations. Topic 3 contains messages regarding the spreading of ransomware over the network indicated by the terms drive, scan, connect, enterprise, fly, and specify. Other keywords such as host, user, administrator, pull, and collect refer to reconnaissance and credential collection. Therefore, it may be argued that reconnaissance and credential collection is performed to identify how to spread the ransomware over the network. Furthermore, it can be observed that this topic has a large prevalence compared to the other topics, similar to the reconnaissance topic in the Jabber chatlogs. This implies that reconnaissance is likely one of the largest parts of Conti's activities for them to strategically plan their ransomware attacks.

Topic 8 contains messages on potential targets and users indicated by terms such as "zealand," "usa," "ro" and "nl." These countries and country codes relate to root servers and networks posted by the responsible Conti member. Terms such as "beavant," "roger" and "Williams" relate to users being observed in the network. These users may be observed for two primary reasons based on the context of other topics, as previously explained. Either Conti is stealing their credentials from their browsers, or administrator roles are observed within the network to find additional entry points.

When observing the most relevant terms of the topics, as shown in table 5.3, it becomes evident that the messages in the Rocket sub-channels predominantly contain technical terms. From the three topics discussed, it becomes evident that the sub-channels can be simplified into the three overarching activities reconnaissance, accessing networks, and account takeover. Terms such as "sharpzerologon," "credman," "sharpweb" and "piggy" all relate to both reconnaissance and gaining access to different network nodes. The ZeroLogon vulnerability is a vulnerability found in 2020 that allows intruders to comprise the domain controller account and control the entire corporate IT infrastructure (Bezzateev et al., 2021). It becomes evident that the ZeroLogon is still being used in July 2021, indicating that their victims' servers are not adequately

Table 5.3: Topics in the Rocket sub-channel chatlogs

#	Topic	Relevancy λ and terms
1	General conversations on attack operations	λ = 1: user, server, session, collect, launch, administrator, open, password, host, domain, computer, hash, service, scan, help, connect, pc, us, group, log
2	???	λ = 1: user, service, session, administrator, server, host, domain, delete, finish, remove, open, fly, dc, launch, credit, computer, access, backup, product, scan
3	Spreading ransomware over the network	λ = 1: host, user, pull, connect, administrator, domain, service, log, support, update, drive, center, enterprise, scan, dc, session, open, fly, server, specify
4	???	λ = 1: computer, connect, user, session, server, administrator, window, host, watch, password, copy, kill, die, pull, lock, domain, shoot, upload, delete, cut
5	???	λ = 1: session, host, user, domain, server, launch, controller, specify, dump, open, service, administrator, connect, password, download, attempt, delete, group, version λ = 0.5: sharpzerologon, correspond, compress, null, session, credman, trnsfer, ipak, redone, ssp, authentication, hashdump, controller, practice, scanned, art, worker, tspkg, wdigest, attempt
6	???	λ = 1: server, host, user, session password, domain, administrator, network, scan, account, backup, dc, car, browser, exe, question, computer, pas, note, encrypt, admins
7	???	λ = 1: user, store, host, session, specify, system, open, line, version, mark, browser, execute, nt, window, pull, argument, pc, service, decrypt, guide λ = 0.5: automate, decrypt, estate, ideas, pn, goal, piggy, styler, protected, exel, tokens, decode, demolition, whitelist, portable, store, sharpweb, flood, pr, bank
8	(potential) Targets and users	λ = 1: root, id, delete, zealand, usa, ff, backup, pull, server, free, dc, host, wine, pm, administrator, progress, rub, open, content, map λ = 0.5: zealand, usa, wps, nick, nl, ro, dj, cy, williams, wtg, rv, bos, thompsona, peel, reu, jerzy, decher, setter, beavant, roger
9	???	λ = 1: update, require, server, inject, follow, note, window, session, point, pass, connection, host, service, crash, register, generate, top, sort, case, error λ = 0.5: announce, injector, bug, belong, update, ahnlab, rozena, pursuit, lifetime, left, reload, signature, scripted, ic, dincheck, dllinstall, appsys, backupuser, loomisindiodb, objectnotfound
10	???	λ = 1: link, question, option, parameter, local, admin, fire, forum, system, program, delete, fly, burn, service, hand, pay, launch, domain, session, user
11	???	λ = 1: host, user, pc, service, administrator, account, server, machine, system, controller, version, request, computer, trust, password, open, log, domain, scan, dump λ = 0.8: pc, service, administrator, trust, host, user, intersection, machine, system, ipconfig, account, controller, version, request, tasked, computer, power, license, composition, rubeus

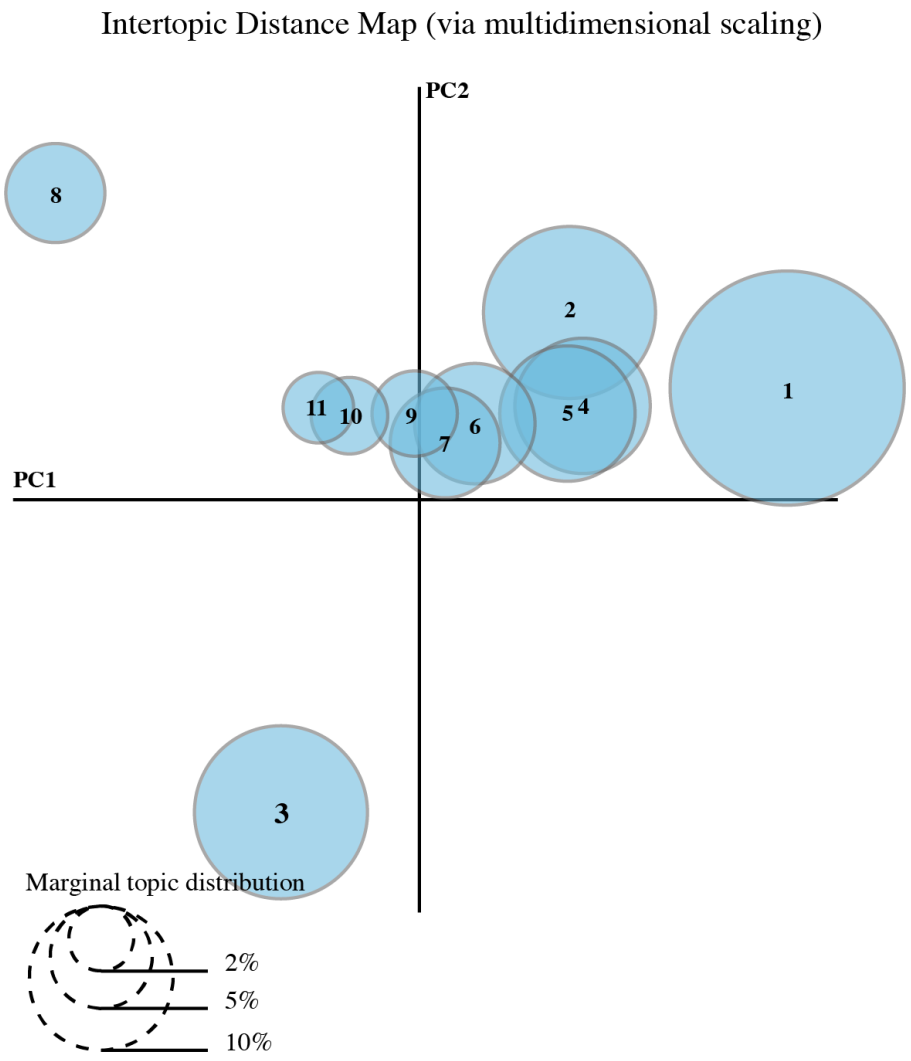

Figure 5.7: Intertopic Distance map Rocket sub-channel LDA model

patched. Sharpweb is an open-source program like SharpChromium that allows cybercriminals to observe browser data and credentials, while Credman refers to credential management which highlights another focus area within the credential collection (Diogenes and Ozkaya, 2019). Finally, Piggy is a trojan that can be used for gaining access, which confirms the thought of Conti utilizing external actors' products and services as commodities to strengthen their attack operations.

Context of topics in the Rocket general chatlog data

Many messages regarding topic 1 contain manuals that explain in detail how each of the different stages in an attack should be performed. For example, a manual on the different parameters and functions of Conti's locker and a detailed manual on entry point detection is presented. We do not discuss these manuals in detail since this can be a study on its own, and therefore, we leave this for future research. Furthermore, from the messages in topic 1 becomes evident that Conti's attack operations heavily rely on Cobalt Strike, as is currently the case for most ransomware gangs (Caroscio et al., 2022; Poudyal and Dasgupta, 2020).

Messages related to spreading ransomware over the network once more provide a detailed manual on how to spread across defenders' networks. Multiple tactics of spreading ransomware across a defender's IT system can be identified when observing the messages related to the spreading of ransomware. Generally, the spreading of ransomware is based on an extensive

reconnaissance of the victim's network, following the earlier explained tactics. That is, the victim's network is first fully scanned on hosts, servers, users, different drives, and sub-domains to determine all possible entry points and locations of confidential files in the network. Furthermore, when looking for confidential files, Conti focuses their search on GDPR-sensitive data and cyber insurance documents. GDPR sensitive data is targeted since they are aware of the regulatory framework regarding GDPR in which companies are fined when GDPR sensitive data gets leaked.

Using this knowledge, having compromised GDPR-sensitive data increases their negotiation strength. Similarly, cyber insurance documents can help better their position in negotiating the ransom. Subsequently, when access is retrieved to all nodes in the network, e.g., through kerberoasting, the ransomware is spread over the network using Cobalt Strike and the destinations coming forward from reconnaissance. Finally, cases can be observed in which Conti uses external platform-based companies, of which the target victim is a customer, to gain access to additional nodes in a victim's network. To illustrate, a case can be observed in which a software platform that focuses on the automation for independent insurance agencies is used to gain access to additional nodes within the network of an insurance broking company.

5.3 CONCLUSION

This chapter aims to determine overarching topics in Conti's internal communication data by applying LDA topic modeling. In section 5.1 the parameter settings of the three models are discussed. This resulted in the Jabber LDA model having 12 topics, the Rocket general LDA model having 11 topics, and the Rocket sub-channel having 14 topics. Based on the discussion of these topics in section 5.2 we can conclude that reconnaissance is one of the more important activities within Conti since it comes forward in both the Jabber- and Rocket general chatlogs as one of the more important topics and relatively has a high prevalence score. Furthermore, the context of the topics indicates that reconnaissance is being used to make an informed decision, for example, for spreading ransomware over the network. In addition, it becomes evident that the Jabber server is used differently from the Rocket server. The topics that come forward from running the LDA model on the Jabber chatlog data generally regard keeping Conti operational from a business perspective. It is, therefore, most likely that the Jabber server is mainly used for communication regarding Conti's strategy and business operations.

In contradiction, the topics coming forward from the Rocket LDA models do not discuss similar topics. More specifically, the topics in the Rocket chatlogs focus on executing attacks in different victim cases and the techniques used, illustrating that the Rocket server is used for operating attacks on victims. Furthermore, it becomes evident that sub-communities as identified in section 4.2 represent the three attack teams that are active in the Rocket server and that the sub-channels represent victim cases. Taking these differences between the chat services into account, it is likely that the strategical decisions and techniques predominantly come forward from the topics in the Jabber server. Conti's procedures and some of the relevant techniques used in attacks may come forward from the topics in the Rocket server. Therefore, in the next chapter, we lay the focus on topics in the Jabber server when mapping topics to Conti's TTP. That is, we choose to discuss all the topics that come forward from the Jabber server that were labeled with certainty when mapping Conti's TTP. In addition, we selectively discuss and map topics from the Rocket server if these give relevant insights into Conti's TTP.

il

6 LEVERAGING CONTI'S TTP TO RECONSTRUCT THE RANSOMWARE ECOSYSTEM

In the previous chapter, overarching topics were determined in Conti's leaked communication data. This chapter aims to map the identified overarching topics to TTP of the Conti ransomware gang. Subsequently, this chapter aims to leverage the mapped TTP to reconstruct the ransomware ecosystem while taking the perspective of Conti establishing interactions with actors in the ransomware ecosystem. Section 6.1 discusses Conti's tactics, techniques, and procedures (TTP), followed by section 6.2, which discusses the reconstruction of the ransomware ecosystem. Finally, section 6.3 presents a brief conclusion of the findings in this chapter.

6.1 MAPPING IDENTIFIED TOPICS TO CONTI'S TTP

In this section, the topics in Conti's internal communication, as discussed in section 5.2, are used to map Conti's tactics, techniques, and procedures (TTP). In other words, the unique characteristics that make Conti successful in its ransomware attacks are empirically determined based on ground truth communication data. While doing so, we focus on the topics coming forward in the Jabber server as these relate to most of Conti's strategical decisions (tactics) and techniques. The topics in the Rocket server are mainly used to give insights into Conti's procedures since the Rocket server is used for communication during attack operations. In addition, the topics in the Rocket server are selectively used to add relevant insights into Conti's tactics and techniques. While mapping the identified topics to TTP, the allocated messages to topics, which put the empirically determined topic in context, are used to come to more detailed TTP. This is particularly helpful for mapping topics to procedures since the mapping of procedures relies on the order of events or activities that are performed. Section 6.1.1 discusses the relevant tactics, section 6.1.2 discusses the relevant techniques, and in section 6.1.3 we discuss two relevant procedures that we identified from the discussion of topics.

6.1.1 Tactics

Business-related tactics

From the topics in Conti's internal communication, it becomes evident that Conti's business strategy revolves around several tactics. First, Conti has a well-defined hierarchical structure which is constructed of different teams within Conti's business units and in which employees have clear roles. This is, for example, indicated by the topics on HR and recruitment, payment infrastructure, cash-out, and the messages allocated to these topics that illustrate how different members have responsibilities. Furthermore, the Jabber communication network, as presented in figure 4.4, illustrates the hierarchies in Conti's organization. In addition, from other messages that reform around the general conversations on Conti operations can be observed how each team has a team lead that is responsible for its team's performance. These team leads frequently report their team's progress to Conti's high-level managers, each having a portfolio of business units and teams. Moreover, when observing the differences between the Jabber- and Rocket server, it becomes evident that Conti's attack operations are kept rather separate from Conti's internal operations. These examples illustrate how Conti tactically uses a divide and conquer approach to split their operations into smaller separate tasks, each performed by different teams.

This contradicts the picture that is often created by scientific authors in RaaS literature in which ransomware gangs are observed as just a group of developers creating the ransomware program (Bayoumy et al., 2018; Meland et al., 2020; Raheem et al., 2021). For example, in the RaaS value chain illustrated in the work of Meland et al. (2020), ransomware gangs are observed as authors of ransomware programs. The findings that illustrate how Conti is a ransomware gang with multiple teams, each with its own responsibility, for example, the HR and recruitment team or the three attack operation teams, contradict previous findings of ransomware gangs being just developers of the ransomware. Conti embedded the RaaS business model in which they help lesser tech-savvy affiliates use their ransomware for a percentage of the ransom. However, it can be observed that Conti is not just a group of developers and that their application of RaaS within their business should be observed differently from how it is often illustrated.

Conti is a large organization that is not only developing ransomware but is also performing many ransomware attacks themselves. To increase their revenue, they let their affiliates license the Conti formula and provide additional help, for example, through manuals and their IT helpdesk. In addition, based on the topics and allocated messages, there are no signs of Conti using vendors or Conti distributing their ransomware through dark web marketplaces, in contradiction to the previously illustrated ransomware value chain (Meland et al., 2020). Hence, it may be argued that ransomware gangs such as Conti take over the role of author, vendor, vulnerability researcher, and distributor within the presented RaaS value chain. This implies that the actual RaaS value chain should, in practice, be observed differently for ransomware gangs.

Building on the divide and conquer approach, it becomes evident from the development topic and its context that Conti does not develop their attack vectors from scratch but rather builds its attack vectors by incorporating combinations of other malware and open-source tools. In addition, both section 4.2 and the terms in the Rocket sub-channel illustrate that Conti is likely to collaborate with other cybercriminal organizations to strengthen their ransomware attack, of which Expiro and Piggy are examples. Regarding the purchase of attack vectors, it becomes evident from the topic in the Jabber chatlogs and examples from allocated messages that Conti externally purchases services such as exploits, crypts, loaders, hacking services, and access to networks. The prices paid for these services are often based on the revenue of victim companies. This shows how Conti further applies the divide and conquer approach by relying on the expertise of other cybercriminals rather than developing ransomware from scratch. This contradicts with how ransomware gangs are observed in the scientific literature as a sort of "tech scale-up" in which the development of the ransomware is central (Bayoumy et al., 2018; Meland et al., 2020; Raheem et al., 2021).

Van Wegberg et al. (2017) illustrated how different parts of the ransomware value chain can be outsourced using the commoditization of cybercrime. It can be observed that the commoditization of cybercrime evolved to a level in which services are provided that together construct parts of the ransomware value chain. This is different because these services do not outsource the different parts of the ransomware value chain but provide sub commodities to allow ransomware gangs to perform these parts themselves. This illustrates how the commoditization of cybercrime allows one of the most notorious ransomware gangs to build sophisticated attack vectors. As Richardson and North (2017) argued that ransomware gangs may use price discrimination based on the willingness to pay of victim defenders, the topics illustrate how price discrimination is similarly applied to illicit service providers. That is, ransomware gangs are only willing to pay based on the added value of the offered commodity, which is illustrated by the purchasing of networks based on company revenue.

The commoditization of cybercrime can also be observed from Conti building on the skills and networks of (future) employees, which is illustrated by the Recruitment and HR topic. Conti intentionally targets parts of its recruitment on dark web forum administrators to access their networks of cybercriminals. In addition, it becomes evident that when targeting candidates through recruitment sites, it is tactically decided to stay anonymous. Conti directly approaches candidates by email and does not allow any traces of Conti on the internet to stay off the

radar of law enforcement agencies. In addition, it can be observed from different topics that Conti intentionally uses several techniques in parallel rather than focussing more on a single technique. Using a mixture of techniques helps Conti to be more successful in their attacks and be more resilient to possible actions of law enforcement.

For example, it becomes evident that Conti does not use one dominant technique for storing files but rather uses a mixture of different services and techniques, such as using public cloud services and decentralized data storage. This is in line with the findings of other authors of ransomware gangs and other cybercriminals focussing on infrastructure and attack vector resilience (Ife et al., 2021; Mansfield-Devine, 2010; Zimba and Chishimba, 2019). Similarly, it becomes evident that Conti's infrastructure is configurated, so it is resilient to law enforcement, often having multiple elements set up in parallel. In addition, Conti does not rely on a standard of attack vectors but uses a variety of options which increases the chance of a successful ransomware attack.

The payment infrastructure topic illustrates how Conti focuses on blockchain-related currencies and exchanges for their payment infrastructure. Furthermore, it becomes evident that Conti may develop their own blockchain protocol to have more control. Conti may be innovating their business model by creating additional revenue streams by incorporating selling compromised victim data, preferably through its own blockchain protocol. However, the topics coming forward do not show signs of Conti having sold any victim data up to the present. Conti aims to become a central entity within the cybercriminal ecosystem by being the first to develop this cybercriminal blockchain. This protocol should form a basis for cybercriminal payments, increase the connectivity among cybercriminals, and needs to allow other cybercriminals to develop their own cybercriminal blockchain projects.

Reconnaissance-related tactics

While observing the topics in Conti's internal communication in its context, it becomes evident that reconnaissance is one of the most important activities that Conti performs to get to a successful ransomware attack. Reconnaissance allows Conti to retrieve information to strategically organize and plan ransomware attacks. Hence, it allows Conti to make informed decisions in the preparation and execution of ransomware attacks. Conti employs several tactics in relation to its reconnaissance. The reconnaissance topic indicates that possible environments such as popular antivirus and endpoint protection software are extensively observed. Conti tests its attack vectors using different versions of these security solutions to determine how attack vectors are detected. It may be argued that Conti continues developing the obfuscation of its attack vectors until those cannot be detected. The attacking Windows and the reconnaissance topics indicate that Conti observes the most recent versions of operating systems such as Windows to find unpatched vulnerabilities and test the defense mechanisms to Conti's attack vectors.

In addition, Conti actively searches for open-source tools and vulnerabilities that are useful to be incorporated within their attack operations. These tools and vulnerabilities are often published for educational purposes by cybersecurity researchers, as illustrated for SharpChromium by the credential collection topic. Hence, reconnaissance is also applied for acquiring additions to Conti's attack vectors. The third and most predominant part of Conti's reconnaissance tactics is designated to observing targeted defenders. For instance, Conti observes defenders' revenues, based on which defenders are targeted, through websites like ZoomInfo and publicly available financial statements. The reconnaissance topic in its context illustrates how Conti generally focuses on businesses with revenue starting from tens of millions, which implies that Conti follows a big game hunting strategy, as explained by Oosthoek et al. (2022).

Furthermore, in its reconnaissance, Conti focuses on observing the users with domain- or enterprise administrator roles, as illustrated by the credential collection topic. While doing so, it tries to apply customized forms of social engineering such as spear-phishing or genuine-looking webpages to lure these users into secretly installing malicious software. The (purchasing) attack

vector topic illustrates how spear-phishing services may be purchased through external actors. Finally, when using reconnaissance towards the location of confidential files, a focus is laid on files containing GDPR-sensitive data and cyber insurance documents since these files strengthen the negotiation position of Conti.

This is in contradiction with the work of McDonald et al. (2022) and Keshavarzi and Ghaffary (2020), in which reconnaissance is not part of the attack chain. Other scholars have acknowledged reconnaissance as being part of ransomware attacks (Al-rimy et al., 2018; Dargahi et al., 2019; Ibarra et al., 2019; Yunus and Ngah, 2021). These scholars highlight reconnaissance as one of the first stages in a ransomware attack in which a victim defender's territory is extensively observed in the preparation of an attack. However, the reconnaissance for open-source tools and vulnerabilities and the reconnaissance activities in which the ransomware is tested in different environments is not yet covered in scientific literature. Hence, the mapped reconnaissance-related tactics give further meaning to the importance of reconnaissance in a ransomware gang's operation and the role of reconnaissance in a successful ransomware attack.

6.1.2 Techniques

Many techniques come forward from the topics in the Jabber and Rocket chatlog data. From these topics, we highlight the most relevant techniques in this section. Multiple topics illustrate how keeping illegal activities hidden from the public is essential for Conti's continuance since this protects against being shut down by law enforcement agencies. Conti uses multiple techniques to keep their identities and illegal activities hidden and support its continuity. First, Conti uses two techniques in parallel for keeping their internal communications private, the PGP and OTR messaging protocols. Second, to hide their botnet servers behind an ever-changing network, they use the Fast Flux evasion technique. Third, to protect its domain names used in illegal activities, such as social engineering practices, from being taken down by law enforcement, Conti uses the EmerDNS protocol.

The EmerDNS protocol makes it unable for any authority to alter, revoke or suspend the domain name records, supporting Conti's infrastructure resistance (EmerCoin, 2022). Furthermore, bulletproof hosting providers are likely used to increase Conti's infrastructure resistance since Conti generally uses multiple techniques in parallel, and it became evident from Conti's infrastructure configuration topic that is likely that this tactic is also applied for hosting. Using these bulletproof hosting providers, Conti protects content from being taken down. However, the use of bulletproof hosting providers does not become evident from observing the topics. Finally, SiaCoin is used to store data decentrally. In practice, this means that the data is split into multiple segments, which are stored at different nodes in the blockchain. Not all segments are needed to reconstruct the data into its original file, meaning that this technique of storing files is harder to take down (Sia, 2022).

Authors have addressed cybercriminals using PGP, although no empirical evidence for ransomware gangs using PGP or OTR for communication has been provided in the scientific literature (Heinl et al., 2020; Kaur and Randhawa, 2020; Orman, 2016). Hence, these findings contribute to the scientific literature by providing empirical evidence of ransomware gangs using these anonymous communication protocols. The findings of Conti using the Fast Flux evasion technique and bulletproof hosting to make it more difficult for botnets and hosting to be taken down is in the line with the work of other scholars that have addressed similar findings for ransomware gangs (Lombardo et al., 2018; Meland et al., 2020; Surjanto and Lim, 2020). Furthermore, Conti using the EmerDNS protocol is in line with the work of Casino et al. (2021), in which they identified ransomware gangs using the Emercoin and Namecoin to protect their domain names against being taken down. However, Conti was not listed as one of the ransomware gangs. To our knowledge, ransomware gangs using Siacoin to decentrally store data is a novel finding that has not yet been covered in the scientific literature.

For reconnaissance, botnets are used to scan the networks for files, ports, hosts, and servers. In addition, browser injections allow Conti to observe stored passwords and cookies. It may be argued that these cookies may help Conti with creating better informed social engineering to lure a defender into installing malicious software since identified topics illustrate how customized social engineering is applied. This may create access to the defender's network. In relation to the progressive intrusion of defender networks, several techniques can be identified. For example, kerberoasting may be used to gain access to the network from a domain administrator or enterprise administrator. Secondly, Conti may exploit unpatched vulnerabilities such as the ZeroLogon vulnerability, providing them further access within the network. Furthermore, backdoor tools such as Gasket are used to create supplementary access to a defender's IT system. Finally, as can be observed from spreading ransomware over the network topic in section 5.2.3, in special cases, platform-based companies of which the victim is a customer may be used to create additional access to the network.

Eventually, Conti uses Cobalt Strike and botnets during the account takeover to distribute the ransomware and encrypt files. Furthermore, the topics on account takeover indicate that a victim's domain may be used to infect a victim's system with ransomware or that these may be compromised to distribute ransomware to other victims. Normally this ransomware consists of two DLL files and one .exe file. Most of these techniques have been covered in previous scientific work (Badhwar, 2021; Maroofi et al., 2020; Meland et al., 2020; Kao and Hsiao, 2018; Oosthoek et al., 2022; Reshmi, 2021; Surjanto and Lim, 2020). However, empirical findings of ransomware gangs using these techniques are often lacking in previous academic work. Hence, these findings contribute to the findings of previous scholars by providing empirical evidence of ransomware gang using these techniques. However, ransomware gangs utilizing browser injections and platform-based companies to create additional access to the targeted network have not yet been covered in the scientific literature. Multiple authors have addressed managed services providers (MSP) being targeted by ransomware gangs to exploit vulnerabilities in remote desk protocols. Still, the usage of platform-based companies for creating additional access within targeted networks is a novel finding regarding ransomware gangs' TTP (Beaman et al., 2021; Meland et al., 2020; Oosthoek et al., 2022).

To launder the obtained ransoms in untraceable assets, Conti uses several techniques. Mixers are used to break the link of cryptocurrencies with their illicit source. Second, the Anti Money Laundry bot (AMLbot) is used to determine to what extent laundered assets are traceable to illegal sources. Third, physical exchangers and loosely regulated crypto exchanges are used to convert the laundered Bitcoins into cash or more anonymous cryptocurrencies such as Monero. Fourth, digital wallets and payment infrastructures such as Webmoney and ADVCash are used, while verification is done with fake passports. Finally, pumping and dumping is used as investment technique to increase the values of earned income. Authors such as Oosthoek et al. (2022) discussed how ransomware gang launder obtained ransoms. However, the ransomware gangs utilizing tools such as the AMLbot to determine the level of success of laundering assets is a novel finding in the scientific literature. Furthermore, when the pumping and dumping technique is put in the context of the work of Galinkin (2021), it becomes evident that in contradiction to defenders, ransomware gangs are, to a certain extent, able to positively influence the value of payments.

6.1.3 Procedures

Many interesting procedures come forward from the topics and allocated messages. However, we will only discuss three procedures and leave the discussion of other procedures for future research. First, it can be observed that Conti's attack chain can be simplified into six stages, as shown in figure 6.1. Other scholars have discussed attack chains for ransomware gangs in previous academic work, e.g., see Dargahi et al. (2019), McDonald et al. (2022), Ibarra et al. (2019), Keshavarzi and Ghaffary (2020), and Yunus and Ngah (2021). The attack chains discussed by these scholars are often observed from a technical perspective observing the ransomware pro-

gram and focusing on the network intrusion and account takeover. Ibarra et al. (2019) and Yunus and Ngah (2021) include reconnaissance in their work, but the importance of reconnaissance in the attack chain is not yet clearly discussed. Other authors such as van Van Wegberg et al. (2017) and Oosthoek et al. (2022) have discussed the cash-out phase of ransomware gangs in their operations. This implies that the purchasing of access to networks and the target selection stages in the attack chain are often overlooked and that none of the discussed attack chains includes all stages. This can be explained due to the lack of research focusing on the internal operations of ransomware gangs, which is identified as a knowledge gap in section 1.2. The simplified attack Chain for the Conti ransomware gang includes all six stages that can be observed from a high level and contribute to the previous work of scholars by including all these stages based on empirical findings.

First, access to companies' networks is bought through external cybercriminals in which the prices paid are based on the companies' revenues within the networks. From these networks, targets are selected, which is done based on the yearly revenue and is retrieved from publicly accessible resources as previously illustrated. As Conti is generally mixing different techniques and tactics, Conti may also find initial access to networks themselves based on their targets, indicating that they may skip phase 1 in certain cases.

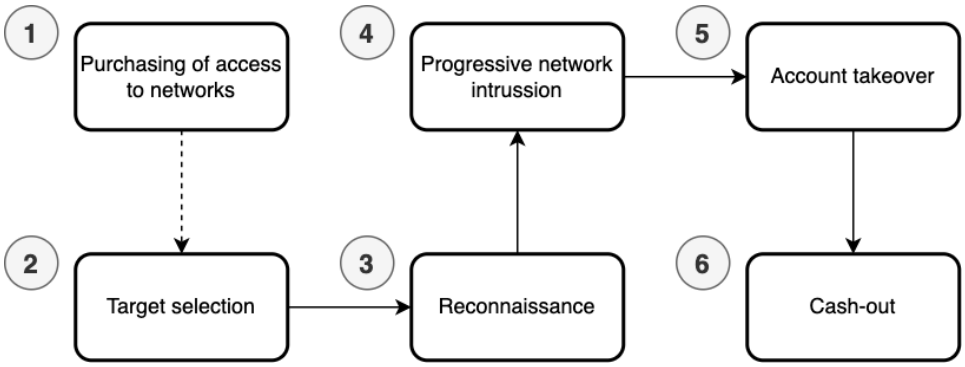

Figure 6.1: Graphical overview of Conti's attack chain

When a specific defender is targeted, Conti focuses on extensively performing reconnaissance to observe the different servers, domains, administrators, passwords, and confidential files, as previously illustrated by the topics. Based on the outcomes of the reconnaissance, Conti progressively intrudes the network until they have fully accessed all nodes within a defender's network. Subsequently, the defender's files are encrypted using Cobalt Strike and botnets. Finally, the paid ransoms are diligently laundered, making it very difficult to trace back the earnings to the source of illegal activities. Since it is evident from the topics that Conti's reconnaissance is a reason for Conti being so successful in its attacks and since the successful cash-out provides incentives for performing ransomware attacks, we discuss these processes in more detail.

As previously argued, reconnaissance has been mentioned in the scientific literature as one of the first steps in the attack chain of ransomware gangs (Dargahi et al., 2019; Yunus and Ngah, 2021). However, the reconnaissance procedure of ransomware gangs has not been studied, and it is often observed as just observing a defender's IT system and the associated vulnerabilities. The identified topics and their context illustrate how Conti follows a reconnaissance procedure that allows them to make informed decisions while executing ransomware attacks. The identified procedure, therefore, adds meaning to the role and importance of reconnaissance in the attack chain. Conti's reconnaissance is generally constructed of five phases and starts with observing the target revenue, as previously explained. A complete overview of Conti's reconnaissance procedure is presented in figure 6.2.

After a victim's revenue is determined and the victim case is issued, its network is scanned using botnets to determine its structure and content. This allows Conti to determine which users they should focus on, often being users with the domain- or enterprise administrator rights.

Subsequently, both browser injections and the social engineering of victims are performed. The order in which these activities are performed could not be determined since both activities amplify each other. More specifically, the use of browser injections helps Conti members to create behavior-based spear-phishing, while social engineering may be used to lure a victim into the secret injection of these observation tools into its browser. It can be observed that from these browser observations often passwords are found that give Conti further access to a victim's IT system, for example, by compromising a domain administrator's account.

That ransomware gangs scan the network for users, and their contact details and that social engineering is used during the reconnaissance is in line with the work of Ibarra et al. (2019). However, that reconnaissance is used for customized social engineering, as indicated in stage 3 in figure 6.2, is a novel finding. While having access to the domain administrator's account, Conti shifts its focus to targeting the Enterprise administrator to gain access to the full network. Techniques that could be employed in this stage are kerberoasting, a more elaborate scanning of the network, and using platform-based companies to gain access to other nodes within the network. That is, from the initial access, Conti progressively intrudes on the network until it has fully accessed the victim's network. We note that the reconnaissance procedures may be different based on the context of the specific case, for example, because the context of the case makes it more efficient to directly focus on the enterprise admin. However, we believe that the procedure that we distinguished is representative of most cases, based on the allocated messages that we observed.

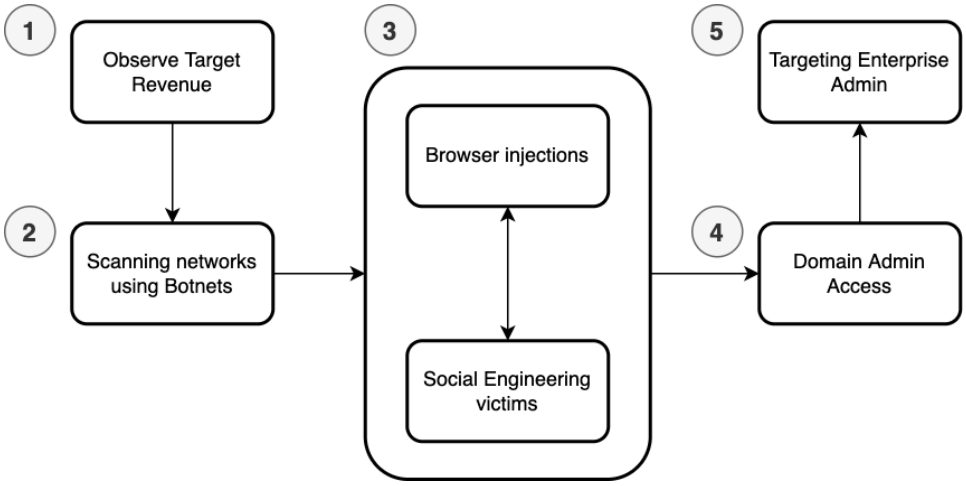

Figure 6.2: Graphical overview of Conti's reconnaissance procedure

The topics indicate that Conti has a structured procedure for laundering their assets, making it hard to trace these to their origin. More specifically, Conti's money laundering procedure consists of six phases, as illustrated in figure 6.3. Each of the three attack teams receives the negotiated ransom amount in a form of cryptocurrency, likely in Bitcoins. These earnings are then mixed using the team's dedicated mixer since the topics and their context illustrated that each team has its own mixer. Subsequently, Conti observes to what extent the mixed cryptocurrencies still relate to illegal activities using the AMLbot to identify to what sources the assets are linked. When the results of this validation show that the assets look like they are being legitimately acquired, the laundered assets are used for Conti's internal cash flow and to pay employees' salaries. Subsequently, Conti's employees cash out their salaries using physical exchanges or loosely regulated exchanges such as audia6.best. These earnings can then be invested in pumping and dumping manipulation schemes to increase their value. Finally, employees use cash, digital wallets, and digital money protocols such as WebMoney to spend their salaries.

Money laundering strategies for ransomware gangs have been discussed in the scientific literature, for example, by Meland et al. (2020) and Oosthoek et al. (2022). In the scientific literature, it is commonly argued that ransomware gangs may use mixers and loosely regulated exchanges

(Laszka et al., 2017; McDonald et al., 2022; Meland et al., 2020; Oosthoek et al., 2022). Different scholars have been studying the financial trails of Bitcoins that were used in ransom payments, e.g., see Bayoumy et al. (2018) and Oosthoek et al. (2022). However, because of the use of these mixers and loosely regulated exchanges, it is difficult for scholars to observe the money laundering procedures of ransomware from an external perspective. In addition, Oosthoek et al. (2022) argue in their work that RaaS ransomware gangs such as Conti have sophisticated laundering procedures, which makes it difficult to identify chokepoints in money laundering. The identified money laundering procedure of the Conti ransomware gang may provide means for identifying chokepoints in the money laundering procedures of RaaS ransomware gangs. In addition, the identified money laundering procedure provides a more detailed overview of money laundering procedures to contribute to previous academic work of scholars that researched money laundering in ransomware gangs from an external perspective.

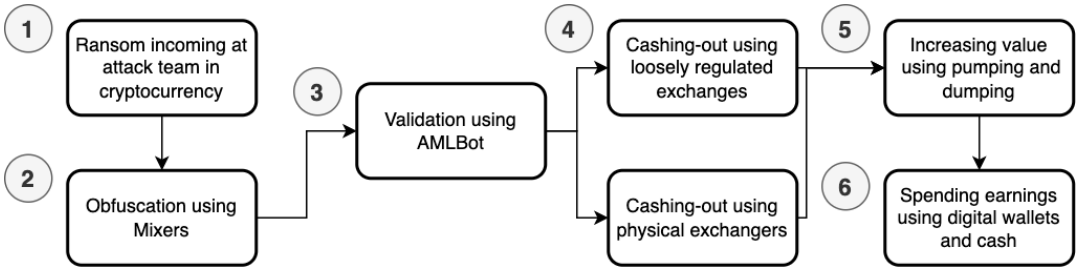

Figure 6.3: Graphical overview of Conti's money laundering procedure

6.2 RECONSTRUCTING THE RANSOMWARE ECOSYSTEM

This section discusses how the identified TTP can be leveraged to reconstruct the ransomware ecosystem. While doing this, we take the perspective of the Conti ransomware gang and lay the focus on how ransomware gangs establish interactions with actors in the ransomware ecosystem to reconstruct each of the three sub-ecosystems. More specifically, we are leveraging the identified TTP as discussed in section 6.1 by putting them into the context of the three sub-ecosystems of the ransomware ecosystem, which are the attacker ecosystem, the defender ecosystem, and the governance ecosystem. Chapter 2 of this thesis discussed what actors comprise the ransomware ecosystem and how these are related. In section 6.2.4. the findings in chapter 2 are synthesized with the reconstructions of the three sub-ecosystem to come to a reconstruction of the ransomware ecosystem.

6.2.1 Attacker ecosystem

From the mapped TTP, it becomes evident that Conti is a large professional organization which is in line with the research of other scholars and cybersecurity companies (Checkpoint, 2022; Figueroa et al., 2022). Ransomware gangs that adopted the RaaS business model are often observed in scientific literature as being the developers of ransomware that is licensed to their affiliates to perform attacks (Bayoumy et al., 2018; Meland et al., 2020). The mapped TTP illustrate a contradicting picture since Conti focuses its development on purchasing tools and services and integrates and adjusts these to form its attack vectors. Furthermore, the TTP illustrate how the commoditization of cybercrime, as identified by van Van Wegberg et al. (2017) evolved to a further extent than outsourcing parts of the ransomware value chain. It becomes evident that Conti establishes interaction with other service-providing cybercriminals in the ransomware ecosystem to use their expertise and resources, which can be observed as sub-commodities to construct parts of the ransomware value chain, rather than outsourcing separate parts of the value chain. Relevant examples that we discussed are the use of criminal actors to purchase access to networks of victims, the distribution of its ransomware through Phishing-as-a-Service, and multiple services involved in Conti's cash-out. Furthermore, Conti establishes

interactions with other malware organizations such as Expiro to be able to strengthen their attack vectors.

Conti being a large professional organization and choosing to establish interactions with service-providing cybercriminals instead of developing the ransomware from scratch illustrates how other cybercriminals comprise the attacker ecosystem. More specifically, it illustrates how the attacker ecosystem consists of other cybercriminals that provide services that have such a quality that Conti is willing to choose to incorporate these services instead of constructing their ransomware value chain completely by themselves. Furthermore, the mapped TTP illustrate how Conti differs from how RaaS ransomware gangs are described in most scientific work Bayoumy et al. (2018); Meland et al. (2020). Conti is a large organization that performs sophisticated ransomware attacks themselves. To increase their revenue, they license the Conti formula to affiliates and provide additional help and service to teach and instruct them. This illustrates how Conti establishes interactions with attackers in the attacker ecosystem in two ways.

On the one hand, Conti interacts with actors in the attacker ecosystem to strengthen its attack vectors and infrastructure. On the other hand, Conti interacts with its affiliates to help them perform successful ransomware attacks, following the Conti formula. The divide and conquer strategy that Conti follows by establishing these interactions indicates that the trend of collaboration is likely to develop only further. In addition, other ransomware gangs are likely following similar trends since it is beneficial for them to create additional revenue streams through RaaS and collaborate with service-providing cybercriminals to strengthen attack vectors. The innovation strategy of Conti developing a blockchain protocol coupled with a cybercriminal social network is a perfect example of how ransomware gangs plan to utilize further interaction with other cybercriminals in the attacker ecosystem. Currently, Conti predominantly establishes interactions with other cybercriminals using the networks of employees and inner-sphere forums. Introducing a blockchain protocol provides anonymous means for actors within the attacker ecosystem to establish interactions with other cybercriminals in the attacker ecosystem.

Hence, it can be argued that the attacker ecosystem comprises large professional ransomware gangs such as Conti, affiliates, inner-sphere forums, and other cybercriminals providing services to ransomware gangs. These large ransomware gangs perform sophisticated ransomware attacks themselves, and following the RaaS business model, they license their formula to affiliates. Rather than developing the ransomware themselves, these large ransomware gangs interact with cybercriminals that provide services to construct their ransomware value chain. These interactions are often facilitated by inner-sphere forums. If ransomware gangs compete or collaborate is an ongoing debate among scholars (Cartwright et al., 2019a; Richardson and North, 2017). The findings do not indicate if Conti competes or collaborates with other ransomware gangs. Hence, there may be competition or collaboration between large ransomware gangs. Still, these larger ransomware gangs interact with other service-providing cybercriminals to collaborate and leverage the commoditization of cybercrime.

The reconstruction of the attacker ecosystem contradicts the findings in the work of Meland et al. (2020) and Bayoumy et al. (2018) in which RaaS is posited as being distributed through dark web markets. As Meland et al. (2020) and van Van Wegberg et al. (2017) argue that easily accessible dark web markets do not play as big of a role as often assumed, the reconstruction of the attacker ecosystems builds on these findings by illustrating a more accurate picture of how the attacker ecosystem is comprised. By observing RaaS ransomware gangs such as Conti in the attacker ecosystem, we believe the business model of RaaS may be better understood than observing RaaS as being distributed through dark web markets. To further reconstruct the ransomware ecosystem, we first turn to reconstruct the defender ecosystem in the next section.

6.2.2 Defender ecosystem

The mapped TTP illustrate how Conti continuously establishes interactions with defenders in the defender ecosystem. Most of these interactions can be appointed to Conti performing recon-

naissance activities on defenders in the defender ecosystem. Conti deliberately establishes interactions with popular security vendors and operating systems to observe new developments in these defense mechanisms. Subsequently, Conti extensively tests its attack vectors within each environment and adjusts its techniques and attack vectors accordingly to prevent them from being detected. Secondly, Conti continuously observes defenders in the defender ecosystem, searching for exploits, open-source tools, and information that allows them to improve their attack operations.

These defenders are publishing these exploits, open-source tools, and information with the goal of strengthening the defense of other defenders in the defender ecosystem, which we may assume has a positive effect. However, publishing these resources results in these defenders unknowingly assisting Conti in strengthening its attack chain. Similarly, tools such as Cobalt Strike are initially designed to help protect defenders against cybercriminal attacks but are currently also used by ransomware gangs in their account takeover. Again, this illustrates how defenders in the defender ecosystem aim to provide services that strengthen the defense of other defenders in the defender ecosystem but indirectly strengthen the attack vectors of ransomware gangs. It, therefore, becomes evident that Conti establishes interactions with defenders in the defending ecosystem that in itself provide a harmless service, but in the context of Conti's operations, conflict damage. Other examples are Siacoin, EmerDNS, and cloud services that generally provide a harmless service but, in the context of ransomware, support ransomware gangs in constructing a resilient infrastructure.

In addition, Conti establishes interactions with victim defenders. Again, most of these interactions are based on the extensive reconnaissance of defender territories. Conti secretly observes defenders' browser activities and their complete IT infrastructure to progressively intrude on defenders' systems. In addition, anonymous interactions with recruitment sites are made to gain access to potential employees. Finally, it can be observed that the connectivity of defenders plays a role in ransomware attacks and has a greater extent than only targeting managed service providers (MSP) to exploit vulnerabilities in remote desk protocols. It becomes evident that Conti is on the lookout for defenders that have customer-supplier relationships with targeted defenders, in which the customer-supplier relationships focus on bringing connectivity, e.g., by being a platform-based company. These defenders may subsequently be used to intrude on additional nodes in the targeted defender's network. These examples of interactions illustrate how ransomware gangs establish interactions within the defender ecosystem to increase their chances of performing a successful ransomware attack.

Hence, the defender ecosystem is comprised of defenders that perform actions to strengthen their defense against ransomware. In addition, these actions may be performed to strengthen the defense of other defenders, such as security vendors developing their environments or publishing open-source tools and vulnerabilities. These defenders may provide generally harmless services such as cloud services to other defenders in the defender ecosystem. Ransomware gangs establish interactions with defenders in the defender ecosystem by observing the actions to strengthen their defense, observing defenders' IT systems, leveraging generally harmless services to strengthen their ransomware attack, leveraging connectivity between defenders, and performing ransomware attacks. Prior literature has barely studied the defender ecosystem since most scholars studied ransomware from a descriptive perspective or studied ransomware as a relation between a defender and ransomware gang (Chen et al., 2021; Galinkin, 2021; Laszka et al., 2017). Moreover, reconnaissance of ransomware is often limited to the reconnaissance of defenders' IT systems (Dargahi et al., 2019; Yunus and Ngah, 2021). Hence, the reconstruction of the defender ecosystem creates a more accurate picture of how ransomware gangs establish interactions to defenders in the defender ecosystem.

6.2.3 Governance ecosystem

When projecting Conti's TTP to the governance ecosystem, it becomes evident that Conti is aware of rules and regulations within the ransomware ecosystem. For instance, Conti observes

the GDPR regulatory frameworks within defender's territories to better their position during the ransom negotiations. Moreover, Conti intentionally establishes interactions with loosely regulated cryptocurrency exchanges in which they are aware of the rules and regulations set by countries that confine the different exchanges. This is further illustrated by Conti members knowing the risk involved in the different methods related to the spending and laundering of their earnings. In other words, Conti is strategically dodging well-regulated territories based on the involved risk. Moreover, this may indicate that Conti is indirectly aware of tactics, techniques, and procedures that law enforcement agencies and governance actors use to investigate and regulate ransomware gangs. However, no ground-truth data is supporting this.

Hence, the governance ecosystem comprises governance actors that create and maintain the governance framework that confines the rights of defenders and cybercriminals. This can be observed from the GDPR regulation in defender territories and from the regulations that confine cryptocurrency exchanges. Ransomware gangs observe the regulatory frameworks in defender territories and strategically dodge these, for example, by choosing loosely regulated exchanges to avoid the risk of being caught. Prior literature has focused on governance actors influencing defenders to adopt better preventive measures or influencing defenders to be transparent on paying ransoms (Chen et al., 2021; Kenneally, 2021; Richardson and North, 2017). However, prior literature has not covered how ransomware gangs can leverage GDPR regulation and how they observe and utilize loosely regulated exchanges. Hence, the reconstruction of the governance ecosystem based on Conti's TTP creates a clearer picture of the role of governance actors from the eyes of ransomware gangs.

6.2.4 A reconstruction of the ransomware ecosystem

As illustrated in the previous sections, ransomware gangs establish interactions within each of the three ecosystems to improve their chances of having a successful ransomware attack. While leveraging the mapped TTP, these three sub-systems were reconstructed in section 6.2.1, 6.2.2, and 6.2.3. These findings are enriched with the identified actors and their interactions that came forward from reviewing the scientific literature in chapter 2. Combining these findings results in a reconstruction of the ransomware ecosystem in which we find that the ransomware ecosystem can be reconstructed as an interconnected ecosystem constructed of the attacker ecosystem, the defender ecosystem, and the governance ecosystem.

Large professional ransomware gangs such as Conti operate from the attacker ecosystem in which they establish interactions with other cybercriminals to purchase services that form commodities in their ransomware value chain. They license their sophisticated formula of performing ransomware attacks to affiliates to increase their revenue which allows them to perform ransomware attacks. These affiliates are provided with detailed help through services and helpdesks. Inner-sphere dark web forums and the networks of employees facilitate the establishment of interactions between ransomware gangs and other cybercriminals in the attacker ecosystem. As Oosthoek et al. (2022) argue, commodity ransomware may have a role in the attacker ecosystem. However, commodity ransomware is observed as just a proving ground for higher-impact utilization of ransomware. Therefore the role of commodity ransomware is likely to be little in the attacker ecosystem (Oosthoek et al., 2022).

The defender ecosystem comprises defenders that strengthen their defense through multiple actions. While doing so, these defenders may strengthen the defense of other defenders. In addition, as Cartwright et al. (2019a) argue, defenders may benefit from other defenders spending more on their defense against ransomware. However, as Pal et al. (2021) argue, defenders may harm other defenders through cascading effects in their supply chain or through found confidential files that allow ransomware gangs to extort other defenders. Furthermore, these defenders may provide generally harmless services to other defenders that ransomware gangs can use to their advantage, such as cloud services and defenders providing services that interconnect defenders. Ransomware gangs and their affiliates establish interactions with defenders in the defender ecosystem by performing ransomware attacks, observing their IT systems through

reconnaissance, and observing how defenders strengthen their defense. These observations are then used to strengthen their ransomware attacks.

The governance ecosystem is comprised of regulatory actors that create and maintain the governance framework that influences the attacker ecosystem and the defender ecosystem. These regulatory actors influence defenders in the defender ecosystem, for example, by creating rules and regulations on GDPR, payment transparency, or by regulating cryptocurrency exchanges. In addition, these governance actors may create policies for cyber-insurers and promote and advocate for not paying ransoms (Chen et al., 2021). However, these governance actors may support ransomware gangs by lacking strict regulation of exchanges or through state-sponsored attacks (Lee et al., 2019; Keshavarzi and Ghaffary, 2020). Ransomware gangs observe the regulatory frameworks that confine and define the rights and responsibilities of defenders in the defender ecosystem. Moreover, ransomware gangs may leverage regulatory frameworks that confine the rights of defenders to strengthen their negotiation position, as is the case for GDPR. In addition, ransomware gangs use their observations of governance frameworks to strategically dodge strict regulated territories, which is illustrated by ransomware gangs choosing loosely regulated exchanges.

As argued in section 1.1.5, cyber-insurers cannot be easily categorized in one of the sub-systems. However, our research contributes to the argument of ENISA (2021), McDonald et al. (2022), and Galinkin (2021) that cyber-insurers fuel the ransomware economy since it becomes evident that ransomware gangs focus on discovering information on this insurance which increases the chances of getting paid and increases the ransom value that can be asked. From the topics in Conti's internal communication and Conti's TTP, no effects of cyber-insurers hindering Conti's attack operations can be observed.

Many scholars have studied the ransomware ecosystem. The ransomware ecosystem in their work is often considered as a collection of ransomware programs or -gangs or collection of ransomware programs or -gangs that attack defenders (Bayoumy et al., 2018; Fang et al., 2020; Huang et al., 2018; Kaptchuk et al., 2017; Laszka et al., 2017; Lee et al., 2019; Mei et al., 2021; Raheem et al., 2021). However, a clear definition or description of the ransomware ecosystem is never presented. Therefore, these scholars leave out the context of the attacker ecosystem, the defender ecosystem, and the governance ecosystem. This can be explained due to the lack of empirical research and since most research has been done from a descriptive or technical perspective. The findings in this thesis illustrate how the context of these sub-ecosystems of the ransomware ecosystem is important for understanding how ransomware gangs operate within the ransomware ecosystem. Moreover, it becomes evident how ransomware gangs establish interactions within each of these sub-ecosystems and how establishing these interactions is an important factor in their success. Hence, we find that the reconstruction of the ransomware ecosystem as presented in this thesis creates a more accurate and complete illustration of how ransomware gangs operate within the ransomware ecosystem.

6.3 CONCLUSION

This chapter aims to map the identified topics to Conti's TTP and leverage the mapped TTP to reconstruct the ransomware ecosystem. From mapping the identified topics to Conti's TTP, it becomes evident that Conti is a large and professional organization that performs ransomware attacks themselves and licenses the Conti formula to their affiliates to increase their revenue. As argued, this differs from how RaaS is often observed in scientific literature. Conti relies on the services of other cybercriminals in the attacker ecosystem to construct their ransomware value chain, which illustrates the quality of these services due to Conti being one of the most dominant ransomware gangs. Furthermore, the mapped TTP illustrate how reconnaissance is one of the most important activities that Conti performs to get to a successful ransomware attack. Moreover, the findings illustrate how the importance of reconnaissance of ransomware gangs has been previously overlooked in the scientific literature. The mapping of Conti's TTP

illustrate novel techniques that Conti uses and have not yet been discussed in scientific literature. In addition, while mapping the topics, we discussed Conti's attack chain-, reconnaissance-, and money laundering procedure. Based on leveraging the mapped TTP, the ransomware ecosystem can be reconstructed from the attack ecosystem, the defender ecosystem, and the governance ecosystem. Ransomware gangs operate from the attacker ecosystem and establish interactions within each of the three ecosystems to get to a successful ransomware attack. The reconstructed ransomware ecosystem creates a more complete and accurate illustration of how ransomware gangs operate from the ransomware ecosystem.

7 DISCUSSION

7.1 DISCUSSING THE RESULTS IN CONTEXT

This thesis researches a ransomware gang's establishment of interactions from ransomware gangs to actors in the ransomware ecosystem. This is done by mapping tactics, techniques, and procedures (TTP) of the Conti ransomware gang based on topics that are empirically determined in their internal communication. These TTP are then used to reconstruct the ransomware ecosystem. To our knowledge, this thesis is the first academic effort that empirically researches a ransomware gang's TTP and their establishment of interactions within the ransomware ecosystem. Since the most dominant ransomware gangs are large and well-structured organizations and their success relies on well-executed communication, it may be argued that their internal conversations entail topics dealing with their tactics, techniques, and procedures. The research goal of this thesis stipulates that it aims to retrieve insights into the internal establishment of a ransomware gang's interactions with actors in the ransomware ecosystem and reconstruct the ransomware ecosystem accordingly. Hence, this section discusses the findings of this thesis from the perspective of the ransomware ecosystem.

The results indicate how Conti uses its Jabber server for conversations regarding business operations and uses its Rocket server for communication on attack operations. In addition, it becomes evident that Conti is a large and professional organization that performs ransomware attacks themselves and licenses its formula to affiliates to increase their revenue, following the RaaS business model. The findings indicate that reconnaissance is one of the most important activities that ransomware gangs perform to get to a successful ransomware attack since it allows for informed decision-making. Finally, the results indicate that the ransomware ecosystem can be reconstructed from the attacker ecosystem, the defender ecosystem, and the governance ecosystem. In the ransomware ecosystem, ransomware gangs operate from the attacker ecosystem and strategically establish interactions within each sub-ecosystem to get to a successful ransomware attack. Examples of interactions that ransomware gangs are establishing are interactions to utilize the commoditization of cybercrime, to observe defenders' IT systems and behaviors, to observe rules and regulations related to their risk, and interactions to improve and extend their attack vectors. The establishment of interactions with actors in the ransomware ecosystem is fully embedded in the ransomware gangs' business strategies, resulting in their tactics, techniques, and procedures forming around the establishment of interactions.

7.1.1 Results in relation to prior scientific work

Prior research on the ransomware ecosystem is descriptive and technical. It takes an external perspective when observing ransomware, and most studies only observe specific parts of the ransomware ecosystem in isolation (Huang et al., 2018; Lee et al., 2019; McDonald et al., 2022). This can be observed from scholars using the ransomware ecosystem concept with different scopes and definitions, while a clear definition of the ransomware ecosystem is never presented (Bayoumy et al., 2018; Fang et al., 2020; Huang et al., 2018; Kaptchuk et al., 2017; Laszka et al., 2017; Lee et al., 2019; Mei et al., 2021; Raheem et al., 2021). This research is novel in comparison to prior research in that it researches ransomware gangs while taking an internal perspective through empirically mapping TTP and observing ransomware in its full context of the ransomware ecosystem. In addition, it reconstructs the ransomware ecosystem based on

these findings, giving definition and meanings to the widely used concept of the ransomware ecosystem. The findings of ransomware gangs establishing interactions in the attacker ecosystem by using services that form sub-commodities to construct a ransomware value chain are in line with the work of McDonald et al. (2022) and van Van Wegberg et al. (2017). McDonald et al. (2022) argue that ransomware gangs will continue evolving tactics to encourage payment. The evolution of tactics can be observed from ransomware gangs performing extensive reconnaissance to adjust and strengthen their attack vectors and from ransomware gangs relying on services provided by other cybercriminals to construct parts of their ransomware value chain.

Van Wegberg et al. (2017) argue that separate parts of the ransomware value chain can be outsourced. The results in this thesis illustrate how ransomware gangs evolved from outsourcing separate parts of their value chain to using different services of other cybercriminals in the attacker ecosystem and combining and adjusting these to construct the parts of the ransomware value chain. Ransomware gangs tactically decide to use these services instead of developing these parts themselves using scarce IT talent. This shows that these services have such a quality that it is beneficial for large ransomware gangs to do so. Moreover, this illustrates how it is likely that other cybercriminals have evolved to provide more specific and higher-quality services. Cartwright et al. (2019a) argue that ransomware gangs are aware of the state of backups of defenders, which is in line with the reconnaissance-related findings in this thesis. Our findings illustrate how ransomware gangs do not only use reconnaissance to observe a defender's IT infrastructure. More specifically, ransomware gangs use sophisticated reconnaissance procedures in which they progressively intrude on defenders' IT infrastructure, use reconnaissance to observe actions by defenders that strengthen their defense and use reconnaissance to observe useful tools and information that strengthens their ransomware attacks. This illustrates how reconnaissance has a more important role in the success of ransomware attacks than is often argued for in the scientific literature. Therefore, these results contradict findings in previous scientific work that observes reconnaissance as just extensively observing a defender's IT system (Al-rimy et al., 2018; Dargahi et al., 2019; Ibarra et al., 2019; Yunus and Ngah, 2021).

As previously illustrated, the results indicate that RaaS should be observed differently than illustrated in the work of Bayoumy et al. (2018) and Meland et al. (2020). Ransomware gangs such as Conti perform sophisticated ransomware attacks themselves and license their formula to affiliates. Subsequently, they provide help through IT helpdesks and manuals to assist these affiliates in performing attacks following this formula. This differs from the work of Bayoumy et al. (2018) and Meland et al. (2020) in which ransomware gangs are observed as just the ransomware developers, and the ransomware is distributed through vendors over dark web marketplaces. The results indicate that ransomware gangs rather take the role of author, vendor, and vulnerability researcher, although vulnerabilities are also externally purchased. The results do not explicitly indicate through which channels RaaS is distributed. However, as Meland et al. (2020) and van Van Wegberg et al. (2018) argue that the effects of public dark web markets are limited on the distribution of RaaS, and the results indicate that targeting dark web forum administrators in recruitment is embedded in Conti's TTP, it may be argued that RaaS is distributed through inner-sphere forums. Hence, the results indicate that the RaaS value chain is, in practice, different for ransomware gangs than previously argued in the scientific literature.

How ransomware gangs establish interactions within the defender ecosystem has barely been studied besides the interactions of performing ransomware attacks. In addition, little scientific research has focused on how ransomware gangs establish interactions within the governance ecosystem. Research on the governance ecosystem has mainly focused on how governance can be applied to the defenders in the defender ecosystem, for example, to support payment transparency or incentivize defenders through cyber-insurers to adopt preventive measures (Kenneally, 2021; Laszka et al., 2017; Richardson and North, 2017). Similarly, research on the defender ecosystem has focused on defenders making decisions for preventative measures (Cartwright et al., 2019a; Chen et al., 2021; Laszka et al., 2017). The lack of research on how ransomware gangs establish interactions within these sub-ecosystems explains how the importance of reconnaissance is previously overlooked in scientific literature since the results in this thesis illustrate

how most of the interactions that ransomware gangs establish within the defender- and governance ecosystem, are designated to reconnaissance.

The contrast between prior scientific research and the results in this thesis can be explained due to the lack of empirical research on ransomware gangs. The lack of empirical research results from a lack of ground truth data since ransomware gangs put in an effort to keep their operations and communication hidden from the outer world. To our knowledge, the leaked communication data of the Conti ransomware gang is the first source of ground truth data of RaaS ransomware gangs. While using this ground truth communication data to empirically determine topics and map these to TTP to describe how ransomware gangs internally operate, this thesis is the first academic work that empirically researches the internal operations of ransomware gangs. Hence, the results in this thesis provide empirical evidence to substantiate the findings of prior scientific work, create a clearer illustration of how ransomware gangs establish interactions within the ransomware ecosystem and are the first findings to provide meaning and definition to the ransomware ecosystem based on empirical findings.

7.1.2 Interpretation of the results

In this thesis, empirically mapped TTP of the Conti ransomware gang are leveraged to reconstruct the ransomware ecosystem. Conti is a representative of ransomware gangs in the ransomware ecosystem since it is the largest and most dominant ransomware gang since the shutdown of REvil. Other ransomware gangs are likely to follow similar TTP because of their success (ENISA, 2021). In addition, in a recent study, Bátrla and Harašta (2022) argue that Conti, REvil, and DarkSide/BlackMatter are the three most representative ransomware gangs in the current ransomware ecosystem. The results indicate that RaaS should be observed as large professional ransomware gangs that are not just a group of ransomware developers but perform sophisticated ransomware attacks themselves and license their formula to affiliates to gain additional revenue. Therefore, ransomware gangs in the ransomware ecosystem can be observed in two categories: large professional RaaS ransomware gangs, such as Conti and Darkside/BlackMatter, and their affiliates. This can be explained by all the most dominant ransomware gangs following the RaaS business model (ENISA, 2021).

The strategic advantage that RaaS ransomware gangs gain from establishing interactions in the ransomware ecosystem can be explained by observing Conti's TTP. By building on the expertise of other cybercriminals in the attacker ecosystem, RaaS ransomware gangs outsource subcommodities and adjust and combine these into the individual parts of their ransomware value chain, as illustrated by van Van Wegberg et al. (2017). This allows RaaS ransomware gangs to improve the quality of their ransomware value chain elements use multiple techniques in parallel, increasing their attack success and resilience. This can, for example, be observed from Conti having complex reconnaissance procedures in which they deploy multiple techniques to gain information about a victim's territory. Based on this information, it is decided how to progress. Using this perspective, it can be argued that generally, RaaS ransomware gangs follow a divide and conquer strategy in which they divide the complex task of performing sophisticated procedures into multiple smaller parts. These smaller elements may then be distributed over different teams or outsourced to external actors and subsequently reconstructed into a complete and sophisticated ransomware operation.

Conti incorporating these services illustrates how the attacker ecosystem has evolved from cybercriminals providing services that allow outsourcing separate parts of the ransomware value chain to providing high-quality sub-commodities for these parts. The fact that the most dominant RaaS ransomware gangs use these services instead of developing them by themselves illustrates the availability and quality of the provided services in the attacker ecosystem. These services are a key factor for RaaS ransomware gangs developing ransomware which has varied and sophisticated attack vectors and is resilient against different defense mechanisms. Ransomware evolved from commodity ransomware to RaaS, to using double extortion and is now leveraging services in the attacker ecosystem to evolve to more sophisticated and effective ran-

somware. Hence, it may be observed that the trend of RaaS ransomware gangs strategically establishing interactions with other cybercriminals in the attacker ecosystem is likely to develop further. This is substantiated by the results that illustrate the need for a central blockchain protocol coupled with a cybercriminal social network.

Leveraging these services allows RaaS ransomware gangs to focus more on their reconnaissance. The findings in this thesis show that performing extensive reconnaissance is one of the major factors for a ransomware gang's success in digitally extorting their victims. RaaS ransomware gangs and their affiliates intentionally interact with defenders in the defender ecosystem to retrieve valuable information on their territories. This information supports making informed tactical decisions during ransomware attacks which illustrate how reconnaissance information is a valuable resource for customizing attacks, increasing the potential for success. In other words, based on the information gained by the reconnaissance, ransomware gangs can adjust their tactics and configurations of attack vectors. This is facilitated by RaaS ransomware gangs incorporating sub-commodities from services of other cybercriminals in the attacker ecosystem since this leads to a more varied set of attack vectors. While performing reconnaissance on victim territories, ransomware gangs may use the digital connectivity between defenders in the defender ecosystem as a resource within their attack operations.

Other interactions that are established with actors in the defender ecosystem are based on observing changes in defense mechanisms, the publishing of open-source tools and exploits, and utilizing legitimate services that increase their infrastructure resilience. This shows how ransomware gangs interact with the defender ecosystem in a twofold manner. On the one hand, ransomware gangs are being hindered by defenders increasing their defense mechanisms but observe these actions and adjust their attack vectors accordingly. On the other hand, ransomware gangs gain important resources through reconnaissance that supports them in planning and constructing a successful ransomware attack. Hence, the interactions between ransomware gangs and defenders can be observed as a strategic game between the two, although ransomware gangs are clearly winning. This can be explained due to ransomware gangs continuously evolving their attack vectors based on their reconnaissance and because of this reconnaissance they have valuable information on defenders' defense.

Since RaaS ransomware gangs such as Conti license their formula to their affiliates, affiliates are not establishing interactions within the attacker ecosystem to adjust and construct these parts of the ransomware value chain themselves. However, they benefit from the RaaS ransomware gang establishing these interactions from the resulting higher quality ransomware. These affiliates do perform reconnaissance, following the sophisticated reconnaissance procedures that are part of the RaaS ransomware gang's formula. This reconnaissance does not entail observing changes in the defense mechanisms of defenders since this is development related but focuses on observing the defenders' territories. Performing these reconnaissance procedures is supported by detailed manuals and help services from the RaaS ransomware gang they are affiliated with, as illustrated by the results. Similarly, it can be argued that using the digital connectivity of defenders may only apply to RaaS ransomware gangs since this requires more technical expertise, and affiliates are often observed as lesser tech-savvy ransomware gangs (Oosthoek et al., 2022; Van Wegberg et al., 2017).

In addition, the results indicate that RaaS ransomware gangs are aware of regulatory frameworks in the ransomware ecosystem. This results from their reconnaissance of observing rules and regulations. Knowing where rules and regulations are positioned can be used to their advantage by adjusting tactics and procedures based on this knowledge. An illustrative example that comes forward from this thesis is selecting exchanges for cashing-out earnings based on their knowledge of the regulatory framework that confines these exchanges. While doing so, ransomware gangs strategically reduce their risk of getting on the radar of law enforcement agencies. As elaborated upon in interim discussions, this may indicate that RaaS ransomware gangs are aware of tactics, techniques, and procedures used by law enforcement agencies and adjust their strategy and procedures based on these. Since these procedures are crucial for the

financial gains of a ransomware attack, and RaaS ransomware gangs benefit from their affiliates succeeding, likely, their affiliates do not perform reconnaissance to observe regulatory frameworks in the ransomware ecosystem. It is more likely that RaaS ransomware gangs perform the observations of regulatory frameworks and construct the findings in procedures for affiliates since affiliates are likely to lack the skill and knowledge for these observations. Therefore, these findings only apply to RaaS ransomware gangs.

Although Conti's TTP can clearly be observed as a representative of other RaaS ransomware gangs in the ransomware ecosystem, some of the findings should be observed more specifically for Conti. All RaaS ransomware gangs are likely to utilize services provided by cybercriminals in the attacker ecosystem since this allows for a more sophisticated and varied set of attack vectors. This can be explained due to the high-level quality of services that are provided, and utilizing these services relieves RaaS ransomware gangs of the burden of hiring scarce IT talent. Therefore, utilizing these services seems to be the dominant strategy for RaaS ransomware gangs in the ransomware ecosystem. Similarly, it is evident that all ransomware gangs in the ransomware ecosystem are aware of regulatory frameworks and strategically use these observations to dodge strictly regulated exchanges. This can be explained due to RaaS ransomware gangs designing money laundering procedures that are crucial for laundering the financial assets gained from ransomware attacks while staying off the radar of law enforcement agencies. As the findings illustrate that Conti is aware of the regulations in place for GDPR and cryptocurrency exchanges, it may be argued that other ransomware gangs observe these regulatory frameworks as well. A simple reason for this is the fact that regulation is often publicly available and the fact that there are many loosely regulated exchanges that allow ransomware gangs to compare exchanges and choose the best option.

Since Conti is the largest and most dominant RaaS ransomware gang and having a largely varying set of attack vectors and techniques requires experience and liquidity, the variety of attack vectors and techniques may therefore be less widespread for smaller RaaS ransomware gangs. In addition, the exact techniques and procedures observed for Conti may have some differences for each RaaS ransomware gang. The observed attack chain for Conti is situated on a higher level and, therefore, can be applied to all ransomware gangs in the ransomware ecosystem. The reconnaissance- and money laundering procedures have more detail, and therefore each RaaS ransomware gang may show some differences in these procedures. For example, using browser injections in combination with social engineering relates to Conti actively searching for open-source tools and vulnerabilities, and other RaaS ransomware gangs may use different techniques. However, Conti's reconnaissance procedure illustrates the importance and sophistication of reconnaissance in all RaaS ransomware gangs' operations.

Similarly, for the money laundering procedure, other ransomware gangs may not show the validation of mixed cryptocurrencies or the pumping and dumping technique, as these findings apply more to Conti. However, the incoming earnings being directly mixed and being cashed out using physical exchangers or loosely regulated exchanges apply to all ransomware gangs, while other steps in the money laundering procedure may differ per RaaS ransomware gang. This can be explained by the findings of Oosthoek et al. (2022) that illustrate how RaaS ransomware gangs use mixers as one of the first steps to stop law enforcement agencies from observing their financial trails. We did not leverage the identified techniques to project these on other ransomware gangs, since Conti uses many techniques and parallel, and some of the novel identified techniques may only apply to Conti. It can also be argued that the techniques used in each ransomware gang is related to the skill of the employees. Since Conti is one of the largest and most dominant ransomware gangs they are able to use many techniques in parallel, while other ransomware gangs may not. Therefore, novel identified techniques may apply to other ransomware gangs, but we cannot say this with certainty based on a single case study.

Lastly, the results illustrate how the ransomware ecosystem can be reconstructed from the three sub-ecosystems in which ransomware gangs establish interactions. Ransomware gangs operate from the attacker ecosystem in which they establish interactions to commoditize sub-

commodities to construct their ransomware value chain. As the results illustrate that each of the three sub-ecosystems affects the success of ransomware gangs, it is implied that the phenomena of ransomware cannot be objectively studied without observing it in the context of the ransomware ecosystem as reconstructed in this thesis. Therefore, the reconstruction of the ransomware ecosystem based on the empirical findings of Conti contributes to the scientific literature by providing a framework in which ransomware should be studied to include possible influences in its success. We do acknowledge that this framework is not covering all possible effects on the success of ransomware gangs. Still, we will further discuss these in the limitations of this thesis and recommendations for future scientific work.

7.2 IMPLICATIONS

7.2.1 Scientific relevance

As discussed in section 7.1.1, this thesis makes several contributions to the prior scientific literature on ransomware. These contributions are summarized in this section, and it is discussed how these address the identified knowledge gaps in the academic literature. In the introduction of this thesis (see section 1.2), two knowledge gaps in the academic literature were identified. The identified knowledge gaps are addressed in this thesis as follows:

> *There is a lack of understanding of how ransomware gangs establish interactions with actors in the ransomware ecosystem*

Three contributions are made that provide an understanding of how ransomware gangs establish interactions with actors in the ransomware ecosystem. Firstly, this thesis is the first academic effort to present a reconstruction of the ransomware ecosystem based on three subsystems, being the first to give a description and definition of the ransomware ecosystem concept. This is done by observing the ransomware ecosystem as being constructed of three subsystems the attacker ecosystem, the defender ecosystem, and the governance ecosystem. In the ransomware ecosystem, ransomware gangs operate from the attacker ecosystem and establish interactions within each of the three sub-ecosystems. By providing meaning and definition to the ransomware ecosystem, future scientific work may use the ransomware ecosystem concept presented in this thesis to study the phenomena of ransomware in its full context.

Secondly, empirical findings are presented on the tactics, techniques, and procedures that the Conti ransomware gang uses in their internal operations, being the first academic effort to empirically study a RaaS ransomware gang from an internal perspective. This resulted in novel identified tactics, techniques, and procedures that have not yet been discussed in the scientific literature in relation to ransomware gangs. These empirically mapped TTP are then leveraged to observe how Conti and other ransomware gangs establish interactions with actors in the ransomware ecosystem. Thirdly, the empirical findings on Conti's TTP and the leveraging of these TTP on the ransomware ecosystem illustrate how ransomware gangs establish interactions with actors in the ransomware ecosystem. This contributes to prior scientific work by providing empirical evidence substantiating the findings in previous work or by building on the work of previous authors to create a clearer illustration of how ransomware gangs establish interactions within the ransomware ecosystem. These findings, for example, illustrate how the commoditization of cybercrime as illustrated by van Van Wegberg et al. (2017) evolved to ransomware gangs utilizing services to outsource sub-commodities and adjust and incorporate these to form parts of their ransomware value chain. In addition, the findings illustrate how the RaaS value chain should be observed differently than illustrated by Bayoumy et al. (2018) and Meland et al.

(2020) and illustrate how reconnaissance is one of the major factors that lead to a successful ransomware attack and is often overlooked in scientific literature.

> *There is a lack of research based on ground truth data*

This thesis addresses this gap in the current scientific knowledge by using the leaked internal communication data of the Conti ransomware gang and empirically determining topics in these chatlog data by following an extensive pre-processing methodology and applying LDA topic modeling. These topics are subsequently mapped to Conti's TTP, which is a proofed methodology to describe a ransomware gang's internal operations. While doing so, as previously argued, this is the first scientific work to empirically study how a RaaS ransomware gang internally operates. The lack of research based on ground truth data is, for example, illustrated by reconnaissance being overlooked in the current scientific literature. The ground-truth-based findings in this thesis contribute to the gap in current scientific literature by providing empirical findings that put prior findings from a more descriptive and technical perspective on ransomware in context. Secondly, this thesis provides a set of topics associated with the most relevant terms in Conti's internal communication. These topics and relevant terms may be used in future scientific work to validate findings based on ground truth data. Finally, the methodology used in this thesis addresses the knowledge gap in current scientific knowledge since it provides a novel methodology for empirically researching large ground truth communication datasets using LDA. These ground truth datasets are not designed for research, so an extensive methodology is needed to work with these datasets. In this thesis, the work of different scholars that have used LDA topic modeling is combined and improved to form a novel methodology that allows scholars in future work to apply our proposed methodology to large and hard-to-study ground truth datasets.

7.2.2 Implications for law enforcement agencies and policymakers

The findings in this thesis have several implications for law enforcement agencies and policymakers. First, the findings illustrate how the Jabber server is used for conversations regarding Conti's business operations while the Rocket server is used for conversations on attack operations. These findings can help law enforcement agencies in their forensic research. For example, if law enforcement agencies are interested in responsible Conti team leads on several of Conti's business units, the empirical findings indicate that they should focus their investigation on the Jabber chatlog data. Similarly, if law enforcement agencies are interested in the exact procedures of money laundering or the exact procedures and techniques used by Conti in their attacks, they should focus on the Rocket chatlog data. In addition, the empirically determined topics in both chat servers and the associated most relevant terms provide law enforcement agencies guidelines on which terms they can use to search for possible evidence related to that topic. Since the results in this thesis provide law enforcement agencies a guideline on how to use these chatlogs in their investigation, it saves law enforcement agencies valuable time and allows them to efficiently use the chatlog data for forensic research and the preparation of prosecutions.

Second, the findings illustrate how ransomware gangs use legitimate services such as cloud services, EmerDNS, Siacoin, and Cobalt Strike to strengthen their attack operations. The vendors of these legitimate services can be observed as defenders whose rights and obligations are confined by the governance framework created by policymakers (governance actors). The results indicate that Conti uses these legitimate services to strengthen their attack operations. Similarly, the findings illustrate how ransomware gangs can retrieve the newest versions of security solutions to test their attack vectors and adjust them accordingly. It may be worthwhile for policymakers to create new policies that hinder cybercriminals from using these legitimate services to their advantage. In addition, the results indicate that published security information and open-source tools are used by ransomware gangs to their advantage. Therefore, policymakers may focus on communicating and educating defenders in the defender ecosystem on

the consequences of publicly publishing security information and open-source tools and create guidelines on how to share this information among the defenders in the defender ecosystem.

Finally, the findings illustrate how it may be worthwhile for law enforcement agencies to focus on targeting inner-sphere forums. The findings illustrate how it is likely that inner-sphere forums are the facilitators of the interactions between RaaS ransomware gangs and other cybercriminals in the attacker ecosystem in which they outsource sub-commodities of the separate parts of their ransomware value chain. It is illustrated in the findings how ransomware gangs utilize these services to come to more sophisticated and more varied attack vectors. Hence, targeting the inner-sphere forums that facilitate these interactions may hinder ransomware gangs from establishing these interactions within the attacker ecosystem. Similarly, it is argued that the distribution of RaaS from RaaS ransomware gangs to affiliates may go through these inner-sphere forums. Since all of the most dominant ransomware gangs follow the RaaS business model, targeting inner-sphere dark web forums could hinder the distribution of RaaS to affiliates, leading to lesser affiliates and, therefore, lesser ransomware attacks.

7.3 LIMITATIONS

This thesis is limited in a few ways. We will discuss these limitations and their implications for the results following the limitations for the research approach, limitations for the methodology & data, and the limitations for the interpretation of the findings.

7.3.1 Research approach

First, it must be noted that only a single case is studied. This has a negative impact on the generalisability and external validity of the findings (Seawright and Gerring, 2008; Yin, 2009). This limits the research in that it hinders determining what techniques are used specifically for Conti and what techniques are used by multiple ransomware gangs. In addition, to find money laundering and reconnaissance procedures that generally apply to all ransomware gangs in the ransomware ecosystem, more cases need to be studied. However, as previously explained, Conti is currently the most dominant RaaS ransomware gang, and other ransomware gangs will likely follow their TTP; therefore, Conti's TTP is representative of the ransomware ecosystem. In addition, the leaked communication data of Conti is the only source of ground truth communication data of RaaS ransomware gangs that is currently available.

Second, this thesis does not include the empirical studying of the establishment of interactions by defenders and governance actors in the ransomware ecosystem. The establishment of interactions by defenders and governance actors in the ransomware ecosystem is included in this thesis by reviewing scientific literature. However, this is barely studied in the scientific literature, as previously illustrated. As the results illustrated how the defender- and governance ecosystem influence ransomware gangs, these interactions are important to study to get a clearer picture of the ransomware ecosystem. Hence, the reconstruction of the ransomware ecosystem is not yet a finished concept but provides a basis from the ransomware gang's perspective and should be further studied to determine how defenders and governance actors establish interactions within the ransomware ecosystem and how this affects ransomware gangs.

7.3.2 Methodology & data

Since not all messages in the used Conti chatlog data are decrypted and translated, and these messages were removed by extending the stop-words, some relevant topics may be excluded from the determination of topics in Conti's internal communication. Since LDA works based on word frequencies, and because word frequencies may be different or important words are left out of the corpus, other topics in the TTP could come forward. However, since the important findings in this thesis are based on topics with relatively high prevalence in the corpus, and the

encrypted and untranslated messages only take up a small part of the chatlog data, this is likely only to have an effect on the less prevalent topics.

Secondly, the topic models that are created, following the methodology in this thesis, are not useful for clustering messages to topics. This can be explained due to LDA assuming a distribution of topics to be represented in each document and the fact that topics are not mutually exclusive. In addition, to cluster the messages the hyperparameters should be further optimized to clustering messages, which may lead to a lesser quality of topics. Therefore, messages allocated to topics could only be observed based on the most relevant terms, meaning that the most relevant terms of a topic indicate that a message with that term is at least allocated to that topic. This implies that some messages should be clustered to topics, although these are currently not since these are clustered on terms that do not occur in the 20 most relevant terms. Hence, the validation and explanation of the topics could be improved by clustered messages. However, since the most relevant terms are the most important terms in that topic, it can be argued that this could only extend the current findings with other relevant TTP and does not has significant effect on the current results.

Lastly, applying LDA topic modeling following the clustering and pre-processing methodology on the Rocket sub-channel chatlog data has shown to be moderately effective due to similarities and technical terms used in the documents used for analysis. This is illustrated by the three out of 11 topics that were labeled and by the other topics that could not be properly labeled. Since the Rocket chatlogs are used for attack operations mainly, being able to properly label these topics could lead to identifying additional techniques and procedures used by Conti. The focus of this thesis was laid on the topics coming forward from the Jabber LDA model. Since the Rocket chatlogs were only used to substantiate findings on Conti's techniques and procedures, this has little consequence for the results on the ransomware ecosystem. However, by using a different methodology for the Rocket sub-channel chatlog data in future research, novel techniques and procedures used by Conti in their attack operations could be identified.

7.3.3 Interpretation of the findings

Since LDA is an unsupervised algorithm, it is unknown upfront what these topics are. These topics are labeled based on the most relevant terms in their context, and these labels are validated using the allocated messages to topics and based on an expert session with the FIOD. However, some of the topics could not be labeled with certainty, and topic 9 in the Rocket general chatlogs could not be meaningful labeled. As previously argued, the Rocket chatlogs are primarily used in this thesis to substantiate findings on techniques and procedures. Therefore this implies that novel techniques and procedures may be identified for applying a different methodology. The Rocket server is used by Conti for their attack operations, and therefore more technical terms are used. Topics identified by the LDA algorithm are determined based on word frequencies. Therefore it could be argued that some identified topics do not have semantic meaning and, therefore, cannot be labeled. However, the interpretation of topics in this thesis is limited to our technical knowledge and the technical knowledge of the attendees of the expert session with the FIOD.

Finally, due to the lack of cases, the projection of Conti's TTP on other ransomware gangs might be misinterpreted. That is, some of the findings of the Conti ransomware gang that are currently projected on all ransomware gangs in the ransomware ecosystem might be findings that only apply to Conti, while some of the findings that are currently designated to only apply for Conti may apply for the ransomware ecosystem. Further research based on ground truth data should indicate to what extent the generalization of findings to the ransomware ecosystem is currently performed correctly. However, since Conti is the largest and most dominant ransomware gang which is due to their TTP, other ransomware gangs are likely to follow similar TTP, meaning that most of their findings as currently projected can be observed as correctly. Further research should validate to what extent there are differences in TTP of ransomware gangs and to what extent the projecting of findings of Conti's TTP to the ransomware ecosystem is done correctly.

7.4 RECOMMENDATIONS FOR FUTURE RESEARCH

Firstly, future research is needed to get further insights into money laundering procedures and reconnaissance procedures that apply to all ransomware gangs in the ransomware ecosystem. This research should focus on empirically studying the internal operations of other ransomware gangs to identify generic procedures and to further validate the projection of Conti's TTP on the ransomware ecosystem. In addition, future studies should research how defenders and governance actors establish interactions within the ransomware ecosystem to further define the ransomware ecosystem concept. By further researching how the ransomware ecosystem can be reconstructed from other perspectives, we come to a more coherent and more accurate definition of the ransomware ecosystem. The ransomware ecosystem concept can then be used to further research how each ecosystem affects the success of ransomware attacks.

Secondly, future research on Conti's internal communication is needed to find novel topics and TTP that Conti uses in their operations. These recommendations are related to the limitations of LDA topic modeling and the limitations of the data used in this thesis. For instance, future research may improve on the methodology in this thesis by applying a classification method such as a support vector machine (SVM) classifier to cluster the messages to the identified topics (Yun and Geum, 2020). Using an SVM classifier, to cluster the messages to topics pairs of users can be determined that discuss similar topics. Consequently, this provides a meaningful set of data that other researchers can use for new studies based on ground truth data. In addition, other text classifier algorithms or an improved methodology using LDA could be applied to the Rocket sub-channel chatlog data to find novel techniques and procedures used by the Conti ransomware gang. Moreover, future research may focus on applying the methodology in this thesis on an improved translation of the chatlog dataset to determine if novel topics come forward that identify novel TTP for the Conti ransomware gang and the ransomware ecosystem. These research efforts may then further strengthen the interpretation of the findings by interpreting topic 9 of the Jabber chatlog data and further interpreting the topics that were labeled with uncertainty. This may be done by validating the most relevant terms in context with multiple expert sessions with attendees from different backgrounds and while using the allocated messages to topics to further validate the labeling.

A few other recommendations can be made for future research. For example, future research should use the ransomware ecosystem concept to study the phenomena of ransomware in its context. The results indicate that ransomware gangs are affected by actors from each of the three sub-ecosystems, and therefore it is a worthwhile contribution to the scientific literature to study ransomware in the ransomware ecosystem. The reconstructed ransomware ecosystem can form a guideline for future scientific work to further study ransomware in its full context. In addition, as the results indicate that the RaaS value chain is in practise different, it is worthwhile to dedicate future research to how the RaaS value chain can be observed for RaaS ransomware gangs and their affiliates. This entails the distribution of RaaS which need further investigation of RaaS affiliates or inner-sphere dark web forums.

8 CONCLUSION

Ransomware has evolved over the years, becoming a greater threat in the cybersecurity threat landscape. To effectively intervene with ransomware, ransomware gangs and their interactions within the ransomware ecosystem should be correctly understood. However, since ransomware gangs put in an effort to keep their communication and operations hidden from the outer world, there is a lack of understanding of how ransomware gangs establish interactions with actors in the ransomware ecosystem. This thesis aims to provide insights into how ransomware gangs internally establish interactions with actors in the ransomware ecosystem and reconstruct the ransomware ecosystem accordingly. To this end, the leaked internal communication data of the Conti ransomware gang is used as a source of ground truth data. A novel methodology is proposed that empirically determines topics in the internal communication of Conti using Latent Dirichlet Allocation (LDA) and maps these to their tactics, techniques, and procedures (TTP). Subsequently, these TTP are leveraged to reconstruct the ransomware ecosystem. As such, we set out to answer the following main research question:

> *To which extent can the ransomware ecosystem be reconstructed using ground truth communication data of the Conti ransomware gang?*

Based on the determined topics and the mapped TTP of Conti, the ransomware ecosystem can be reconstructed from three sub-systems which are the attacker ecosystem, the defender ecosystem, and the governance ecosystem. Ransomware gangs establish interactions within the three sub-ecosystems while operating from the attacker ecosystem. In the attacker ecosystem, ransomware gangs establish interactions with other service-providing cybercriminals to utilize the commoditization of cybercrime by outsourcing sub-commodities of parts of their ransomware value chain. This allows ransomware gangs to strengthen their attack vectors by relying on the expertise of others. Using these services, ransomware gangs can use multiple techniques in parallel, which increases their attack success and resilience. In addition, RaaS ransomware gangs such as Conti perform sophisticated ransomware attacks themselves, and to increase their revenue, they license their formula to affiliates. They establish interactions with these affiliates to support them in performing ransomware attacks following their sophisticated formula.

The defender ecosystem is comprised of defenders that defend themselves against ransomware attacks through multiple actions. These defenders may influence other defenders by helping them strengthen their defense mechanisms. In addition, defenders may harm other defenders in their supply chain through cascading effects or found confidential files or vulnerabilities. Furthermore, defenders may provide generally harmless services to other defenders in the defender ecosystem. Ransomware gangs establish interactions with defenders in the defender ecosystem by performing ransomware attacks and reconnaissance. The reconnaissance in the defender ecosystem can be observed as observing defenders' IT systems, observing how defenders strengthen their defense and observing important information and open-source tools published. These observations are subsequently used to strengthen their ransomware attacks. However, development-based reconnaissance does not apply to affiliates of RaaS ransomware gangs.

The governance ecosystem comprises governance actors that create and maintain the governance framework that influences both the attacker ecosystem and the defender ecosystem. These gov-

ernance actors, for example, regulate defenders through GDPR and by regulating cryptocurrency exchanges. Ransomware gangs establish interactions within the governance ecosystem by leveraging regulatory frameworks such as the GDPR to strengthen their negotiation position. In addition, ransomware gangs observe governance frameworks to strategically dodge strictly regulated territories based on the involved risk, which illustrates how ransomware gangs are aware of regulatory frameworks in the ransomware ecosystem. For example, it becomes evident that ransomware gangs strategically choose loosely regulated cryptocurrency exchanges. By doing so, ransomware gangs adjust their TTP according to the risk involved.

The results in this thesis indicate that Conti is a large and professional organization with such a high level of professionality that it chooses to incorporate services from service-providing cybercriminals in the attacker ecosystem. These services then form sub-commodities to construct the individual parts of their ransomware value chain. Conti choosing for these services over developing these sub-commodities themselves using scarce IT talent illustrates the quality and availability of the provided services in the attacker ecosystem. The quality and availability of these services and the added value to ransomware gangs show how other ransomware gangs are likely to use similar tactics. In addition, from this thesis, it can be concluded that ransomware gangs such as Conti should not be observed as just the developers of ransomware but that they rather work on constructing a ransomware formula that they subsequently license to affiliates. This formula includes sophisticated tactics, techniques, and procedures in which support to affiliates is provided through IT helpdesks and detailed manuals. Hence, RaaS ransomware gangs follow a different RaaS value chain, as often illustrated in the scientific literature.

Reconnaissance is one of the most important activities that Conti and other ransomware gangs perform to get to a successful ransomware attack since it allows them to make informed tactical decisions during ransomware attacks and allows for further development of their attack vectors. That is, based on the information gained by reconnaissance, ransomware gangs can adjust their tactics and configurations of attack vectors. This is facilitated by ransomware gangs incorporating sub-commodity services from service-providing cybercriminals in the attacker ecosystem. This leads to a more varied set of attack vectors and allows them to focus more on reconnaissance. In addition, it became evident how Conti uses their Jabber chat service for conversations on business operations and how the Rocket chat service is used for attack operations. While empirically studying the internal communications in these chat services, several novel TTP are identified, such as Conti using browser injections for reconnaissance of defenders and Siacoin being used to store data centrally. Of these TTP we highlighted Conti's attack chain, reconnaissance procedure, and money laundering procedure. The attack chain identified is more complete than those in previous scientific work and applies to all ransomware gangs. Finally, as the results in this thesis illustrate that each of the three sub-ecosystems of the ransomware ecosystem affects the success of ransomware gangs attacking defenders, ransomware cannot be objectively studied without observing it in the context of the ransomware ecosystem.

This thesis is the first academic effort that provides a description and definition of the ransomware ecosystem based on empirical findings and the first to empirically study the internal operations of a ransomware gang based on ground truth communication data. As such, this thesis contributes to the work of previous scholars by providing empirical findings that substantiate the findings in previous work or creates a more complete and more accurate picture of findings in previous scientific work on ransomware observed from a technical and descriptive perspective. Future work may improve on our research by researching how the identified money laundering procedures and reconnaissance procedures apply to all ransomware gangs in the ransomware ecosystem or by improving on the ransomware ecosystem concept by studying the establishment of interactions of defenders and governance actors. In addition, it may use the ransomware ecosystem concept to further study the phenomena of ransomware in its full context.

BIBLIOGRAPHY

Al-rimy, B. A. S., Maarof, M. A., and Shaid, S. Z. M. (2018). Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. *Computers & Security*, 74:144–166.

Amon (2022). Your all in one card. https://amon.tech/card.

Badhwar, R. (2021). Advanced active directory attacks and prevention. In *The CISO's Next Frontier*, pages 131–144. Springer.

Bateman, J. (2020). *War, Terrorism, and Catastrophe in Cyber Insurance: Understanding and Reforming Exclusions*. Carnegie Endowment for International Peace.

Bátrla, M. and Harašta, J. (2022). 'releasing the hounds?'1 disruption of the ransomware ecosystem through offensive cyber operations. In *2022 14th International Conference on Cyber Conflict: Keep Moving!(CyCon)*, volume 700, pages 93–115. IEEE.

Bayoumy, Y. F. F., Meland, P. H., and Sindre, G. (2018). A netnographic study on the dark net ecosystem for ransomware. In *2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*, pages 1–8. IEEE.

Beaman, C., Barkworth, A., Akande, T. D., Hakak, S., and Khan, M. K. (2021). Ransomware: Recent advances, analysis, challenges and future research directions. *Computers & Security*, 111:102490.

Bezzateev, S., Fomicheva, S., and Zhemelev, G. (2021). Agent-based zerologon vulnerability detection. In *2021 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, pages 1–5. IEEE.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Borisov, N., Goldberg, I., and Brewer, E. (2004). Off-the-record communication, or, why not to use pgp. In *Proceedings of the 2004 ACM workshop on Privacy in the electronic society*, pages 77–84.

Broadhurst, R., Lord, D., Maxim, D., Woodford-Smith, H., Johnston, C., Chung, H. W., Carroll, S., Trivedi, H., and Sabol, B. (2018). Malware trends on 'darknet'crypto-markets: Research review. *Available at SSRN 3226758*.

Caroscio, E., Paul, J., Murray, J., and Bhunia, S. (2022). Analyzing the ransomware attack on dc metropolitan police department by babuk. In *2022 IEEE International Systems Conference (SysCon)*, pages 1–8. IEEE.

Cartwright, A., Cartwright, E., and Xue, L. (2019a). Investing in prevention or paying for recovery-attitudes to cyber risk. In *International Conference on Decision and Game Theory for Security*, pages 135–151. Springer.

Cartwright, E., Hernandez Castro, J., and Cartwright, A. (2019b). To pay or not: game theoretic models of ransomware. *Journal of Cybersecurity*, 5(1):tyz009.

Casino, F., Lykousas, N., Katos, V., and Patsakis, C. (2021). Unearthing malicious campaigns and actors from the blockchain dns ecosystem. *Computer Communications*, 179:217–230.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., and Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 22.

Checkpoint (2022). Leaks of conti ransomware group paint picture of a surprisingly normal tech start-up... sort of. https://research.checkpoint.com/2022/leaks-of-conti-ransomware-group-paint-picture-of-a-surprisingly-normal-tech-start-up-sort-of/.

Chen, K.-Y., Wang, J., and Lang, Y. (2021). Coping with digital extortion: An experimental study of benefit appeals and normative appeals. *Management Science*.

Cisco Talos (2021). Translated: Talos' insights from the recently leaked conti ransomware playbook. https://blog.talosintelligence.com/2021/09/Conti-leak-translation.html.

Coinbase (2022). What is a dex? https://www.coinbase.com/learn/crypto-basics/what-is-a-dex.

Conti, M., Gangwal, A., and Ruj, S. (2018). On the economic significance of ransomware campaigns: A bitcoin transactions perspective. *Computers & Security*, 79:162–189.

Dargahi, T., Dehghantanha, A., Bahrami, P. N., Conti, M., Bianchi, G., and Benedetto, L. (2019). A cyber-kill-chain based taxonomy of crypto-ransomware features. *Journal of Computer Virology and Hacking Techniques*, 15(4):277–305.

Diogenes, Y. and Ozkaya, E. (2019). *Cybersecurity–Attack and defense strategies: Counter modern threats and employ state-of-the-art tools and techniques to protect your organization against cybercriminals*. Packt Publishing Ltd.

Egloff, F. J. and Smeets, M. (2021). Publicly attributing cyber attacks: a framework. *Journal of Strategic Studies*, pages 1–32.

EmerCoin (2022). Community documentation: Emerdns introduction. https://emercoin.com/en/documentation/blockchain-services/emerdns/emerdns-introduction/.

ENISA (2021). *ENISA threat landscape 2021 : April 2020 to mid-July 2021*. European Network and Information Security Agency.

Falcone, R., Hinchliffe, A., and Cooke, Q. (2021). Mespinoza ransomware gang calls victims "partners," attacks with gasket, "magicsocks" tools. https://unit42.paloaltonetworks.com/gasket-and-magicsocks-tools-install-mespinoza-ransomware/.

Fang, R., Xu, M., and Zhao, P. (2020). Should the ransomware be paid? *arXiv preprint arXiv:2010.06700*.

Figueroa, M., Bing, N., and Silvestrini, B. (2022). The conti leaks - insight into a ransomware unicorn. https://www.breachquest.com/conti-leaks-insight-into-a-ransomware-unicorn/.

Galinkin, E. (2021). Winning the ransomware lottery. In *International Conference on Decision and Game Theory for Security*, pages 195–207. Springer.

Goncharov, M. (2015). Criminal hideouts for lease: Bulletproof hosting services. *Forward-Looking Threat Research (FTR) Team, A TrendLabsSM Research Paper*, 28.

Heinl, M. P., Giehl, A., and Graif, L. (2020). Antipatterns regarding the application of cryptographic primitives by the example of ransomware. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*, pages 1–10.

Huang, D. Y., Aliapoulios, M. M., Li, V. G., Invernizzi, L., Bursztein, E., McRoberts, K., Levin, J., Levchenko, K., Snoeren, A. C., and McCoy, D. (2018). Tracking ransomware end-to-end. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 618–631. IEEE.

Ibarra, J., Butt, U. J., Do, A., Jahankhani, H., and Jamal, A. (2019). Ransomware impact to scada systems and its scope to critical infrastructure. In *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, pages 1–12. IEEE.

Ife, C. C., Shen, Y., Murdoch, S. J., and Stringhini, G. (2021). Marked for disruption: tracing the evolution of malware delivery operations targeted for takedown. In *24th International Symposium on Research in Attacks, Intrusions and Defenses*, pages 340–353.

Kaminska, M., Broeders, D., and Cristiano, F. (2021). Limiting viral spread: automated cyber operations and the principles of distinction and discrimination in the grey zone. In *2021 13th International Conference on Cyber Conflict (CyCon)*, pages 59–72. IEEE.

Kao, D.-Y. and Hsiao, S.-C. (2018). The dynamic analysis of wannacry ransomware. In *2018 20th International conference on advanced communication technology (ICACT)*, pages 159–166. IEEE.

Kaptchuk, G., Miers, I., and Green, M. (2017). Managing secrets with consensus networks: Fairness, ransomware and access control. *IACR Cryptol. ePrint Arch.*, 2017:201.

Kaur, S. and Randhawa, S. (2020). Dark web: a web of crimes. *Wireless Personal Communications*, 112(4):2131–2158.

Kenneally, E. (2021). Ransomware: a darwinian opportunity for cyber insurance. In *Kenneally, Erin." Ransomware: A Darwinian Opportunity for Cyber Insurance." Connecticut Insurance Law Journal Fall Symposium Edition*, volume 28.

Keshavarzi, M. and Ghaffary, H. R. (2020). I2ce3: A dedicated and separated attack chain for ransomware offenses as the most infamous cyber extortion. *Computer Science Review*, 36:100233.

Kharraz, A., Robertson, W., Balzarotti, D., Bilge, L., and Kirda, E. (2015). Cutting the gordian knot: A look under the hood of ransomware attacks. In *International conference on detection of intrusions and malware, and vulnerability assessment*, pages 3–24. Springer.

Kigerl, A. (2018). Profiling cybercriminals: Topic model clustering of carding forum member comment histories. *Social Science Computer Review*, 36(5):591–609.

Kim, J., Park, M., Kim, H., Cho, S., and Kang, P. (2019). Insider threat detection based on user behavior modeling and anomaly detection algorithms. *Applied Sciences*, 9(19):4018.

Krebs, B. (2022a). Conti ransomware group diaries, part i: Evasion. https://krebsonsecurity.com/2022/03/conti-ransomware-group-diaries-part-i-evasion/.

Krebs, B. (2022b). Conti ransomware group diaries, part ii: The office. https://krebsonsecurity.com/2022/03/conti-ransomware-group-diaries-part-ii-the-office/.

Kshetri, N. and Voas, J. (2022). Ransomware as a business (raab). *IT Professional*, 24(02):83–87.

Laszka, A., Farhang, S., and Grossklags, J. (2017). On the economics of ransomware. In *International Conference on Decision and Game Theory for Security*, pages 397–417. Springer.

Lee, S., Kim, H. K., and Kim, K. (2019). Ransomware protection using the moving target defense perspective. *Computers & Electrical Engineering*, 78:288–299.

Lemmou, Y. and Souidi, E. M. (2017). An overview on spora ransomware. In *International Symposium on Security in Computing and Communication*, pages 259–275. Springer.

Li, Z. and Liao, Q. (2022). Preventive portfolio against data-selling ransomware—a game theory of encryption and deception. *Computers & Security*, 116:102644.

Lin, X. (2017). Expiro infects, encrypts files to complicate repair. https://www.mcafee.com/blogs/other-blogs/mcafee-labs/expiro-infects-encrypts-files-to-complicate-repair/.

Lombardo, P., Saeli, S., Bisio, F., Bernardi, D., and Massa, D. (2018). Fast flux service network detection via data mining on passive dns traffic. In *International Conference on Information Security*, pages 463–480. Springer.

MacColl, J., Nurse, J. R., and Sullivan, J. (2021). Cyber insurance and the cyber security challenge. *RUSI Occasional Paper*.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., et al. (2018). Applying lda topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3):93–118.

Mansfield-Devine, S. (2010). Battle of the botnets. *Network Security*, 2010(5):4–6.

Maroofi, S., Korczyński, M., Hesselman, C., Ampeau, B., and Duda, A. (2020). Comar: Classification of compromised versus maliciously registered domains. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 607–623. IEEE.

McDonald, G., Papadopoulos, P., Pitropakis, N., Ahmad, J., and Buchanan, W. J. (2022). Ransomware: Analysing the impact on windows active directory domain services. *Sensors*, 22(3):953.

Mei, R., Yan, H.-B., and Han, Z.-H. (2021). Ransomlens: Understanding ransomware via causality analysis on system provenance graph. In *International Conference on Science of Cyber Security*, pages 252–267. Springer.

Mekdad, Y., Bernieri, G., Conti, M., and El Fergougui, A. (2021). The rise of ics malware: A comparative analysis. In *European Symposium on Research in Computer Security*, pages 496–511. Springer.

Meland, P. H., Bayoumy, Y. F. F., and Sindre, G. (2020). The ransomware-as-a-service economy within the darknet. *Computers & Security*, 92:101762.

Morato, D., Berrueta, E., Magaña, E., and Izal, M. (2018). Ransomware early detection by the analysis of file sharing traffic. *Journal of Network and Computer Applications*, 124:14–32.

Mos, M. A. and Chowdhury, M. M. (2020). The growing influence of ransomware. In *2020 IEEE International Conference on Electro Information Technology (EIT)*, pages 643–647. IEEE.

Northwave Security (2022). The complete translation of leaked files related to conti ransomware group. https://github.com/NorthwaveSecurity/complete_translation_leaked_chats_conti_ransomware.

Oosthoek, K., Cable, J., and Smaragdakis, G. (2022). A tale of two markets: Investigating the ransomware payments economy. *arXiv preprint arXiv:2205.05028*.

Orman, H. (2016). Evil offspring-ransomware and crypto technology. *IEEE Internet Computing*, 20(5):89–94.

Pal, R., Huang, Z., Lototsky, S., Yin, X., Liu, M., Crowcroft, J., Sastry, N., De, S., and Nag, B. (2021). Will catastrophic cyber-risk aggregation thrive in the iot age? a cautionary economics tale for (re-) insurers and likes. *ACM Transactions on Management Information Systems (TMIS)*, 12(2):1–36.

Paquet-Clouston, M., Haslhofer, B., and Dupont, B. (2019). Ransomware payments in the bitcoin ecosystem. *Journal of Cybersecurity*, 5(1):tyz003.

Parmar, B. (2012). Protecting against spear-phishing. *Computer Fraud & Security*, 2012(1):8–11.

Popli, N. K. and Girdhar, A. (2019). Behavioural analysis of recent ransomwares and prediction of future attacks by polymorphic and metamorphic ransomware. In *Computational Intelligence: Theories, Applications and Future Directions-Volume II*, pages 65–80. Springer.

Porter, K. (2018). Analyzing the darknetmarkets subreddit for evolutions of tools and trends using lda topic modeling. *Digital Investigation*, 26:S87–S97.

Poudyal, S. and Dasgupta, D. (2020). Ai-powered ransomware detection framework. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1154–1161. IEEE.

Raheem, A., Raheem, R., Chen, T. M., and Alkhayyat, A. (2021). Estimation of ransomware payments in bitcoin ecosystem. In *2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, pages 1667–1674. IEEE.

Rehurek, R. and Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2):2.

Ren, A. L. Y., Liang, C. T., Hyug, I. J., Broh, S. N., and Jhanjhi, N. (2020). A three-level ransomware detection and prevention mechanism. *EAI Endorsed Transactions on Energy Web*, 7(27).

Reshmi, T. (2021). Information security breaches due to ransomware attacks-a systematic literature review. *International Journal of Information Management Data Insights*, 1(2):100013.

Richardson, R. and North, M. M. (2017). Ransomware: Evolution, mitigation and prevention. *International Management Review*, 13(1):10.

Rocket.Chat (2022). Communications platform you can fully trust. https://rocket.chat/.

Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408.

Seawright, J. and Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political research quarterly*, 61(2):294–308.

Severi, G., Meyer, J., Coull, S., and Oprea, A. (2020). Exploring backdoor poisoning attacks against malware classifiers.

Sia (2022). What sia does. https://docs.sia.tech/get-started-with-sia/sia101, journal=Sia 101 - Sia Docs.

Sievert, C. and Shirley, K. (2014). Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70.

Subedi, K. P., Budhathoki, D. R., and Dasgupta, D. (2018). Forensic analysis of ransomware families using static and dynamic analysis. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 180–185. IEEE.

Surjanto, W. and Lim, C. (2020). Finding fast flux traffic in dns haystack. In *International Conference on Critical Information Infrastructures Security*, pages 69–82. Springer.

Tajalizadehkhoob, S., Gañán, C., Noroozian, A., and Eeten, M. v. (2017). The role of hosting providers in fighting command and control infrastructure of financial malware. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 575–586.

Tuttle, H. (2021). Ransomware attackers turn to double extortion. *Risk Management*, 68(2):8–9.

Van Wegberg, R., Klievink, A., and Van Eeten, M. (2017). Discerning novel value chains in financial malware. *European Journal on Criminal Policy and Research*, 23(4):575–594.

Van Wegberg, R., Tajalizadehkhoob, S., Soska, K., Akyazi, U., Ganan, C. H., Klievink, B., Christin, N., and Van Eeten, M. (2018). Plug and prey? measuring the commoditization of cybercrime via online anonymous markets. In *27th USENIX security symposium (USENIX security 18)*, pages 1009–1026.

Waal, A. d., Venter, J., and Barnard, E. (2008). Applying topic modeling to forensic data. In *IFIP International Conference on Digital Forensics*, pages 115–126. Springer.

Wang, K., Huang, C.-Y., Lin, S.-J., and Lin, Y.-D. (2011). A fuzzy pattern-based filtering algorithm for botnet detection. *Computer Networks*, 55(15):3275–3286.

Yan, X., Guo, J., Lan, Y., and Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1445–1456.

Yeboah-Ofori, A. and Islam, S. (2019). Cyber security threat modeling for supply chain organizational environments. *Future internet*, 11(3):63.

Yin, R. K. (2009). *Case study research: Design and methods*, volume 5. sage.

Yun, J. and Geum, Y. (2020). Automated classification of patents: A topic modeling approach. *Computers & Industrial Engineering*, 147:106636.

Yunus, Y. K. B. M. and Ngah, S. B. (2021). Ransomware: stages, detection and evasion. In *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, pages 227–231. IEEE.

Zimba, A. and Chishimba, M. (2019). Understanding the evolution of ransomware: paradigm shifts in attack structures. *International Journal of computer network and information security*, 11(1):26.

A LIST OF STOP-WORDS USED

something, ok, like, right, right, way, even, yes, hi, cool, yet, okay, anything, hey, already, thanks, bro, see, get, fuck, maybe, fucking, fucked, let, everything, would, seems, hell, well, still, hello, yeah, also, guys, well, one, everyone, nothing, two, new, plz, better, much, look, take, think, else, anyone, somewhere, old, cannot, said, know, want, back, ask, go, come, soon, always, enough, say, first, tell, looking, many, second, sorry, oh, without, lot, really, thing, possible, could, damn, great, ones, someone, later, please, sure, another, needs, done, needs, etc, sent, went, found, tried, need, find, away, needed, kind, came, means, gave, pliz, got, going, exactly, necessary, exactly, least, work, write, give, wait, list, make, add, throw, today, create, start, use, change, load, throw, time, give, take, begin, know, day, turn, send, good, ready, today, month, build, run, show, add, try, check, end, happen, hii, minute, drop, check, fix, ehe, three, chill, chat, past, year, set, name, say, want, put, close, main, wrong, fine, normal, talk, clear, read, add, right, shit, understand, leave, know, see, every, brother, remember, look, go, as, dick, think, seem, work, tomorrow, task, somehow, mean, minute, probably, look, likely, see, thought, day, almost, keep, either, use, morning, appear, understood, talk, raise, sit, move, next, hour, apparently, usually, definitely, ago, week, long, saw, must, threw, although, yesterday, thank, ah, listen, become, never, immediately, mine, sometimes, guess, wow, answer, hour, week, include, sec, vayvyvayfv, chvv, fywa, waifa, fall, count, little, receive, agree, decide, call, several, whatever, yo, aaa, feel, may, via, eyjhbgcioijb, anyway, believe, often, whole, since, generally, easier, around, lol, continue, dear, greeting, pwvhdafnrwcx, dsqwv, bhmwy, vyvayva, ehsitrpsvxha, xcjhnhleby, nvzdt, vyva, fyva, wyfy, yfv, yfwa, vayvyvvayfv, different, completely, nice, strange, dude, remain, bring, hang, place, whether, stay, place, remain, attract, miss, fyv, fv, result, eto, ne, ky, ti, po, kak, na, bydesh, ya, tyt, otpishi, man, temt, nm, ghgh, sednya, tut, asd, bl, iva, vap, five, fiv, loris, iva, waifwa, wa, fva, ayf, wyfy, ivyf, lip, elm , phao, psy, cdgljfo, nmth, aa, vmvm, vcbb, clearly, accordingly, suitable, grin, wish, slightly, crap, merry, christmas, boys, chest, desire, joke, near, early, monday, early, heart, jump, frog, shall, feast, besides, constantly, precisely, happy, major, personally, principle, desire, ordinary, eat, non, late, unambiguously,ku, curity, rofl, gather, yell, considers, hye, fell, finally, otherwise, reach, depend, break, play, stand, dmcuywe, mtratadizef, vgmgzo, zolgeo, ephjjmkt, xpacqygfftmw, mqapipvv, nuance, stop, correctly, pxu, sorwfuosmm, mkultktvohtdwfl, real, oi, correctly, remind, hd, entirely, nearest, significant, perceive, specifically, ugh, heard, fb, allow, overall, climb, choose, unfortunately, neither, sometime, occur, yl, afy, ivyfyfv, shipaesh, prinyal, afy, wif, va, dnie, budet, est, slova, tolko, etogo, vozmojnosti, voobhe, mne, bilo, kart, poka, chto, esli, zavtra, otpsihi, bydehs, eshe, ghbdtn, hb, folk, fhjdgen, jiv, hto, dnei, otpihi, davat, bullshit, fva, napisano, negative, twenty, yus, ka, uefi,slat, vt,responsible, faggot, bitch, whore, yep, per, lit, ps, firmly, behalf, sincereley, aside, issuing, terribly, ridonly, thumbsup, anywhere, honestly, jkijigsfjer, albeit, yah, nifiga, tst, tebe, tebya, ay, budu, vot, privet, pozje, kuda, doroge, seychas, ko, skha, sha, nid, whoop, stuff, nn, pp, rw, eyjhbgcioijbm, dt, wj, bi, xx, mm, mi, tm, kq, lk, oooo, wu, qltb, bq, nrxuoz, qujhty, vgdzrqd, lzgpt, ldmhu, edpxxfx, amamxf, xlwdiz, afgky, gipox, ankh, psexesvc, wo, bo, uu, xo, te, jw, ym, di, rq, ai, ala, psomotrim, tween, tak, delo, eti, etot, toje, kidal, tol, priviet, che, ostavlayesh, toka, sednya, vot, chepe, sdohli, vse, shipai, ili, priv, cherez, pryam, zaebis, kak, ti, beresh, obichno, pary, sdelaesh, authnoencap, zuul, kwjqi, sg, td, otletela, eqdc, adcs, hih, lv, lc, uz, jh, vaf, rkn, zero, cunt, mattermost, lpr, cuckoo, dnr, utf, mmm, mmmm, pipet, eq, hpe, asreproas, adv, gpj, han, yf, gbkp, yf, ccskre, crbym, coupe, four, bad, alf, kaif, polnie, zaebok, especially, grinning, latts, bydy, takie, nyanci, esli, otkativayu, terayte, borra, munoz, dde, pizdyuly, ooooh, tenth, regt, blat, stsuko, hz, hiss, narayana, ascus, rsa, duid, dhcpv, vdski, fl, gamarjoba, th, za, bp, zvh, tyr, liter-

ally, eis, rep, shcha, uac, bdehaven, jul, rce, nj, eyjhbgcioijbp, cw, vxg, knew, ac, bd, hh, pw, vp, iaid, hp, dhcpv, jk, jt, qd, km, qa, qn, rch, lj, ey, auf, oshchy, absolute, pyasn, kj, pf, xb, hk, anc, droa, gtym, snu, spr,xc, ls, qweasdzxc, brsib, vpnu, lb, nq, dfbvcdertgh, qi, pv, thfdc, qweasdzxc, speak, nobody, slcc, clr, stl, dachshunds, uo, zt, grateful, nevertheless, kuku, jdem, rr, ehel, vaf, vpa, chil, couple, iitst, nis, cal, dfg, kryptanem, nichego, fx, iiiig, kriptani, harosh, nachnut, otpadat, uspevau, sidim, pol, shalom, shk, oka, dyl, staba, vacatak, odejy, tolko, toka, smotri, palke, davat, mon, vidish, ahhaha, togo, broo, plfhjdf, triple, cho, wkv, zev, dim, mthv, xuhtwjw, odejdy, vaf, cceqv, vseh, kiday, smotri, potom, delai, horosho, deleash, pishi, linki, chasi, zoloto, awk, nc, stuicht, toxtbxnm, col, conv, errno, papa, cxcx, sm, gb, xq, etih, ge, nihao, wmciygc, etu, gcix, acf, fj, eta, zr, novya, utochnit, novih, vx, doljna, adminka,vrode, uje, prosil, uqu, tvteu, mbjnc, ntplwp, onp, idzcly, chcybl, rrrv, bolee, proto, kakay, privot, tyta, bout, pishy, obehaet, jivuchestiu, spameram, nikakih, teryaetsya, paytnicu, kajdiy, menaytsya, dvijeniy, kaktysy, prihod, gonu, oststyk, kaktusu, obidelsya, lts, bydet, seoganuy, bez, brod, vsem, razvel, odnoi, skzal, deeply, teaby, bol, komy, mult, horsy, moya, viplatam, kstate, ondu, himan, sraslos, hjr, rykah, prava, kacahet, proverayt, kriptovaniy, daem, tscl, oo, however, dp, yp, grynya, htf, dostavlen, mnogie, seven, dobavil, chtobi, together, part, spisok, govoril, skoro, soh, tebay, bydhs, mediki, dtsp, ymret, leviy, pakuy, simkoy, kupi, rabochuu, sumki, viezdu, poprobuite, otpdihi, soobjenie, jee, ystraivaet, poraylcay, budesh, dat, test, complete, ee, edi, last, pcie, whoami,zany,unlikely, current,wh, od, suck, mw, gh, bj, hq, cthdfrb, happily, lt, gt, stupid, hmm, yak, bother, ia, ea, gwn, faylomoyku, daite, resilio, proksya, disins, vi, vim, process, file, men, share, hess, half, hint, ad, mt, fe, pcie, namely, yuzak, heh, howklmw, aevt, dptwmb, mrckk, croltiny, conf, boring, yorgar, kilny, concentrada, copfps, hauant, command, ping, karoch, ti, obviously, haha, asshole, laugh, recognize, discuss, forget, love, drink, reason, qg, nada, kwwka, qq, his, kryptani, palit, volvhvb, msie, bigint, included, chelovek, absurd, mem, niks, conveniently, hahaha, peredai, particularly, razbore, bous, prozovn, poryadke, pereskin, morrow, takoe, surreal, ita, gjcvjnh, zvcxdyazrwj, nimi, gotovo, glyanu, otpishus, rozberemsya, palit, popo, viasky,

cf, bl, et, blah, eb, mn, rk, uv, lw, fm, yt, oz, ln, bv, nh, dq, tx, sq, jm, dm, aj, zc, em, ed, eb, ou, cn, ef, efd, ec, dd, bf, bb, da, yc, fa, eh, kv, ebu, kd, ao, tc, vg, yg, sv, far, arrive, night, story, ab, krb, tgs, ef, sp, el, jl, ci, cm, wt, ut, jz, al, mq, fq, yy, mg, ax, ce, af, aif, wafy, aifa, yfva, fif, wafy, fd, df, fc, cc, ca, cbc, wmwjg, ob, rt, oa, gq, gx, sw, fgducw, hzdysto, tsijsra, grp, corp, rms, hy, pj, zp, ek, da,pt, er, ca, ni, tam, ego, vse, sn, md, gde, mp, ot, cd, zg, gk, ri,gg, pg , ju, xm, cq, ev, dh, uj, xym, cj, xf, jx, qv, ih, zm, dx, ij, cu, bw, yvz, ei, jq, bg, omg, ry, fy, nu, qz, fk, tg, apu, dyncheck, lf, unk, iw, iq, otstuk, podgurzka, tj, pe, rf, kl, xv, zo, mlz, tf, gi, yj, mk, es, ph, pe, wb, zb, gw, ccf, lc, hc, ja, jg, ze, qts, hsm, bda, yx, fz, un, zw, lrc, kvd, ss, nego, mqtdb, nzqy, pxu, ntcn, ttl, nas, mnt, fp, lo, jn, cx, li, je, ow, ta, tp, mc, st, si, nd, mx, wi, sf, xd, tb, sa, hm, qy, ly, szp, nv, xr, iy, nx, yv, xw, wf, nf, kr, bmr, ew, ud, rl, ilo, bu, uat, yani, oc, fo, erlang, bdsrv, wbrz,

B SUMMARY OF EXPERT SESSION FIOD

An expert session with members of the cybercrime team of the Fiscal Information and Investigation Service (FIOD) was held. These members have experience in investigating cybercrime and ransomware organizations from a financial perspective. The purpose of this expert session was to validate the labelled topics since the experts present in the session are more competent with technical terms and can help to put the terms into context. Based on the expert session changes were made to the labelling of topics. For the topics coming forward in the Jabber data, the changes made are presented in table B.1 and for the topics forward in the Rocket general data the changes are presented in table B.2. Based on the expert session it became evident that the topics coming forward in the Rocket sub-channels are difficult to label and therefore it is collaboratively decided on that topics 1, 3 and 8 should be labelled as they are and that other topics have too much overlap to be labelled. Finally, it was mentioned that it is worthwhile to investigate if the SharpZeroLogon exploit is still being used after 2020 since this would imply that many servers are not correctly patched.

Table B.1: Changes in Jabber topics based on FIOD expert session

Initial topic	Validated topic
General conversations on specific business units	General conversation regarding business units
Attack vectors	Reconnaissance
Network intrusion?	Development
Ransomware attack operations?	(Purchasing) attack vectors
Anonymous communication?	IT helpdesk
Network block / ARP spoofing + ???	PGP messaging setup
Infrastructure?	Infrastructure configuration

Table B.2: Changes in Rocket-general topics based on FIOD expert session

Initial topic	Validated topic
General conversations on target intrusion	Target selection?
Observing target networks	Malware hosting
Credential forcing?	Credential collection
DNS hijacking?	Conti's cloud
Customer support?	Communication infrastructure?

C LARGER VERSIONS OF COMMUNICATION NETWORKS

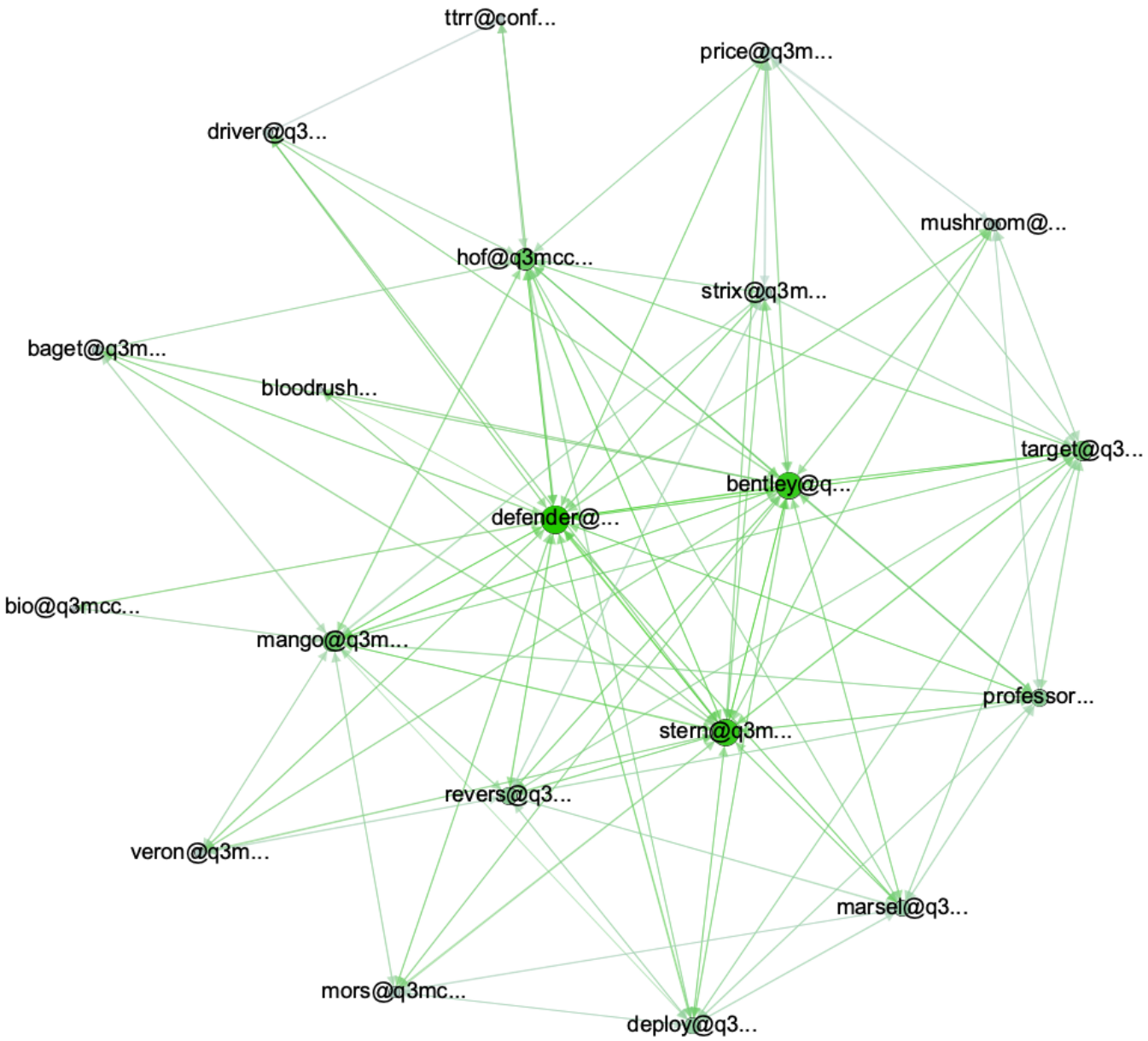

Figure C.1: Larger version of Jabber communication network

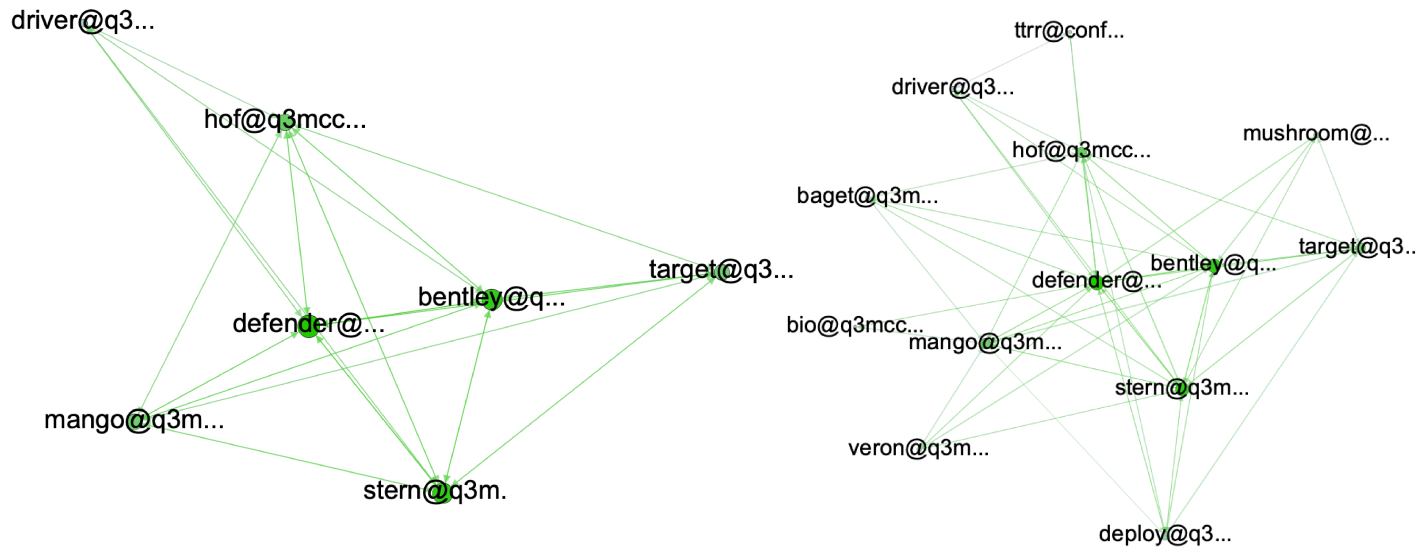

Figure C.2: Larger version of deconstructed Jabber communication network

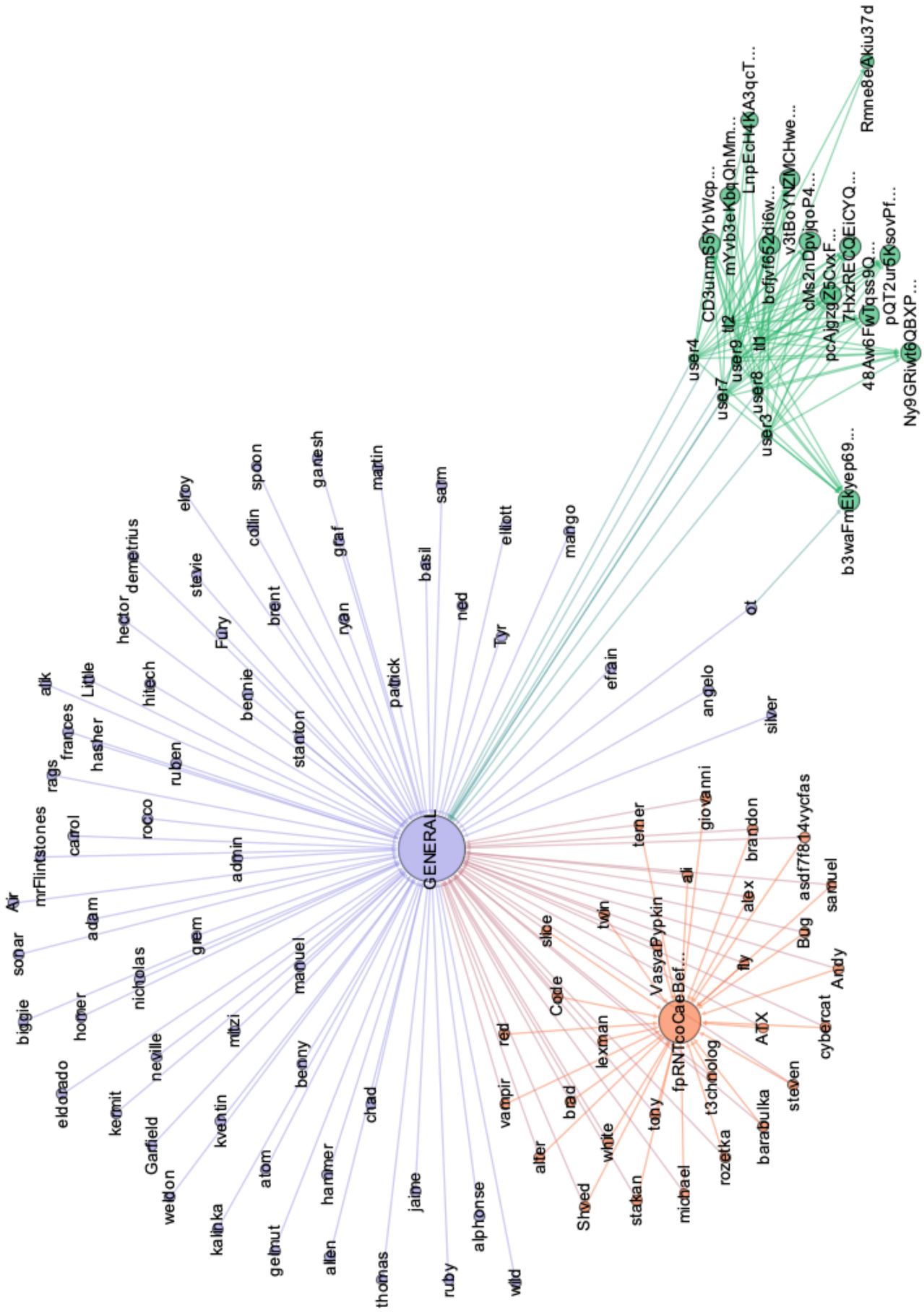

Figure C.3: Larger version of Rocket communication network