Double-Extortion Ransomware: A Study of Cybercriminal Profit, Effort, and Risk

Tom Meurs

This page is intentionally left blank.

Double-Extortion Ransomware: A Study of Cybercriminal Profit, Effort, and Risk

DISSERTATION

to obtain
the degree of doctor at the Universiteit Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Friday 24 January 2025 at 14.45 hours

by

Thomas Willem Arnold Meurs

born on 23 January 1992
in Utrecht, the Netherlands

This dissertation has been approved by:

promotor
prof. dr. M. Junger

co-promotors
dr. A. Abhishta
dr. ir. E. Tews

Type set with LaTeX. Printed by IPSKAMP printing.
Cover design: Esther Beekman (www.estherontwerpt.nl)
ISBN (print): 978-90-365-6417-5
ISBN (digital): 978-90-365-6418-2
DOI: 10.3990/1.9789036564182

Doctorate Board:

Chairman/secretary:	prof. dr. T. Bondarouk
Promotor:	prof. dr. M. Junger
Co-promotors:	dr. A. Abhishta
	dr. ing. E. Tews

Committee Members:
dr. L. Allodi, TU Eindhoven, The Netherlands
prof. dr. E.R. Leukfeldt, Leiden University, The Netherlands
prof. dr. L. Spierdijk, University of Twente, The Netherlands
prof. dr. R.M. Van Rijswijk-Deij, University of Twente, The Netherlands

Referee:
prof. dr. E. Cartwright, De Montfort University, Leicester, United Kingdom

Funding source:

Dutch Police Program Digitization & Cybercrime
Dutch Police Unit East-Netherlands

Contents

List of Tables

List of Figures

Glossary

AIC Akaike Information Criterion.

BIC Bayesian Information Criterion.

C&C server Command and Control server.

CaaS Cybercrime-as-a-Service.

CBS Statistics Netherlands (Centraal Bureau voor de Statistiek).

CI Confidence Interval.

CIA Confidentiality, Integrity, Availability.

CRC Capture-Recapture.

DDoS Distributed Denial of Service.

DoS Denial of Service.

ICT Information and Communications Technology.

IKC Intrusion Kill Chain.

IoT Internet of Things.

IR Incident Response.

IT Information Technology.

LE Law Enforcement.

LLM Large Language Model.

MICE Multiple Imputation by Chained Equations.

MSE Multiple System Estimation.

NAS Network Attached Storage.

RaaS Ransomware-as-a-Service.

RAT Routine Activity Theory.

RBF Replace-By-Fee.

RCM Rational Choice Model.

RCT Rational Choice Theory.

RDP Remote Desktop Protocol.

SCP Situational Crime Prevention.

SME Small and Medium-sized Enterprises.

TOR The Onion Router.

VPN Virtual Private Network.

WTP Willingness To Pay.

If I have seen further, it is by standing on the shoulders of giants

~ Isaac Newton

Acknowledgements

I extend my deepest gratitude to the following individuals whose support has been invaluable throughout my doctoral journey. First and foremost, I am immensely thankful to my supervisor, Marianne Junger. Marianne, your guidance has profoundly shaped my perspective on crime, its motivations, and prevention strategies. Your support during my challenging times as a PhD candidate was invaluable, and for this, I am eternally grateful. Our shared meals with Hans and Sylvia were thoroughly enjoyable.

I am also grateful to my daily supervisors, Abhishta and Erik Tews. Abhishta, your patience and our conversations about coffee and table tennis were delightful, and it was a pleasure meeting Letizia. Erik, your expertise and technical approach were important in my research, and I truly appreciate it.

Raphael, joining the team after two years was a turning point, and I can hardly imagine my PhD journey without you. Our discussions on Teams about complex topics were always enlightening, and your integrity and baking skills are remarkable. I look forward to our paths crossing again.

My appreciation also goes to Edward and Anna Cartwright; our sessions working through difficult game theory puzzles were incredibly stimulating. Harold, your insights into bargaining theory and your exceptional teaching skills have left a lasting impression on me. My sincere thanks to Damon McCoy for your invaluable contributions to the intervention paper and the warm welcome to NYC.

I must acknowledge my colleagues at the police. Emma, thank you for the wonderful start; Cees, your innovative and person-centered approach was inspiring. Mariska and Wolter, thank you for your support throughout my PhD; your growth as leaders has been impressive. Additionally, I extend my thanks to the amazing Cybercrimeteam East-Netherlands: Jørg, Martijn, Rob, Femke,

Sharon, Fatma, Bert-Jan, Bart, Dennis, Johnny, Jodrik, Eric, Peter, Hanneke, Jessica, Marjolijn, Ido, Anouk, Joost, Merijn, Hanko, Roel, and others. You feel like family.

To my intervision group — Johan, Jildau, Gerard, and Marianne—thank you for being there to share both the good and the bad. Thanks also to the Ransomware Taskforce: Bob, Matthijs, Sander, Marcia, Robin, Maurice, Rembrandt, Anouk, Ramon, Erwin, Marije, Daniël, Arie, Max, Gijs, Mark, and many others, for their collaboration and insights, which greatly enriched my research. Gratitude also goes to the team at National High Tech Crime Unit, more specifically the awesome X-RAY and COPS teams.

I am equally thankful to my colleagues with whom I developed the RIB project: Maud, Tim, Joris, Gies, Gijs, and Melissa. Your teamwork and dedication were pivotal in making this project a success. Finally, a big thank you to Anne Jan Oosterheert and Theo van der Plas for enabling this important project. Your support and leadership were crucial in bringing our ideas to fruition.

The cybersecurity community has also been a cornerstone of my experience. The initial meeting where we compared voice recordings eventually led to an amazing project called Melissa. Petra, Liesbeth, Pim, Arie, Gert, Yorick, Rickey, Joeri, Esther, Pascal, Dave, Danja, Lars, Lodi, John, and many others have made this journey remarkable. A special thanks to Northwave for the fruitful collaboration.

I am grateful to the staff at the university— thank you Elke and Gea, for facilitating my PhD journey, and Yasir, with whom I shared memorable moments in Genoa despite the challenging flights. Roland, I am grateful for the insightful discussions we had during my Qualifier.

A special thanks to my favorite TV shows: Temptation Island, Ex on the Beach Double Dutch, Below Deck, the Bachelor, and Married at First Sight Australia, for being my guilty pleasures.

To my family, your unwavering support has been my foundation. Mum, your love is a constant in my life. Dad, thank you for encouraging me to pursue this path and your availability for numerous phone calls. Max and Lisa, your support means the world to me.

Finally, Sylvia, waking up next to you every day is a blessing I cherish deeply. You are my guiding star.

Abstract

The increasing reliance on internet-based technologies has provided cybercriminals with numerous opportunities to exploit vulnerabilities in IT infrastructure. Among the various cyberattacks, double-extortion ransomware has emerged as particularly damaging and lucrative. Double-extortion ransomware involves both encrypting and exfiltrating the victim's data, with the goal of publishing the stolen information if the ransom is not paid.

Despite the growing prevalence of double-extortion ransomware, there is limited empirical research on the decision-making processes of ransomware offenders. In this dissertation we address this gap by applying Rational Choice Theory (RCT), which states that the decision-making processes of ransomware offenders is best understood through evaluating profitability, effort, and risks when conducting double-extortion ransomware attacks. Our main research question is:

> *How do double-extortion ransomware attacks influence profitability, effort, and risk for offenders?*

Our first contribution is the development of a theoretical framework that explains the profit, effort, and risk of double-extortion ransomware with the concept of crime chains (Chapter 2). First, we introduce crime chains to define coordinated attacks and discuss how interconnected malicious activities enhance cybercriminals' operations. Afterwards, we conduct a systematic literature review, to examine the advantages of online crime chains for offenders in terms of profitability, effort, and risks. Finally, we provide a theoretical framework to understand the trade-offs involved in double-extortion ransomware, showing how data exfiltration increases both effort and profits.

Our second contribution is identifying how offender and victim characteristics, as well as contextual variables, influence the profitability of double-extortion

ransomware (Chapters 3–5). First, we combine data from police reports, incident response companies, and leak pages, to provide an empirical estimation of ransomware prevalence in the Netherlands between 2019 and 2022. Our analysis indicates that approximately 60% of ransomware attacks on medium- and large-sized companies go unreported, with an even higher rate for small companies. Afterwards, we use data from the police and incident response companies to examine how various offender and victim characteristics influence ransomware profitability. For instance, larger companies are targeted more frequently due to their ability to pay higher ransoms. Additionally, double-extortion ransomware tends to result in higher ransom payments compared to encryption-only ransomware, although it requires more effort from offenders.

Our third contribution is differentiating ransomware targeting Network Attached Storage (NAS) devices from other types of ransomware in terms of modus operandi, victim characteristics, and timeline (Chapter 6). NAS ransomware, which primarily involves individuals with lower financial resources, typically results in smaller ransom demands. However, these attacks are highly automated and involve fewer stages compared to ransomware attacks targeting businesses. This automation allows offenders to compensate for lower profits per attack by maximizing returns through volume rather than high-value targets.

Our fourth contribution is identifying the incentive for offenders to bluff about data exfiltration during ransomware attacks (Chapters 7 and 8). By analysing a signaling game model, we show how information asymmetry can increase offender profits without much additional effort. Offenders may falsely claim to have exfiltrated sensitive data to inflate ransom demands. Victims, often unaware of whether data has been exfiltrated due to misconfigured or deleted monitoring logs, may be more inclined to pay to prevent the publication of sensitive data, thereby increasing offender profits. However, when offenders themselves are unsure of the value of the stolen data, ransom payments may be mitigated, reducing profits. To conclude, information asymmetry illustrates how offenders may manipulate victims' risk perception to enhance profitability with minimal extra effort.

Our fifth contribution is demonstrating how various law enforcement interventions have led ransomware groups to publish fewer and less significant victims on leak pages (Chapter 9). We evaluate the impact of interventions such as arrests, cryptocurrency freezes, and leak page server takedowns on the profitability, effort, and risks for ransomware offenders. We assess the effectiveness of law enforcement interventions by analyzing the number and type of victims listed on offenders' leak pages before and after interventions. Our findings show a reduction in the number and significance of victims published after interven-

tions. The results indicate that decreasing the profits (e.g., decryptor releases and cryptocurrency freezes), increasing the effort (e.g., leakpage server takedown), or increasing the risk (e.g., sanctions and arrests) could effectively reduce activity of ransomware groups.

In conclusion, this dissertation demonstrates that ransomware offenders' decision-making is best understood by evaluating the elements of RCT: profit, effort, and risk. Additionally, the relationship between these elements offers insights into how interventions can be designed to make ransomware attacks less attractive, ultimately reducing their prevalence. We are confident that the findings in this dissertation will directly help policymakers and law enforcement agencies to combat ransomware more effectively.

Samenvatting

De toenemende afhankelijkheid van internetgebaseerde technologieën heeft cybercriminelen tal van mogelijkheden geboden om kwetsbaarheden in IT-infrastructuur te misbruiken. Onder de verschillende vormen van cyberaanvallen is *double-extortion ransomware* bijzonder schadelijk en lucratief gebleken. Double-extortion ransomware omvat zowel het versleutelen als het exfiltreren van de gegevens van het slachtoffer, met als doel het dreigen met de publicatie van de gestolen informatie als het losgeld niet wordt betaald.

Ondanks de groeiende prevalentie van double-extortion ransomware is er weinig empirisch onderzoek gedaan naar de besluitvormingsprocessen van ransomware-aanvallers. In dit proefschrift onderzoeken we deze besluitvormingsprocessen door gebruik te maken van de *Rational Choice Theory (RCT)*. RCT stelt dat de besluitvorming van ransomware-aanvallers het beste kan worden begrepen door de winst, inspanning en risico's te evalueren bij het uitvoeren van double-extortion ransomware-aanvallen. Onze hoofdvraag is:

Hoe beïnvloeden double-extortion ransomware-aanvallen de winst, inspanning en risico's voor aanvallers?

Onze eerste bijdrage is de ontwikkeling van een theoretisch kader dat de winst, de inspanning en het risico van double-extortion ransomware verklaart met het concept van crime chains (Hoofdstuk 2). Eerst introduceren we crime chains om gecoördineerde aanvallen te definiëren en bespreken we hoe onderling verbonden kwaadaardige activiteiten de operaties van cybercriminelen versterken. Vervolgens voeren we een systematisch literatuuronderzoek uit om de voordelen van online crime chains voor daders te onderzoeken in termen van winstgevendheid, inspanning en risico's. Ten slotte bieden we een theoretisch

kader om de afwegingen bij double-extortion ransomware te begrijpen, waarbij we laten zien hoe data-exfiltratie zowel de inspanning als de winst verhoogt.

Onze tweede bijdrage is het identificeren van hoe kenmerken van daders en slachtoffers, evenals contextuele variabelen, de winstgevendheid van double-extortion ransomware beïnvloeden (Hoofdstukken 3–5). Eerst combineren we gegevens van politierapporten, incident response-bedrijven en *leak pages* om een empirische schatting te geven van de prevalentie van ransomware in Nederland tussen 2019 en 2022. Onze analyse geeft aan dat ongeveer 60% van de ransomware-aanvallen op middelgrote en grote bedrijven niet wordt gerapporteerd, met een nog hogere onderrapportage voor kleine bedrijven. Vervolgens gebruiken we gegevens van de politie en incident response-bedrijven om te onderzoeken hoe verschillende kenmerken van daders en slachtoffers de winstgevendheid van ransomware beïnvloeden. Bijvoorbeeld, grotere bedrijven worden vaker aangevallen vanwege hun vermogen om hogere losgelden te betalen. Bovendien leidt double-extortion ransomware meestal tot hogere losgeldbetalingen in vergelijking met alleen-versleutelingsransomware, hoewel het meer inspanning van de daders vereist.

Onze derde bijdrage is het onderscheiden van ransomware die zich richt op Network Attached Storage (NAS)-apparaten van andere soorten ransomware in termen van modus operandi, slachtofferkenmerken en tijdlijn (Hoofdstuk 6). NAS-ransomware, die voornamelijk individuen met minder financiële middelen betreft, resulteert doorgaans in kleinere losgeldverzoeken. Deze aanvallen zijn echter sterk geautomatiseerd en omvatten minder stappen in vergelijking met ransomware-aanvallen die zich op bedrijven richten. Deze automatisering stelt daders in staat om lagere winsten per aanval te compenseren door de opbrengst te maximaliseren via volume in plaats van zich te richten op waardevolle doelwitten.

Onze vierde bijdrage is het identificeren van de prikkel voor daders om te bluffen over data-exfiltratie tijdens ransomware-aanvallen (Hoofdstukken 7 en 8). Door het analyseren van een signaling game-model laten we zien hoe informatieasymmetrie de winst van daders kan vergroten zonder veel extra inspanning. Daders kunnen ten onrechte beweren gevoelige gegevens te hebben geëxfiltreerd om het losgeld te verhogen. Slachtoffers, die vaak niet weten of er gegevens zijn geëxfiltreerd vanwege verkeerd geconfigureerde of verwijderde bewakingslogboeken, zijn mogelijk eerder geneigd te betalen om de publicatie van gevoelige gegevens te voorkomen, waardoor de winst voor daders toeneemt. Echter, wanneer daders zelf onzeker zijn over de waarde van de gestolen gegevens, kunnen losgeldbetalingen worden verminderd, wat de winsten vermindert. Kortom, informatieasymmetrie illustreert hoe daders de risicoperceptie van

slachtoffers kunnen manipuleren om de winstgevendheid met minimale extra inspanning te vergroten.

Onze vijfde bijdrage is het aantonen van hoe verschillende handhavingsmaatregelen van wetshandhavingsinstanties ertoe hebben geleid dat ransomwaregroepen minder en minder belangrijke slachtoffers publiceren op leak pages (Hoofdstuk 9). We evalueren de impact van interventies zoals arrestaties, het bevriezen van cryptovaluta en het neerhalen van leak page servers op de winstgevendheid, inspanning en risico's voor ransomware-daders. We beoordelen de effectiviteit van handhavingsmaatregelen door het aantal en het type slachtoffers dat op de leak pages van daders staat vóór en na de interventies te analyseren. Onze bevindingen tonen een afname in het aantal en de significantie van slachtoffers die na interventies worden gepubliceerd. De resultaten geven aan dat het verlagen van de winst (bijv. decryptor releases en het bevriezen van cryptovaluta), het verhogen van de inspanning (bijv. neerhalen van leak page servers), of het verhogen van het risico (bijv. sancties en arrestaties) effectief de activiteiten van ransomwaregroepen kan verminderen.

Concluderend laat dit proefschrift zien dat de besluitvorming van aanvallers het best kan worden begrepen door de elementen van RCT te evalueren: winst, inspanning en risico. Bovendien biedt de relatie tussen deze elementen inzichten in hoe interventies kunnen worden ontworpen om ransomware-aanvallen minder aantrekkelijk te maken, wat uiteindelijk de prevalentie vermindert. We zijn ervan overtuigd dat de bevindingen in dit proefschrift beleidsmakers en wetshandhavingsinstanties zullen helpen om ransomware effectiever te bestrijden.

Part I

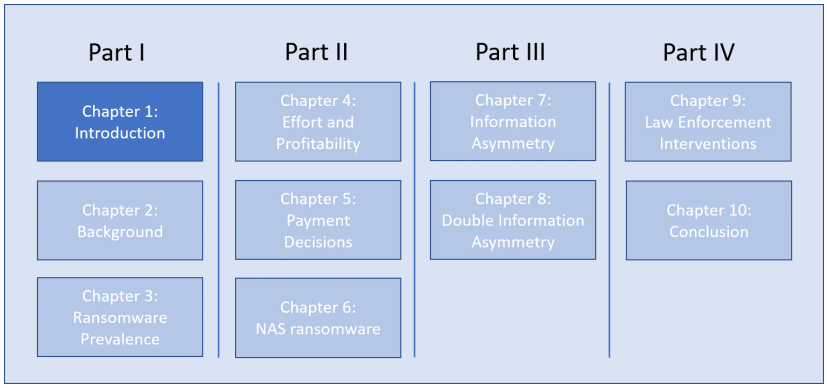
Introduction and Background

Relax - it's worse than you think

~ Case in Point

Chapter 1

Introduction



In this Ph.D. thesis, we focus on double-extortion ransomware. First, we explain that ransomware is a complex phenomenon and an important societal issue. Second, we describe the criminological and economic theories used to examine ransomware: Routine Activities Theory, Rational Choice Theory, Game Theory, and Situational Crime Prevention. Third, we outline the scope of the research, presenting the main research question, identifying research gaps, and formulating the sub-questions. The chapter concludes with an overview of the thesis, highlighting the main contributions of each chapter.

1.1 Scope and motivation

The widespread use of the internet, computers, and IoT devices has increased opportunities for committing various crimes [44]. This may explain the recent decline in offline crime and the simultaneous rise in online crime [61, 11, 75]. Many victims are inadequately protected, and technical IT expertise is easily accessible online. This accessibility could allow individuals, particularly in poorer countries, to target wealthier nations with minimal effort. Additionally, data leaks and accessible infrastructure further facilitate cybercrime.

This dissertation will focus on one of the most complex forms of cybercrime: crypto-ransomware, commonly known as ransomware [63, 40]. Ransomware is a type of malicious software that encrypts a user's data, rendering it inaccessible until a ransom is paid [63]. Typically, the infected device displays a screen with payment instructions, and until the ransom is paid, the files remain inaccessible. Ransomware attacks often involve multiple steps and can be coordinated with other types of cyber-attacks [44]. For example, a ransomware attack might begin with an initial phishing attempt to gain access.

In recent years, ransomware has had an impact on society. For example, in 2017 the ransomware family Wannacry led to the global encryption of million of computers with worm-like capabilities [6]. The ransom was 300 dollars and doubled if not paid in 3 days[6, 35]. IT professionals consider that probably North-Korea was behind the attacks [35]. Additionally, ransomware victims often report losses after an attack. These losses occur in several ways: paying the ransom, facing downtime, and hiring expensive IT consultants to recover files and secure systems against future attacks [4]. Furthermore, the emotional impact of a ransomware attack is often severe [18].

Moreover, ransomware attacks have increasingly targeted companies rather than individuals [19]. Two main factors might contribute to this shift: First, infecting a company might be easier. Enterprises have multiple entry points for infections, while individuals primarily face threats from spam and phishing emails [79, 35]. Common entry points for enterprise infections are network interfaces with the internet, third-party IT solutions, and individual employees through spam and phishing [79]. Second, companies can afford to pay higher ransoms. This financial capability may also explain why offenders have targeted sectors such as healthcare, government institutions, and education, where data is highly valuable and revenues are often substantial [41].

Additionally, ransomware offenders often achieve greater success when they use a coordinated set of attack techniques [66]. For instance, on December 23, 2019, Maastricht University in the Netherlands experienced a ransomware at-

tack after employees clicked on phishing email attachments, giving offenders access to the network [70]. The university paid a ransom of 197,000 euros to regain access to their data. Similarly, Glen Dimplex Home Appliances, attacked in October 2020, paid the ransom only after several Distributed Denial-of-Service (DDoS) attacks [55]. These incidents demonstrate that separate attacks, like DDoS, phishing, or ransomware, can be interconnected parts of a larger coordinated assault. Coordinated cyber-attacks, involving various attack types in a single event, are aggressive and common [78].

A notable type of ransomware that could be considered a coordinated attack is double-extortion ransomware. This involves both data encryption and exfiltration [28, 65]. In this extortion-scheme, offenders encrypt files and threaten to publish exfiltrated data on their blog or leak pages if the ransom is not paid. If negotiations fail, they first list the victim's name on the leak page, followed by publishing the data after a delay. Stolen data may also be sold to other offenders for subsequent attacks [53, 52].

Double-extortion ransomware has become increasingly popular, proving more lucrative than traditional ransomware [16, 28, 53, 65]. Offenders target victims who highly value their data, thereby increasing the attack's impact [36]. Furthermore, in practice, victims are often uncertain about whether data exfiltration has occurred due to missing, misconfigured, or deleted monitoring logs [37]. As a result, victims might pay a premium for data exfiltration even if no data was actually exfiltrated, which increases the profitability of ransomware and therefore worsen the ransomware problem for organisations.

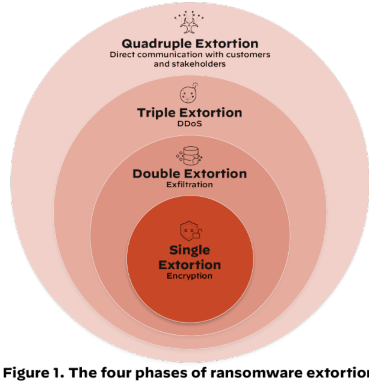


Figure 1.1: The four variants of ransomware extortion [59].

The potential profitability of double-extortion ransomware has led to the development of other multi-extortion schemes. For instance, triple-extortion ransomware involves pressuring the victim with data encryption, data exfiltration, and DDoS attacks [59, 65]. Some offenders further increase pressure by calling customers or employees of the victim's company, a tactic known as quadruple extortion (see Figure 1.1). However, these multi-extortion ransomware schemes seem less common than double-extortion ransomware [65].

In summary, ransomware is a complex crime with an important societal impact. There is a shortage of systematic empirical studies on ransomware offenders' decision-making processes. It is unclear why offenders tend to favor double-extortion ransomware over encryption-only schemes, while more intricate schemes like triple and quadruple extortion are relatively rare. Additionally, information asymmetry may influence the profitability of ransomware for offenders. Large-scale empirical research on ransomware could provide insights into these phenomena.

1.2 Theoretical framework

In this section, we examine criminological and economic theories. Our main focus is on Rational Choice Theory (RCT), as it offers the most direct insight into offender decision-making by analyzing the cost-benefit calculations of ransomware attacks. RCT is complemented with Routine Activity Theory, Game Theory, and Situational Crime Prevention. These theories are interrelated with RCT: Routine Activity Theory (RAT) helps contextualize opportunities for attacks, game theory aligns with RCT by exploring strategic interactions between attackers and victims, and Situational Crime Prevention (SCP) provides strategies for reducing attack opportunities. Each theory provides useful concepts and causal explanations for offender decision-making in conducting ransomware attacks. For an overview, see Figure 1.2.

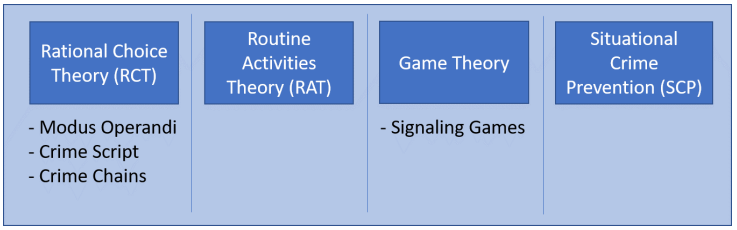


Figure 1.2: Theories and concepts in this dissertation.

The Rational Choice Theory (RCT) of crime states that offenders are rational decision-makers. Crime is purposeful behaviour designed to meet the offender's commonplace needs for such things as money, status, sex, and excitement. Offenders are reasoning actors who weigh means and ends, costs and benefits, and make a rough rational choice for the course of action that seems to yield the most benefits [20, 22]. In contrast to a more economical view of rational choice [72], offenders' choices are not long-term wise decisions but can best be described as bounded rationality: their decisions are highly constrained by factors such as offenders' abilities and the availability of relevant information [21]. The Rational Choice Theory (RCT) of crime focuses on rational decisions and modus operandi to attain offender's short-term goals.

Research supports the rational choice model of crime, for offline crime [12, 76] and online crime [2, 77]. Most relevant, experiments show that policy measures that influence the costs and benefits of crime, by increasing the effort and the risks, and decreasing the potential benefits, generally prevent crime offline [12] and online [8].

In the context of ransomware, RCT suggests that offenders might be more inclined to conduct double-extortion attacks if the average additional profits from data exfiltration outweigh the (average) additional efforts involved. This trade-off is based on the assumption that the actors carry out multiple ransomware attacks and not every double-extortion attack create extra profits compared to encryption-only attacks. Another implication of RCT is that offenders exploit favorable opportunities to gain extra profits. For example, large companies are more likely to afford higher ransom payments than small companies, incentivizing offenders to demand larger ransoms from large companies compared to small ones.

RCT helps explaining the offender's decision-making, but it is equally important to consider the choices made at each stage of a crime. The modus operandi, or crime script, outlines these stages and illustrates how offenders evaluate rewards, effort, and risks throughout the offense [57]. Understanding the crime script for a specific crime is essential, as it reveals the key decisions made at each step.

Crime scripting, a method from crime science, systematically studies these stages, and has been applied to simulated cyber-attacks [67] and online stolen market data [42]. Furthermore, crime scripting helps explain how offenders adapt their strategies depending on the situation. In double-extortion ransomware, for example, offenders may choose to encrypt only the most valuable parts of a network to maximize profits while minimizing effort and risks.

For ransomware, many crime scripts have been developed [46, 35, 60]. These scripts typically include stages such as development, Ransomware-as-a-Service (RaaS) and collaboration, access, lateral movement, data exfiltration, encryption, communication, payment, blackmail, and cash-out. Each stage has specific goals, and understanding these stages helps in identifying points where different actions might be coordinated to achieve the overall objective of the attack. For instance, the goal of the access stage in a ransomware attack might be achieved through a phishing attack [35, 71].

We summarize the steps based on previous literature as follows (see Figure 1.3) [46, 35, 60]:

1. **Development:** Initially, the offender needs to set up the necessary infrastructure and develop the malware. This infrastructure supports the delivery of the malware and helps conceal network traces from the victim's system back to the offender.

2. **RaaS (Ransomware-as-a-Service) and Collaboration:** Individuals or groups lacking expertise may utilize Ransomware-as-a-Service, which involves renting ransomware from other offenders. This service enables affiliates with minimal technical skills to launch sophisticated ransomware attacks, potentially simplifying the execution process and allowing specialization in different aspects of the attack.

3. **Access:** Gaining and maintaining access to a victim's computer or network is achieved by distributing the ransomware through various means such as phishing emails, malicious apps, exploit kits, or vulnerabilities in the victim's systems.

4. **Lateral Movement:** This involves moving across the network to assess files and gain control over the entire network, aiming to maximize the impact of the attack. Furthermore, it is important for the offender to be able to establish persistence, maintaining a foothold in the victim's network, even if they change their password or shutdown their computer.

5. **Data Exfiltration:** Offenders need to find, collect, package, and exfiltrate sensitive data. Subsequently, the offenders need to have infrastructure to host the victim's data and publish it on a leak page if the victim is not willing to pay.

6. **Encryption:** The core of the ransomware attack involves encrypting the victim's files, making them inaccessible without a decryption key.

Figure 1.3: The crime script of ransomware [46, 35, 60].

7. **Communication:** Offenders need to establish communication with the victim to negotiate the ransom, often starting with a ransom note that could be personalized based on the victim's characteristics like annual revenue. The negotiations can be done by e-mail, although some of the larger ransomware groups communicate through a chat platform hosted on a darknet page.

8. **Payment:** At this stage, victims must decide whether to pay the ransom or rely on other means to restore their files, like backups. Negotiations with the offenders might involve determining a ransom amount, potentially with the help of an incident response company.

9. **Blackmail:** Offenders may employ additional extortion tactics like DDoS attacks or contacting the victim's clients or employees to increase pressure on the victim.

10. **Cash-out:** Finally, the ransom payment is processed, typically involving money laundering through mixers or money mules, after which the decryption keys are provided to the victim to restore access to their files.

Double-extortion ransomware adds complexity to the encryption-only ransomware crime script, influencing the offender's approach and victim's response, specifically in step 5 of the previously elaborated crime script. During this step, offenders must set up a robust infrastructure to exfiltrate the data. During the attack, the offenders have to find, collect, package, and exfiltrate sensitive data. Finally, they have to find a way to publish the data if the victim does not cooperate. Subsequently, the offenders have to display a file directory of the exfiltrated data to the victims to prove data is exfiltrated. This action allows victims to assess the value of the stolen data, potentially affecting the ransom amount they are willing to pay. Taken together, data exfiltration during a ransomware attack requires a lot of effort in addition to the effort required for an encryption-only ransomware attack.

From the victims' perspective, the threat of publication of data in addition to the encryption of their data extends the attack beyond the unavailability of data (availability aspect of the CIA triad) to include the potential public disclosure of confidential information (confidentiality aspect of the CIA triad). The CIA triad is an information security framework used to identify which aspect of data is compromised in a cyberattack: confidentiality, integrity, or availability. Given that double-extortion ransomware affects two dimensions of the CIA triad (confidentiality and availability), whereas encryption targets only one, it seems logical that double-extortion ransomware often increases the willingness to pay a ransom compared to attacks that only involve data encryption [53].

The ransomware crime script illustrates the different steps during a ransomware attack. Since these steps might involve other types of offenses, it also illustrates another concept: crime chains. Crime chains are series of different types of offenses that occur together or in a certain order and are related to each other to perform a coordinated set of actions [27, 44]. Crime chains are a generalization of the concept of coordinated attacks mentioned in the previous section. Offenders behind crime chains could deliberately coordinate their own crimes, but they could also respond opportunistically to crimes of other offenders. According to the previously mentioned RCM [23], these benefits should enhance profit or reduce costs, risks, and effort. Evidence shows that crime chains occur offline [27, 5]. Despite limited focus in research, several studies indicate that (long) crime chains also occur online [47, 78] and are more severe than small crime chains or isolated attacks [66].

Figure 1.4: Illustration of the Crime Chain Concept.

Various mechanisms enable a crime chain (see Figure 1.4) [27]:

1. **Necessity.** One crime requires another. For example, most offenders have a strong incentive to cover up the first offense, and that might require committing a second one. A burglary starts out nonviolent, but the offender has a reason to assault someone who discovers him or threatens to turn him in. In ransomware attacks, after initially encrypting data, offenders may find it necessary to use proxies to stay anonymous in order to avoid attention from law enforcement. To gain these proxies that might exploit other computer systems, thereby committing another cybercrime to remain anonymous [56].

2. **Disinhibition.** One crime disinhibits another. Illegal substances often disinhibit people, after which they might commit crimes they would not have committed otherwise. In the context of ransomware, this phenomenon can be seen when offenders gain confidence due to successful high-profile attacks and a perceived lack of law enforcement activity. This confidence might lead to additional attacks or the use of more aggressive ransomware tactics, thus escalating their malicious activities. An example of overconfident ransomware actors is the regularly give interviews, although they know they should stay anonymous [45].

3. **Advertisement.** One crime advertises another. Drug corners and drug markets are known in the community and might attract other types of crime. When ransomware offenders successfully extort a high-profile company and publicize this on their leak sites, it serves as an advertisement to other potential offenders that the victim is vulnerable. This might lead to the victim facing new attacks, leading to repeated victimization [24]. Thus, the initial crime 'advertises' the possibility of additional crimes.

4. **Enticement.** One crime entices another. There is an overlap between offender and victim populations. Offenders spend more time in risky situations and therefore might end up victims themselves. Offenders involved in ransomware groups might fall victim to scamming. A common scam is known as the 'exit scam,' where leaders of a ransomware group falsely claim they are being arrested and must cease their online activities [43]. In reality, they are not being arrested; instead, they use this as an excuse to disappear with the ransom profits, denying other group members their share. This scam illustrates how ransomware offenders can become victims themselves.

5. Setup. One crime sets up for another. A burglar enters a house, thinking only about the loot. But he finds a young lady there alone and, with no prior planning, rapes her [50]. This is an example of one crime setting up another, despite the offender's lack of initial intent. Similarly, in a ransomware context, an offender might initiate a ransomware attack intending to encrypt the data. However, upon discovering highly sensitive data within the victim's systems, the offender may decide to escalate the attack by exfiltrating and selling this data to competing companies or other offenders [28].

6. Escalation. One crime escalates into another. Some disputes involve a cycle of revenge, where a violent crime can lead to retaliation and perhaps escalation. Even an innocent victim might take the law into his own hands and retaliate at another time. In some cases, if a ransomware victim refuses to pay and publicly denounces the offenders, it can lead to an escalation. The offenders might respond by initiating denial-of-service attacks or releasing a portion of the stolen data to pressure the victim into paying [52].

7. Victimization Cycle. One crime can start a victim chain. One of the more subtle facts about crime is when a victim becomes an offender. For example, Jan van Dijk found that bicycle theft in the Netherlands might lead victims to steal another bike, essentially becoming an offender [27]. In a ransomware context, offenders might be scammed by other offenders, and to save money, scam other offenders as well [63]. This chain of offenders scamming each other might become a victimization cycle.

8. One co-offender attacks another. Many crimes are committed by groups of co-offenders who may later conflict over how to divide their loot. Ransomware operations often involve various participants, including developers, affiliates, initial access brokers, and money launderers. Ransomware developers create the malware, affiliates purchase and use it to execute attacks, initial access brokers sell access to victims, often to affiliates, and money launderers are specialized actors hired to launder the ransom after payment. Disputes can arise over the distribution of ransom payments or geopolitical affiliations [33]. These conflicts can lead to the split-up of a ransomware group or the leaking of sensitive information to authorities or competitors, like internal communication chats [33, 30].

These mechanisms show how crimes can lead to a series of other crimes, creating a crime chain. To better understand double-extortion ransomware, it

is helpful to explore crime chains, as they highlight the extra effort and higher rewards involved in combining data encryption and exfiltration, compared to only encrypting data. The concept of crime chains suggests that criminal activities tend to multiply, so that the disruption or removal of a crime might set in motion a chain reaction of crime reduction [27]. Therefore, understanding double-extortion ransomware through crime chains might help develop effective strategies to prevent ransomware.

RCT implies that people commit crime due to favourable situations or opportunities [26]. According to Routine Activity Theory (RAT), these opportunities arise as a result of the convergence of time and space where there is a suitable target, a motivated offender, and the absence of a capable guardian [15, 58]. [51] discuss the possibilities and challenges of applying RAT to cybercrime. They review both theoretical reflections and empirical data-sets for the use of different elements of RAT. Their conclusion is that it is unclear whether RAT could be used as an analytical framework for cybercrime. [31] nevertheless concluded that RAT provides enough explanatory power for cybercrimes. One example of RAT in the online environment is the online presence of risky places. In the context of ransomware, RAT suggests that companies heavily reliant on IT infrastructure, such as managed service providers, are more likely to be targeted by ransomware attacks compared to companies with minimal IT infrastructure, such as those in the leisure sector.

Another approach to studying double-extortion ransomware is through game theory. This framework is particularly relevant because it analyzes the strategic decisions of the involved actors, such as whether the victim should pay the ransom [10, 48, 29]. The features of the game are well-defined, with clear roles for the offender and the victim, and decision options and payoffs that are primarily monetary. By applying a game-theoretic model, we can determine the existence of a stable equilibrium and identify potential interventions to shift this equilibrium in ways that could enhance social welfare.

Research applying game-theoretic frameworks on double-extortion ransomware align with previously mentioned empirical research: Combining data encryption with exfiltration leads to higher profits for offenders compared to encryption alone [52, 49, 54, 52]. However, research on the profitability of double-extortion ransomware using game theory often ignores an important aspect: victims' uncertainty about whether data has been exfiltrated. This information asymmetry can be effectively modeled using signaling games [62].

Signaling games are widely used in economics and evolutionary biology to analyze situations where one party has more information than the other and needs to signal a desirable attribute. For example, job seekers signal productivity

through education or grades [73], and donors signal generosity through public donations [32]. In our context, offenders might signal data exfiltration to maximize ransom payments.

Signaling game analysis has shown that costly signaling can force actors to incur high costs to prove the desirable attribute. For example, university education as a costly signal of ability [7]. [1] demonstrated how information asymmetry between buyers and sellers of used cars can lead to market failure, which can also apply to our context [49]. This framework suggests that data exfiltration may not always benefit criminals. In this dissertation, we apply signaling games to understand the decision-making of offenders and victims in the context of double-extortion ransomware and information asymmetry.

Finally, we examine Situational Crime Prevention (SCP), a criminological approach that is based on RCT and states that crime could be prevented by changing situations [15, 25, 26]. More specifically, effective interventions should alter the cost-benefit trade-off of these circumstances [22]. Five general strategies guide SCP: Increase the Effort, Increase the Risks, Reduce the Rewards, Reduce Provocations, and Remove Excuses [13, 14, 38, 34]. These strategies aim to deter potential offenders by making crimes more difficult or less appealing. The effectiveness of SCP strategies in combating various crimes has been studied [9, 39, 38]. Studies evaluating SCP measures against cybercrime are scarce [38]. In this dissertation we study offenders' decision-making processes during ransomware attacks and develop SCP strategies for law enforcement and policymakers.

The theories discussed in this section were chosen for their ability to explain offender decision-making in ransomware attacks, all based on the assumption that ransomware actors make rational trade-offs between costs and benefits. Rational Choice Theory (RCT) serves as the central framework, analyzing how attackers balance profits, effort, and risks. Routine Activity Theory (RAT) highlights the role of opportunity, showing how offenders exploit vulnerabilities. Game theory complements RCT by examining strategic interactions, particularly during negotiations. Finally, Situational Crime Prevention (SCP) explores how altering conditions can reduce attack opportunities, offering practical preventive measures.

1.3 Research scope

The central theme of this dissertation is double-extortion ransomware, where attackers encrypt victims' files and exfiltrate sensitive data, demanding a ransom for both the decryption key and a promise not to release the data. This research focuses exclusively on encryption-based ransomware, excluding other

forms like locker ransomware, wiper malware, and digital extortion (e.g., account takeovers) [35, 67]. By narrowing the scope to file-encrypting ransomware, we aim to provide a deeper understanding of the decision-making processes involved in these specific attacks.

As described in the previous section, the theoretical foundation for this dissertation is mostly based on theories from crime science, such as Rational Choice Theory, Routine Activities Theory, and crime scripting [23, 34]. These theories have in common that they consider crime as a result of opportunity arising from a favorable situation. Situational-based theories contrast with crime science theories that focus more on the background and personal circumstances of offenders [34]. In this dissertation, we will not focus on the background, childhood, psychological well-being, and personal circumstances of ransomware offenders. In line with RCT, we view offenders as rational actors who make decisions based on evaluating the trade-offs between the potential profits, the efforts required, and the risks involved in an attack, as suggested by [23].

RCT's assumption that offenders are rational actors seeking to maximize profits while minimizing risks and effort, aligns well with the assumption of rational actors in economic models. Therefore, we use game theory to quantify offender decision-making. More specifically, we focus on signaling games to analyze the strategic interactions between ransomware attackers and victims during ransom negotiations. By using signaling games, we can explore how information asymmetry regarding the data exfiltration influences ransomware negotiations and profitability.

Another concept which aligns well with RCT is crime chains. Crime chains allow us to examine how the modus operandi of double-extortion ransomware might influence the RCT elements (profit, effort, and risk). Chapter 2 offers a broad overview of various crime chains, including opportunistic and coordinated crime chains. In subsequent chapters, the focus narrows to double-extortion ransomware crime chains, which involve multiple actors, such as data account managers who are responsible for the exfiltrated data during double-extortion ransomware attacks. A collaboration between malicious actors to conduct a double-extortion ransomware attack might create a sequence of linked criminal activities, from initial access to the final ransom demand. In this dissertation, the concept of crime chains helps explain the emergence of different types of ransomware, such as double-extortion ransomware, encryption-only ransomware, and ransomware targeting NAS devices, based on their respective profits, effort, and risks.

Finally, this dissertation adopts an interdisciplinary approach, integrating criminological and economic theories with computer science to account for the

technical elements of ransomware attacks. This approach enables a more comprehensive analysis of the choices offenders make during ransomware attacks, such as whether to put in additional effort to exfiltrate data and how this impacts both the offenders and the victims.

A potential complication of adopting an interdisciplinary approach is that definitions and terminology may vary across different scientific disciplines. For example, RCT links effort and profit for offenders, described in computer science literature as a "work-averse criminal" [2, 3]. In Information Security, Intrusion Kill Chains (IKCs) detail attack steps, similar to "crime scripts" in criminology. [67] concludes that crime scripts and IKCs are essentially the same, outlining the stages of an attack. Throughout this dissertation, terms like "offender," "actor," and "criminal" will be used interchangeably depending on the context of the discussion and the target audience of the papers.

1.4 Research questions

As discussed in the previous sections, double-extortion ransomware has an impact on society, yet there is a lack of empirical research examining the decision-making of ransomware offenders. Drawing on Rational Choice Theory (RCT), we formulate our main research question as follows:

> **Main Research Question:** *How do double-extortion ransomware attacks influence profitability, effort, and risks for offenders?*

To address this main question, it is essential to consider gaps in the existing literature. Identifying these gaps allows us to develop relevant subquestions that guide our investigation.

Research Gap 1: Limited Research on Online Crime Chains. As described previously, the concept of crime chains helps understanding the sequence and connection between various malicious activities. However, crime chains are seldom discussed in the context of online crime [44]. We pose the following research question:

> **RQ1:** *In what ways do crime chains affect the profitability, effort, and risks of attacks?*

To address this question, we conduct a systematic literature review, collecting studies from both computer science and criminology journals. We then develop a theoretical framework comparing the profitability, effort, and risks associated

with different crime chains. This framework helps us understand the trade-offs between these factors in different types of ransomware, like double-extortion ransomware, encryption-only ransomware, and ransomware targeting NAS devices.

Research Gap 2: Lack of Empirical Research on the Influence of Offender and Victim Characteristics on Ransomware Profitability. There is a lack of systematic, empirical research on ransomware profitability [17]. Most existing studies focus on blockchain analysis of ransom payments [64], but this approach does not account for how offender and victim characteristics influence ransom demands, payments, and financial losses. Access to victim data would provide more detailed insights, but due to its sensitivity, it is difficult to obtain. However, through a special collaboration with the Dutch Police, we gained access to ransomware attacks reported by victims to the police between 2019 and 2023. The police data allowed us to examine the profitability of double-extortion ransomware based on offender and victim characteristics. To address potential underreporting to the police, additional data was obtained from an incident response company [74]. We pose the following research question:

> **RQ2:** *How does combining data encryption with data exfiltration alter the profitability of ransomware attacks?*

First, we estimate the potential biases in the datasets by applying multiple system estimation to calculate the number of unobserved ransomware attacks in the Netherlands between 2019 and 2022. The estimation is based on considering the intersection and disjunction between victims in police data, incident response data, and leak pages. These estimates help us better interpret the analyses focusing on ransomware profitability in subsequent chapters and provide a clearer understanding of the scope of ransomware attacks in the Netherlands.

Afterward, we analyze the ransomware cases reported to both the Dutch Police and incident response companies, focusing on the relationship between offender and victim characteristics and metrics associated with ransomware profitability, such as ransom demands, payments, and financial losses.

Research Gap 3: Limited Research on Ransomware Targeting NAS Devices. To our knowledge, there is a lack of research specifically focusing on ransomware attacks targeting Network Attached Storage (NAS) devices, defined as NAS ransomware. NAS devices are frequently used by individuals rather than companies [68]. Individuals are less likely to hire incident response companies

for recovery after a ransomware attack due to limited financial resources, making them more inclined to report incidents to the police. This makes our police dataset particularly valuable for analyzing NAS ransomware, which would be difficult to study using cybersecurity company data alone. The police dataset also enables a comparison between NAS ransomware and other types of ransomware.

Additionally, comparing the modus operandi and profitability of NAS ransomware to other forms of ransomware may provide insights into how offenders trade off the key RCT elements: profits, effort, and risks. Since individuals generally have fewer financial resources than companies, the potential profits for offenders are lower. According to Rational Choice Theory (RCT), this reduced profitability influences the effort offenders are willing to invest, resulting in a distinct modus operandi for NAS-targeted ransomware compared to regular ransomware attacks. We pose the following research question:

> **RQ3:** *How does the profitability of ransomware attacks targeting NAS devices compare to the profitability of regular ransomware attacks?*

To address this question, we examine ransomware attacks reported to the Dutch Police, of which a subset of attacks was targeting NAS devices.

Research Gap 4: Limited Application of Game-Theoretic Frameworks to Model Information Asymmetry in Ransomware Attacks. As described in the previous section, some literature addresses the use of game-theoretic models to understand the decision-making of offenders and victims in ransomware attacks [54, 10, 69]. These studies primarily focus on comparing the profits of double-extortion ransomware with encryption-only ransomware. However, these models overlook the critical role of information asymmetry during ransomware negotiations as observed in our police data. When victims are uncertain whether their data has been exfiltrated, it affects both the ransom they are willing to pay and the overall profitability of double-extortion ransomware. Incorporating information asymmetry into these models would provide a better understanding of double-extortion ransomware profitability.

We address this gap by exploring a signaling game to model information asymmetry during ransomware attacks. Signaling games are particularly well-suited for capturing the strategic interactions between offenders and victims when one party has more information than the other. Moreover, the assumptions of signaling games align closely with Rational Choice Theory (RCT), making our analysis directly relevant to the RCT elements -profits, effort, and risk— that are central to this thesis. We pose the following research question:

> **RQ4:** *How does information asymmetry regarding data exfiltration affect the dynamics between offenders and victims in double-extortion ransomware attacks?*

To address this question, we develop and apply a signaling game to model the situation where offenders know whether data is exfiltrated or not and whether victims know the importance of the data if it is exfiltrated. We focus on factors such as ransom demands, ransom payments, and the role of information asymmetry.

Research Gap 5: Limited Research on the Evaluation of Law Enforcement Interventions Against Ransomware. There is a lack of studies evaluating the effectiveness of law enforcement interventions against ransomware [38]. Understanding the impact of various strategies is crucial for developing effective responses to the evolving tactics of cyber offenders. Situational Crime Prevention (SCP) provides a framework for evaluating interventions that reduce profits, increase effort, or raise risks for offenders, potentially decreasing ransomware attacks.

Our research highlights the potential of using leak pages to assess the impact of interventions. By analyzing the number and type of victims published by ransomware groups before and after an intervention, we can measure the effectiveness of law enforcement interventions.

> **RQ5:** *What intervention strategies could law enforcement employ to combat double-extortion ransomware attacks?*

To address this question, we examine interventions aimed at reducing profits, increasing effort, and raising risks for offenders—such as arrests, decryptor releases, crypto-asset freezes, leak page takedowns, and sanctions. We assess their effectiveness by comparing the published victims of ransomware groups before and after these interventions.

1.5 Thesis outline and contributions

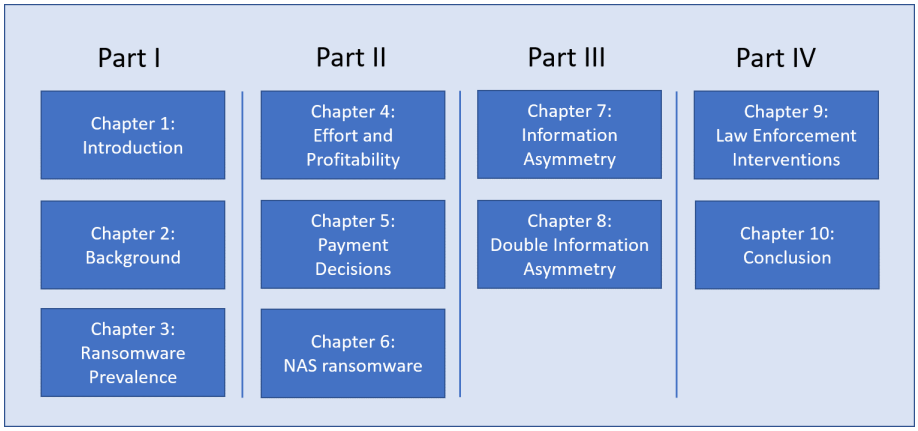


Figure 1.5: Outline of this dissertation.

Figure 1.5 shows the outline of this dissertation.

In **Part I**, after the introduction we start with the background and context for crime chains in Chapter 2. These will prove helpful understanding the profitability, risks, and effort of double-extortion ransomware, thus addressing RQ1. The chapter is based on:

> Meurs, T., Junger, M., Abhishta, A., Tews, E., & Ratia, E. (2022). COORDINATE: A model to analyse the benefits and costs of coordinating cybercrime. *Journal of Internet Services and Information Security*, 12(4).

In Chapter 3, we analyze the prevalence of ransomware attacks in the Netherlands from 2019 to 2022, using three datasets: police reports, incident response reports, and leak pages. By applying multiple system estimation to these combined datasets, we estimate the number of unobserved ransomware attacks. Additionally, this chapter allows us to assess the validity of these datasets for use in the subsequent chapters. This chapter is based on the following study:

> Meurs, T., Junger, M., Cruyff, M., & van der Heijden, P. G. M. (2024). Estimating the Number of Unobserved Ransomware Attacks. *Available at SSRN 4942706*.

In **Part II**, we will provide the payment decisions of victims, which indirectly describes the profitability of double-extortion ransomware for offenders. We start by describing the crime script, considering payment decisions and financial losses reported by victims. More specifically:

In Chapter 4, we analyze the dynamics between offenders and victims in ransomware scenarios, examining ransom requests, payment dynamics, and financial losses reported to Dutch law enforcement to address RQ2. The chapter is based on:

> Meurs, T., Junger, M., Tews, E., & Abhishta, A. (2022, November). Ransomware: How attacker's effort, victim characteristics and context influence ransom requested, payment and financial loss. In *2022 APWG Symposium on Electronic Crime Research (eCrime)* (pp. 1-13). IEEE.

In Chapter 5, we extend Chapter 4 by incorporating an extended dataset. We will use a hurdle model to simultaneously estimate both the decision to pay a ransom and the amount to be paid, further addressing RQ2. The chapter is based on:

> Meurs, T., Cartwright, E., Cartwright, A., Junger, M., Hoheisel, R., Tews, E., & Abhishta, A. (2023). Ransomware economics: a two-step approach to model ransom paid. In *18th Symposium on Electronic Crime Research, eCrime 2023*.

In Chapter 6, we examine a specific type of ransomware attack: those targeting network-attached storage (NAS) devices. NAS devices are typically used by individuals rather than companies, which alters the interaction between offenders and victims. This analysis provides insights into the offender-victim dynamic in a business context as well. The findings contribute to RQ3. This chapter is a translated and extended version of the following study:

> Meurs, T., Junger, M., Tews, E., & Abhishta, A. (2022). NAS-ransomware: hoe ransomware-aanvallen tegen NAS-apparaten verschillen van reguliere ransomware-aanvallen. *Tijdschrift voor Veiligheid*, 21(3-4), 69-88.

In **Part III**, we will apply a game theoretical framework to study the interaction between victim and offender during double-extortion ransomware. More specifically:

In Chapter 7, we will consider an information asymmetry where offenders know whether data is exfiltrated or not, but the victim does not, to address RQ4. We model this information asymmetry with a signaling game. The chapter is based on:

> Meurs, T., Cartwright, E., Cartwright, A., Junger, M., & Abhishta, A. (2024). Deception in double-extortion ransomware attacks: An analysis of profitability and credibility. *Computers & Security*, 138, 103670.

In Chapter 8, we extend the model from Chapter 7, but now also consider the cases where the offender does not know whether the data is valuable to the victim, whereas the victim does know the sensitivity of the exfiltrated data, if exfiltrated. We define this as a double-information asymmetry, and use another signaling game to model this interaction. The analysis help contribute to RQ4. The chapter is based on:

> Meurs, T., Cartwright, E., & Cartwright, A. (2023). Double-sided information asymmetry in double-extortion ransomware. In *International Conference on Decision and game theory for Security* (pp. 311-328). Cham: Springer Nature Switzerland.

In **Part IV**, we will consider law enforcement interventions to combat ransomware. More specifically:

In Chapter 9, we will use double-extortion ransomware attacks to study the effects of law enforcement interventions on ransomware group operations, by considering the victims they publish on leak pages prior and post-intervention, to address RQ5. The chapter is based on:

> Meurs, T., Hoheisel, R., Junger, M., Abhishta, A., & McCoy, D. (2024). What to do against ransomware? Evaluating law enforcement interventions. In *2024 APWG Symposium on Electronic Crime Research (eCrime)* (pp. 1–13). IEEE.

We will conclude the thesis in Chapter 10, where we summarize key findings, discuss implications, and propose ideas for future research. On a final note, this dissertation is a compilation of independent papers. Therefore, there might be some overlap and repetition in different chapters.

This page is intentionally left blank.

Bibliography

[1] G. A. Akerlof. 'The market for "lemons": Quality uncertainty and the market mechanism'. *The quarterly journal of economics* 84.3, 1970, pp. 488–500.

[2] L. Allodi, F. Massacci and J. Williams. *The work-averse cyber attacker model: Theory and evidence from two million attack signatures*. Available at SSRN 2862299. Retrieved from `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2862299`. 2017.

[3] L. Allodi. 'Risk-Based Vulnerability Management. Exploiting the economic nature of the attacker to build sound and measurable vulnerability mitigation strategies.', 2015.

[4] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. V. Eeten, M. Levi, T. Moore and S. Savage. 'Measuring the cost of cybercrime'. *The economics of information security and privacy*. 2013, pp. 265–300.

[5] M. Bac. 'Crime chains'. *American Economic Journal: Microeconomics* 14.4, 2022, pp. 680–722.

[6] M. Bada and J. R. Nurse. 'The social and psychological impact of cyberattacks'. *Emerging cyber threats and cognitive vulnerabilities*. Elsevier, 2020, pp. 73–92.

[7] K. Bedard. 'Human capital versus signaling models: university access and high school dropouts'. *Journal of political economy* 109.4, 2001, pp. 749–775.

[8] N. L. Beebe and V. S. Rao. 'Using situational crime prevention theory to explain the effectiveness of information systems security'. *Proceedings of the 2005 software conference*. Las Vegas, 2005.

[9] R. Brewer, M. de Vel-Palumbo, A. Hutchings, T. Holt, A. Goldsmith, D. Maimon and D. Maimon. 'Situational crime prevention'. *Cybercrime Prevention: Theory and Applications*. 2019, pp. 17–33.

[10] E. Cartwright, J. H. Castro and A. Cartwright. 'To pay or not: game theoretic models of ransomware'. *Journal of Cybersecurity* 5.1, 2019, tyz009.

[11] Centraal Bureau voor de Statistiek. *Online veiligheid en criminaliteit 2022*. `https://www.cbs.nl/nl-nl/publicatie/2023/19/online-veiligheid-en-criminaliteit-2022`. Accessed: 2024-06-19. 2023.

[12] R. V. Clarke. 'Situational crime prevention'. *Environmental Criminology and Crime Analysis*. Ed. by R. Wortley and L. Mazerolle. London, UK: Willan, 2008, pp. 178–194.

[13] R. V. Clarke. 'Situational crime prevention: Its theoretical basis and practical scope'. *Crime and justice* 4, 1983, pp. 225–256.

[14] R. V. Clarke. 'Situational crime prevention: Theoretical background and current practice'. *Handbook on crime and deviance*. Ed. by M. D. Krohn, A. J. Lizotte and G. P. Hall. New York, NY: Springer New York, 2009, pp. 259–276.

[15] L. E. Cohen and M. Felson. 'Social change and crime rate trends: A routine activity approach'. *American Sociological Review* 44.4, 1979, pp. 588–608.

[16] A. Y. Connolly and H. Borrion. 'Reducing ransomware crime: analysis of victims' payment decisions'. *Computers & Security* 119, 2022, p. 102760.

[17] L. Y. Connolly, D. S. Wall, M. Lang and B. Oddson. 'An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability'. *Journal of Cybersecurity* 6.1, 2020, tyaa023.

[18] L. Connolly, M. Lang, P. Taylor and P. Corner. 'The Evolving Threat of Ransomware: From Extortion to Blackmail', 2021.

[19] L. Connolly, D. Wall, M. Lang and B. Oddson. 'An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability'. *Journal of Cybersecurity* 6.1, 2020, pp. 1–18.

[20] D. B. Cornish and R. V. Clarke. 'Introduction'. *The Reasoning criminal: rational choice perspectives on offending*. Ed. by D. B. Cornish and R. V. G. Clarke. New York, NY, US: Springer-Verlag, 1986, pp. 1–18.

[21] D. B. Cornish and R. V. Clarke. 'Rational choice perspective'. *Environmental Criminology and Crime Analysis*. Ed. by R. Wortley and L. Mazerolle. Abingdon, UK: Willan, 2008.

[22] D. B. Cornish and R. V. Clarke. *The reasoning criminal: Rational choice perspectives on offending*. Transaction Publishers, 2014.

[23] D. B. Cornish and R. V. Clarke. 'Understanding crime displacement: An application of rational choice theory'. *Criminology* 25.4, 1987, pp. 933–948.

[24] G. Farrell and K. Pease. 'Preventing repeat and near repeat crime concentrations'. *Handbook of crime prevention and community safety*. Routledge, 2017, pp. 143–156.

[25] M. Felson. 'Linking criminal choices, routine activities, informal control, and criminal outcomes'. *The reasoning criminal*. Routledge, 2017, pp. 119–128.

[26] M. Felson and R. V. Clarke. *Opportunity makes the thief*. 1998.

[27] M. Felson and M. A. Eckert. *Crime and everyday life: A brief introduction*. Sage Publications, 2018.

[28] M. Fuentes, F. Hacquebord, S. Hilt, I. Kenefick, V. Kropotov, R. McArdle, F. Mercês and D. Sancho. *Modern ransomware's double extortion tactics and how to protect enterprises against them*. Trend Micro Research. 2021.

[29] E. Galinkin. 'Winning the Ransomware Lottery: A Game-Theoretic Approach to Preventing Ransomware Attacks'. *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Virtual Event, October 25–27, 2021, Proceedings 12*. Springer. 2021, pp. 195–207.

[30] T. Garkava and D. Ashmore. *Inside the Yanluowang leak: Organization, members, and tactics*. Darktrace Blog. Retrieved May 25, 2024, from `https://darktrace.com/blog/inside-the-yanluowang-leak-organization-members-and-tactics`. 2022.

[31] A. K. Ghazi-Tehrani and H. N. Pontell. 'Phishing evolves: Analyzing the enduring cybercrime'. *Victims & Offenders* 16.3, 2021, pp. 316–342.

[32] A. Glazer and K. A. Konrad. 'A signaling explanation for charity'. *The American Economic Review* 86.4, 1996, pp. 1019–1028.

[33] I. W. Gray, J. Cable, B. Brown, V. Cuiujuclu and D. McCoy. 'Money Over Morals: A Business Analysis of Conti Ransomware'. *2022 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2022, pp. 1–12.

[34] P. H. Hartel, M. Junger and R. J. Wieringa. *Cyber-crime science= crime science+ information security*. Tech. rep. TR-CTIT-10-34. CTIT, University of Twente, 2010.

[35] N. Hassan. *Ransomware revealed*. Springer, 2019.

[36] J. Hernandez-Castro, A. Cartwright and E. Cartwright. 'An economic analysis of ransomware and its welfare consequences'. *Royal Society Open Science* 7.3, 2020, p. 190023.

[37] Hiscox. *Data Exfiltration During Ransomware Attacks*. `https://www.hiscox.co.uk/sites/uk/files/documents/2020-07/20816-Data-exflitration-guide-final.pdf`. Accessed: 2024-06-23. 2020.

[38] H. Ho, R. Ko and L. Mazerolle. 'Situational Crime Prevention (SCP) techniques to prevent and control cybercrimes: A focused systematic review'. *Computers & Security* 115, 2022, p. 102611.

[39] T. Hodgkinson and G. Farrell. 'Situational crime prevention and Public Safety Canada's crime-prevention programme'. *Security Journal* 31, 2018, pp. 325–342.

[40] K. Huang, M. Siegel and S. Madnick. 'Systematically understanding the cyber attack business: A survey'. *ACM Computing Surveys* 51.4, 2018, pp. 1–36.

[41] M. Humayun, N. J. A. Alsayat and V. Ponnusamy. 'Internet of things and ransomware: Evolution, mitigation and prevention'. *Egyptian Informatics Journal* 22.1, 2021, pp. 105–117.

[42] A. Hutchings and T. J. Holt. 'A crime script analysis of the online stolen data market'. *British Journal of Criminology* 55.3, 2015, pp. 596–614.

[43] S. Ikeda. *Under Increasing Federal Scrutiny, BlackCat Ransomware Gang Pulls Exit Scam on Its Way Out*. Cybersecurity Magazine. Retrieved on 05 May 2024, from `https://www.cpomagazine.com/cyber-security/under-increasing-federal-scrutiny-blackcat-ransomware-gang-pulls-exit-scam-on-its-way-out/`. 2024.

[44] M. Junger, A. Abhishta and L. J. M. Nieuwenhuis. *Crime chain: het verband tussen DDoS-aanvallen en Phishing*. Tech. rep. Retrieved from Apeldoorn, Nl.: `https://www.politieenwetenschap.nl/publicatie/pw_verkenningen/2021/crime-chain-364/`. Politie en Wetenschap, 2021.

[45] Kela. *LockBit 2.0 Interview with Russian OSINT*. Retrieved on 05 May 2024 from `https://www.kelacyber.com/lockbit-2-0-interview-with-russian-osint/`. 2021.

[46] M. Keshavarzi and H. R. Ghaffary. 'I2CE3: A dedicated and separated attack chain for ransomware offenses as the most infamous cyber extortion'. *Computer Science Review* 36, 2020, p. 100233.

[47] S. H. Kim, Q.-H. Wang and J. B. Ullrich. 'A comparative study of cyberattacks'. *Communications of the ACM* 55.3, 2012, pp. 66–73.

[48] A. Laszka, S. Farhang and J. Grossklags. 'On the economics of ransomware'. *Decision and Game Theory for Security: 8th International Conference, GameSec 2017, Vienna, Austria, October 23-25, 2017, Proceedings*. Springer. 2017, pp. 397–417.

[49] A. Laszka, E. Panaousis and J. Grossklags. 'Cyber-insurance as a signaling game: Self-reporting and external security audits'. *Decision and Game Theory for Security: 9th International Conference, GameSec 2018, Seattle, WA, USA, October 29–31, 2018, Proceedings 9*. Springer. 2018, pp. 508–520.

[50] J. L. LeBeau. 'The journey to rape: Geographic distance and the rapist's method of approaching the victim'. *Applications of Geographical Offender Profiling*. Routledge, 2017, pp. 155–168.

[51] E. R. Leukfeldt and M. Yar. 'Applying routine activity theory to cybercrime: A theoretical and empirical analysis'. *Deviant Behavior* 37.3, 2016, pp. 263–280.

[52] Z. Li and Q. Liao. 'Game Theory of Data-selling Ransomware'. *J. Cyber Secur. Mobil.* 10.1, 2021, pp. 65–96.

[53] Z. Li and Q. Liao. 'Preventive portfolio against data-selling ransomware—A game theory of encryption and deception'. *Computers & Security* 116, 2022, p. 102644.

[54] Z. Li and Q. Liao. 'Ransomware 2.0: to sell, or not to sell a game-theoretical model of data-selling ransomware'. *Proceedings of the 15th International Conference on Availability, Reliability and Security*. 2020, pp. 1–9.

[55] Lifars.com. *Gangs launch DDoS attacks to push victims into paying ransom*. `https://lifars.com/2020/11/gangs-launch-ddos-attacks-to-push-victims-into-paying-ransom/`. Last checked on Jul 21, 2021. 2020.

[56] A. Mani, T. Vaidya, D. Dworken and M. Sherr. 'An extensive evaluation of the internet's open proxies'. *Proceedings of the 34th Annual Computer Security Applications Conference*. 2018, pp. 252–265.

[57] S. R. Matthijsse, M. S. van 't Hoff-de Goede and E. R. Leukfeldt. 'Your files have been encrypted: A crime script analysis of ransomware attacks'. *Trends in Organized Crime*, 2023, pp. 1–27.

[58] F. Miró. 'Routine activity theory'. *The Encyclopedia of Theoretical Criminology*. 2014, pp. 1–7.

[59] P. A. Networks. *What is multi-extortion ransomware?* Retrieved June 5, 2024, from `https://www.paloaltonetworks.com/cyberpedia/what-is-multi-extortion-ransomware`. n.d.

[60] M. N. Olaimat, M. A. Maarof and B. A. S. Al-Rimy. 'Ransomware anti-analysis and evasion techniques: A survey and research directions'. *2021 3rd International Cyber Resilience Conference (CRC)*. 2021, pp. 1–6.

[61] Openbaar Ministerie. *Het OM in cijfers*. `https://magazines.openbaarministerie.nl/jaarinverhaal/2023/1/het-om-in-cijfers`. Accessed: 2024-06-23. 2023.

[62] M. J. Osborne et al. *An introduction to game theory*. Vol. 3. 3. Oxford university press New York, 2004.

[63] H. Oz, A. Aris, A. Levi and A. S. Uluagac. *A survey on ransomware: Evolution, taxonomy, and defense solutions*. arXiv preprint arXiv:2102.06249. 2021.

[64] M. Paquet-Clouston, B. Haslhofer and B. Dupont. 'Ransomware payments in the bitcoin ecosystem'. *Journal of Cybersecurity* 5.1, 2019, tyz003.

[65] B. Payne and E. Mienie. 'Multiple-extortion ransomware: The case for active cyber threat intelligence'. *ECCWS 2021 20th European Conference on Cyber Warfare and Security*. Vol. 331. Academic Conferences Inter Ltd, 2021.

[66] C. G. J. Putman, A. Abhishta and L. J. M. Nieuwenhuis. 'Business model of a botnet'. *Proc. of the 26th IEEE Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*. Valladolid, Spain: IEEE, 2018, pp. 441–445.

[67] A. Rege, Z. Obradovic, N. Asadi, B. Singer and N. Masceri. 'A temporal assessment of cyber intrusion chains using multidisciplinary frameworks and methodologies'. *2017 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (Cyber SA)*. 2017, pp. 1–7.

[68] E. Rodríguez, S. Verstegen, A. Noroozian, D. Inoue, T. Kasama, M. van Eeten and C. H. Gañán. 'User compliance and remediation success after IoT malware notifications'. *Journal of Cybersecurity* 7.1, 2021, tyab015.

[69] P. Ryan, J. Fokker, S. Healy and A. Amann. 'Dynamics of targeted ransomware negotiation'. *IEEE Access* 10, 2022, pp. 32836–32844.

[70] Security.nl. *Universiteit Maastricht werd besmet via phishingmail en verouderde software*. `https://www.security.nl/posting/642452/Universiteit+Maastricht+werd+besmet+via+phishingmail+en+verouderde+software`. Last checked on Jul 21, 2021. 2020.

[71] K. Shah, T. Shenvi, K. Desai, R. Asrani and V. Jain. 'Phishing: An evolving threat'. *International Journal of Students' Research in Technology & Management* 3.1, 2015, pp. 216–222.

[72] H. A. Simon. 'A Behavioral Model of Rational Choice'. *The Quarterly Journal of Economics* 69.1, 1955, pp. 99–118.

[73] M. Spence. 'Competitive and optimal responses to signals: An analysis of efficiency and distribution'. *Journal of Economic theory* 7.3, 1974, pp. 296–332.

[74] S. G. Van de Weijer, R. Leukfeldt and W. Bernasco. 'Determinants of Reporting Cybercrime: A Comparison Between Identity Theft, Consumer Fraud, and Hacking'. *European Journal of Criminology* 16.4, 2019, pp. 486–508.

[75] F. Weerman. 'Criminaliteit, digitalisering en de online sociale wereld: dezelfde processen in een nieuwe sociale context'. *Tijdschrift voor Criminologie* 61.4, 2019, pp. 395–404.

[76] R. Wortley and M. Townsley. 'Environmental criminology and crime analysis: Situating the theory, analytic approach and application'. *Environmental criminology and crime analysis*. Routledge, 2016, pp. 20–45.

[77] Z. Xu and Q. Hu. 'The Role of Rational Calculus in Controlling Individual Propensity toward Information Security Policy Non-Compliance Behavior'. *Proceedings of the 51st Hawaii International Conference on System Sciences*. 2018.

[78] V. Yegneswaran, P. Barford and J. Ullrich. 'Internet intrusions: global characteristics and prevalence'. *ACM SIGMETRICS Performance Evaluation Review* 31.1, 2003, pp. 138–147.

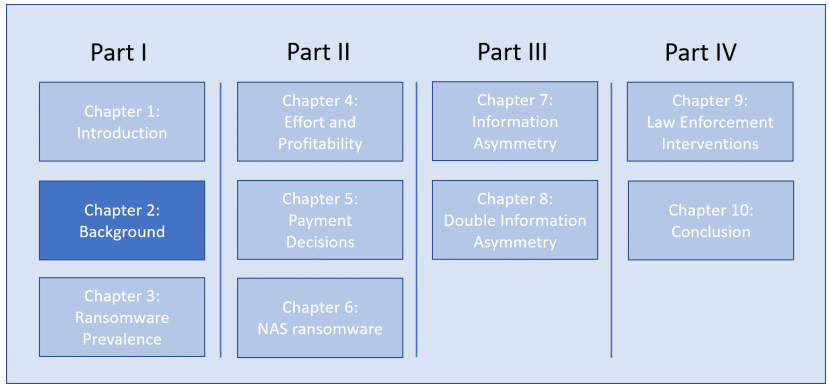
[79] A. Zimba and M. Chishimba. 'On the economic impact of crypto-ransomware attacks: the state of the art on enterprise systems'. *European Journal for Security Research* 4.1, 2019, pp. 3–31.

Crime multiplies

~ Marcus Felson

Chapter 2

Background: A Literature Review



Recent leaks (such as Conti) have provided greater insights on the working of cybercriminal organisations. Just like any other business, these malicious actors strategically manage their processes in order to maximise their revenues. Coordinating different types of cybercrimes as part of a single attack campaign provides another opportunity to these criminal groups to improve the efficiency of their attacks. To investigate the promise of this "coordination" between cybercrimes in improving the financial gains realised by cybercriminals, we take a two-step approach. First, we perform a bibliometric analysis of past scientific literature discussing the concept of "coordination" w.r.t to cybercrime. Second, as a case study, analysing the attack chains of DDoS, phishing and ransomware attacks, we identify vantage points for potential coordination from an attacker's perspective.

2.1 Introduction

Cybercriminals can achieve greater success in their endeavours by using a coordinated set of attack techniques in their strategy [72]. Maastricht University in the Netherlands was struck by a serious ransomware attack which led to attackers gaining access to the computers on December 23rd, 2019. The criminals obtained initial access by sending two phishing emails, where two employees clicked on the attachment [79]. Subsequently, the university decided to pay a ransom of $197,000$ euros to get access to the data encrypted by criminals. However, not all victims of ransomware pay ransom when the demands are first made. For instance, when Glen Dimplex Home Appliances got attacked in October 2020 and they paid the ransom only when the attackers pressured the company by performing a Distributed Denial-of-Service (DDoS) attacks [51]. These examples show that some attacks that at first sight may appear different attack events, but may be part of the same attack event. According to [98], these type of attack events are among the most aggressive and prevalent. We define coordination as *the use of different attacks or crimes for a single attack event*. Understanding coordination is essential to find effective and successful prevention strategies against cybercrime.

Most work on coordination of cyberattacks from a computer science perspective [44], focus on attack coordination and orchestration. To the best of our knowledge, this is mostly theoretical and does not focus on specific cybercrimes. Another part of research literature focuses on the cooperation of criminal actors from an economical [95, 6] or criminological [48] perspective. However, in our view, collaboration and cooperation are different from coordination. As suggested previously, coordination is *the use of different attacks or crimes for a single attack event*. On the contrary, collaboration is *when a group of malicious actors work on a shared objective*. For example, when malware developers and blackhat pentesters working together within a ransomware group [18]. Cooperation is *when a group of malicious actors are working together to help accomplish the goal of one of the groups*. Cooperation is a subset of collaboration. For example, a phishing group helping a ransomware group to get access to a network to install their ransomware. Collaboration and cooperation focus on the relationship between actors, whereas we are interested in the relationship between crimes. Also, we would like to stress that collaboration and cooperation are not mutually exclusive: within a single attack event, both can occur independently of each other.

Although coordinated attacks have been described by cybersecurity companies and blogs [80, 38, 23], to our knowledge no previous scientific research has

systematically investigated the coordinated attacks from an attackers perspective using specific cybercrimes. Additionally, in this study we will argue that coordinated attacks could be more beneficial for the attacker and more severe for the victim than regular types of attack, and that the evolving cybercrime ecosystem will facilitate coordinated attacks in the future. Therefore, this study will focus on coordinated attack events.

We explore coordinated attack events by performing a systematic literature review of coordinated cyberattacks using a bibliometric mapping. Subsequently, we use that information to perform a case study on the coordination of three relatively frequent cybercrimes: DDoS, phishing and ransomware attacks. Previous research has focused on the understanding and prevention of these individual crimes and not their interaction [64, 67, 36]. We illustrate cases of coordination of DDoS, phishing and ransomware as described by the security industry and identify possible vantage points for attackers to coordinate these attacks. Subsequently, we propose COORDINATE: a model to describe different types of coordinated attacks and the benefits and costs for an attacker to decide to coordinate an attack.
Overall, our work focuses on addressing the following research questions:

(i) What is the current state of literature on the coordination and collaboration of cybercrimes?

(ii) What are the costs and benefits for an offender to decide to perform a coordinated attack or not?

The contributions of this work are twofold:

1. A bibliographic mapping of previous academic literature on coordination and collaboration of cybercrimes;

2. Second contribution can be divided into three parts:

   2.a. Introduce a case study of coordinating DDoS, phishing and ransomware and identify potential vantage points for attackers to coordinate these attacks;

   2.b. Identify recent developments in the cybercrime ecosystem and analyse, why they facilitate coordinated attacks;

   2.c. Integrating points 1 and 2.a. into a conceptual model COORDINATE. COORDINATE describes four types of coordination and provides testable hypothesis of the pros and cons of coordination from the criminal's perspective.

The remainder of this chapter is organised as follows. First, we elaborate in Section 2.2 on previous academic literature on coordination and cooperation of cybercrimes. We introduce in Section 2.3 a case study: the coordination of DDoS, phishing and ransomware. We explain in Section 2.4 how the evolving cybercrime ecosystem facilitates the coordination of cybercrimes in the future. Considering these points, we deduce a hypothetical model to describe different types of coordinated attacks and suggest testable predictions for future empirical studies. Finally, in Section 2.5 we summarise our key findings.

2.2 Bibliometric mapping

In this section, we discuss the results of bibliometric analysis of previous academic literature on "coordination" and "collaboration" in relation to cybercrime. First, we discuss the methodology used to perform the bibliometric mapping. Then we present our key findings.

To find the relevant keywords to search for academic literature that discussed "cooperation" and "coordination" with relation to cybercrime, we follow the method described by [7]. They suggest a four step protocol:

(i) Decompose the research question into individual elements.

(ii) Obtain key-words from primary studies.

(iii) Identify synonyms for the main terms.

(iv) Construct search strings using Boolean "AND" to join the main terms and "OR" to include synonyms.

Afterwards, the boolean search string was used to query the literature database Scopus. Subsequently, the literature from the field of mathematics, medical, physics and astronomy sciences were excluded as they are not relevant for studying cooperation and coordination of cybercrimes. We use VOS viewer [20] to identify clusters within resulting literature. We use bibliometric coupling (a measure that represents the number of references shared between two publications) to identify these clusters. Hence, publications within the same cluster, have a substantial overlap in the reference list. We analyse the abstracts of each cluster by using the wordcount of each word in the abstract. Using the top 20 most occuring words within the abstracts of a cluster we identify the clusters which are most relevant to concepts "coordination" and "collaboration" of cybercrimes and cybercriminals. If synonym of these concepts were present in these 20 words we further investigate the content of these clusters. The studies within

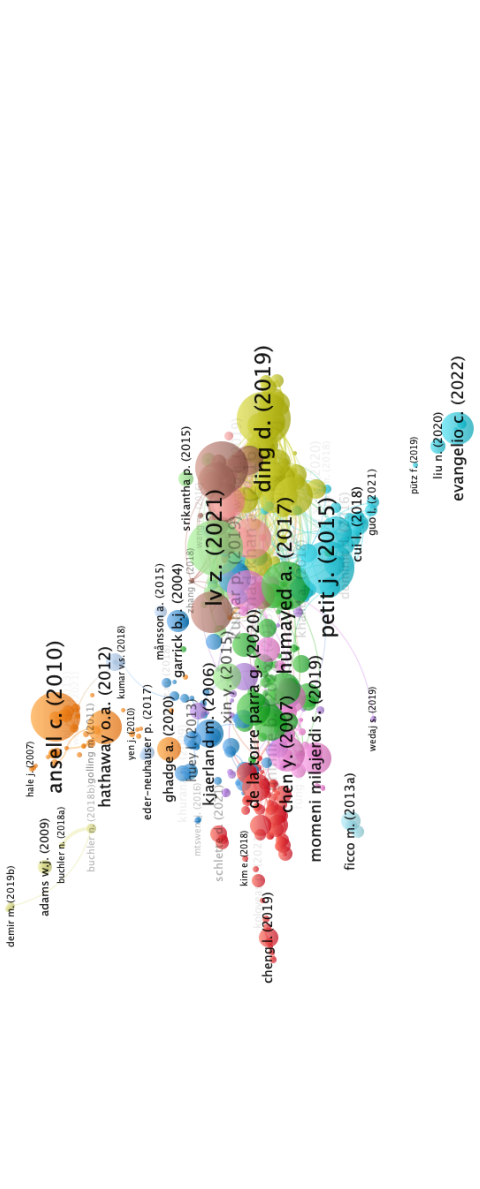


Figure 2.1: Clusters of selected literature. The different colors represent the different clusters of academic literature on coordination of cybercrimes.

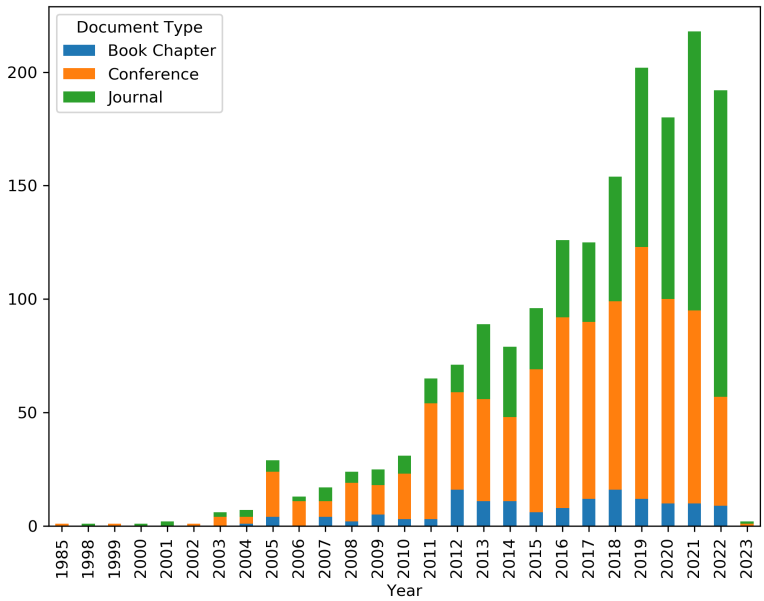


Figure 2.2: Yearly #publications indexed by Scopus.

these clusters were compared to our research objective as described in Section 2.1.

Using the methodology as described above, the main terms of our query were cybercrime and cyberattack, coordination, collaboration, business model and co-operation. We search Scopus database using the following query:

('cyber') AND ('crime' OR 'attack') AND ('coordinat*' OR 'collabora*' OR 'business model' OR 'cooperat*')

Using the Scopus database we found 2341 articles as a result of this query. We excluded publications from the fields of mathematics, medical sciences, physics and astronomy and as a result obtain 1762 articles. The yearly distribution of these publications are shown in Figure 2.2.

These 1762 articles were used for the bibliometric mapping in VOS Viewer. As described above, we used bibliometric coupling as a measure to identify clusters

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Number of studies	64	61	44	42	36	33	31	25	24	32	26	17	15	13	7	2

Table 2.1: Number of studies per cluster found in VOS viewer with the extracted studies from Scopus.

(a) 20 most occurring words in Cluster 10. (b) 20 most occurring words in Cluster 11.

Figure 2.3: Word cloud of 20 most occurring words in abstracts of Cluster 10 (a) en Cluster 11 (b).

of publications related to a similar topic. All the identified clusters are shown in Figure 2.1. Table 2.1 shows the number of studies we found for each cluster. Using the 20 most occurring words of each cluster, we find the clusters most relevant to our study. We identify clusters 1, 7, 9, 10 and 11 as related to concepts of coordination and/or collaboration of cybercrimes. We analyse the studies in each of these clusters to find the connection of these clusters with concepts of coordination/collaboration.

Cluster 1: Keywords: network, framework, attack, security, data. This cluster describes studies were defensive systems coordinate to deter cybercrime. For example, [58] studies the collaboration of different IDPS to detect botnets. [87] develops a honeypot for collaborative defense against distributed attacks of interconnected attackers. Unfortunately, in this study the authors do not explain what distributed attacks of interconnected attackers look like.

Cluster 7: Keywords: vehicle, system, attack, safety, communication. This cluster describes coordination of different systems in a vehicle or several (autonomous) vehicles to defend against cyberattacks. For example, [70] and [93] study cyberattacks against connected autonomous vehicles.

Cluster 9: Keywords: attack, system, power, based, grid. This cluster describes coordinated attacks on a power grid system. The focus is on coordination of the same type of cyberattack. For example, [53] explores distributed smart grid attack strategies to destabilise power system components. The authors consider the objective of the attacker to disrupt the power system by taking control over breakers and coordinating attacks. Subsequently, a strategy is formulated for the opponent to leverage variable structure system theory to attack.

Cluster 10: Keywords: attack, network, model, security, attacker. This cluster describes different cyberattack models, coordinated and collaborated attacks. For example, [5] use a game theoretic approach to model the dynamic behaviour between attacker and defender. The authors argue that each actor adjust his strategy based on costs, potential gain and/or damage and effectiveness of participating the opponent's strategy. [22] develops a canonical model for cyberoperation by advanced attackers. They assume an isolated attack by an individual attacker of homogeneous group. [4] constructs a detection method which can recognise coordinated attacks, by building a 'requires/provides' model. The authors test their model on the multi-stage attack of the Zeus botnet. [97] presents a high-level framework of defending against a cyberattack collaborated by interconnected attackers. The framework consists of five attributes of a coordinated attack: time-aspect, space-aspect, effect of an attack, information change during an attack and the privacy aspect.

Cluster 11: Keywords: system, attack, proposed, model, cyber-physical. Coordination of power grid systems. For example, [28] considers cyber-physical coordinated attacks against power grid and how to formulate a defensive strategy to defend. [92] develops a estimation-based anomaly detection method to defend against cyber-physical smart grid systems. With cyber-physical translates to both cyber as physical security of power systems. This cluster seems highly related to cluster 9.

We can conclude that cluster 10 is most interesting considering the objectives from Section 2.1. Most studies out of cluster 10 are theoretical or consider high-level frameworks of coordinated attacks [5, 97], as for example the canonical model for cyberoperations by advanced attackers [22]. In this study take a different approach: we focus on the costs and benefits of conducting coordinated attacks compared to isolated attacks from the attackers perspective. In the next sections based on a case study we argue the importance to not only consider how

coordinated attacks are performed, but also why attackers have incentives to do so.

2.3 Case Study: Coordinating DDoS, Phishing and Ransomware attacks

In this section we present a case study of coordinating DDoS, phishing and ransomware attacks. First, we performed a small literature review on whether examples of coordination of these three crimes have been studied. In Section 2.3.1 the methodology of finding relevant literature is examined. On the basis of this literature, we present a brief description of DDoS, phishing and ransomware in Section 2.3.2. In Section 2.3.3 we examine possibilities of coordination of the specific crimes, based on the characteristics of the crimes themselves as described in Section 2.3.2. Finally, we consider the repetition of a specific crime as a specific case of coordination.

2.3.1 Methodology

To find specific use cases of coordination in combination with DDoS, phishing and ransomware in the academic literature, we use the following literature databases: Scopus and Web of Science. We have considered the articles/papers published in English language. Since the field of cybercrime is evolving very quickly and we were interested in the most recent modus operandi, we considered literature from the past four years (published since 2017). We also exclude any papers from the field of Medicine. For ransomware the keyword was 'ransomware', for phishing 'phishing' and for DDoS 'DDoS OR denial-of-service'. The results of the search and filtering are shown in Table 2.2.

Table 2.2: Search results of DDoS (DDoS OR denial-of-service), phishing and ransomware on different databases. Hits are the total number of hits with the query. Unique is the amount of unique articles from Scopus and Web of Science, where duplicates are removed and only attributed to Scopus.

Crime	DDoS		Phishing		Ransomware	
Database	Scopus	Web of Science	Scopus	Web of Science	Scopus	Web of Science
Hits	229	481	307	322	263	350
Unique	229	460	307	256	263	253

This resulted in 1765 articles, 689 for DDoS, 563 for phishing and 513 for ransomware. After removing the duplicates we selected articles based on the ab-

stracts which described the modus operandi, victims, offenders, infrastructure or coordination. Articles concerning machine learning models or other automated defense strategies were excluded. This resulted in 244 articles: 97 of ransomware, 94 of phishing and 53 of DDoS. These articles were fully read and used for describing DDoS, phishing and ransomware in Section 2.3 and understanding the cybercrime ecosystem in Section 2.4. If the article referenced to other articles with relevant information about coordination, these other articles were also read, even if the article has been published before 2017. Finally, we add grey literature about coordination based on industry reports related to 'coordination cybercrime', 'DDoS phishing', 'DDoS ransomware', or 'phishing ransomware'. The end date of these queries was 13 September, 2021. This resulted in 16 articles from the security industry used in this chapter. Based on these findings we first give a short description of the modus operandi of the specific crimes in the following section.

2.3.2 Overview DDoS, Phishing and Ransomware

Distributed Denial-of-service (DDoS) is a denial-of-service attack where attackers keep users from accessing a networked system, service, website, application, or other resource [77, 91]. A DDoS attack works by using all available network bandwidth or resources on a target network. Often this is done by using a botnet - entire networks of computers which are infected by malware and under control of a command and control (C&C) server, which is controlled by a botmaster [6]. Often, IoT devices are used for the botnet since they are hardly secured and available in abundance [89, 41]. Anyone with a website or network publicly accessible is prone to DDoS attacks. [91] indicate that 55% of DDoS attacks targeted financial services and web hosting companies. Other obvious targets are retail and e-commerce websites, whose revenue is highly dependent upon their website being available and responsive [55]. For more information about DDoS attacks we refer to [1, 55, 77].

Phishing is the sending of messages with the main objective to gather personal data of users [34, 45]. It is a popular method for stealing credentials, committing fraud and distributing malware. Phishing is based on social engineering: by using methods of persuasion the attacker tries to circumvent a victim's critical thinking and let him perform the action which the phisher wants to accomplish, like giving credentials or installing malware [34]. There are 3 types of targets for phishing: general/indiscriminate, semi-targeted and spear phishing [94]. Different types of phishing target different types of victims [19]: Indiscriminate phishing is when the attacker targets many unrelated victims hoping

at least some will take the bait. Semi-targeted attacks focus on a specific organization or group. With spear phishing a specific individual (often C-level or IT-administrator) is targeted. For more information about phishing attacks we refer to [24, 25, 66].

Ransomware is a category of malicious software that prevents users from accessing their computing device resources by encrypting them [67]. Typically it prevents users from accessing their computing device or files, it shows a screen to provide a way for the victim to pay the ransom. Until the victim pays, the computing device is unusable. Often a deadline is mentioned and an anonymous payment method requested. Ransomware demands used to be typically between 300 to 2000 dollar per target, but is currently much higher [26, 83]. The attack targeting has shifted from individuals to companies [15, 26]. The reasons are twofold: First, targeting has shifted to the healthcare sector, government institutions, and education, because their data is most precious and they often pay high ransoms [32, 30]. Second, it is easier to infect a company than an individual. For more information about ransomware attacks we refer to [12, 26, 15].

2.3.3 Coordinating DDoS, Phishing and Ransomware attacks

(i) **Coordination of ransomware and phishing:** A first type of coordination is between ransomware and phishing. For ransomware to take place, an attacker has to gain access to a network or system. [30, 54, 26] indicate the importance of phishing to gain access to a network, which is than used to install ransomware and perform a ransomware attack. [26] mentions that email phishing accounts for 59% of initial access in ransomware attacks. [100] make the distinction between targeted and bulk ransomware. When the attack is indiscriminate, spam emails are a common way to attack. If the attack is targeted, (spear)phishing and the use of exploits are more typical.

Not only is phishing used to facilitate the installation of ransomware, also ransomware is increasingly used to indirectly steal credentials, which sometimes lead to more phishing [50, 82]. Another way ransomware leads to phishing is in which the content of the phishing email seems more credible by addressing a recent or on going ransomware attack. After the University of Maastricht faced a ransomware attack, it was targeted by a phishing campaign. The emails addressed the ransomware attack, and provided context and credibility to the malicious email [79].

A third way for ransomware to possibly lead to phishing was described by [50]. [50] studied different factors contributing to maximizing profit of

a ransomware attack. Their conclusion was that combining ransomware with data-stealing is in general more profitable than ransomware without stealing the data, and that selling the stolen data is always more profitable than threatening to leak the data. Leaked data is often used for semi-targeted and spear-phishing [82]. Therefore this new method of stealing data during a ransomware attack provides additional opportunities for (targeted) phishing.

(ii) **Coordination of ransomware and DDoS:** A second type of coordination is ransomware and DDoS. Several studies indicate different ways to coordinate ransomware and DDoS. [85] mentions that DDoS is used as retribution for not being able to enter a network, to possibly install ransomware. Furthermore [61] and [3] mention that DDoS is increasingly used as leverage when victims of a ransomware attack decide not to pay the ransom, as was mentioned in the introduction. As example, ransomware gangs like Avaddon group and SunCrypt are mentioned [3]. [39] actively scanned darknet forums and found ransomware actors to actively look for botmasters. This would suggest that ransomware actors do not use easy-to-buy booterservices, but want to possess their own infrastructure to conduct DDoS attacks. Additionally, REvil attackers told in an interview that they want to increase the use of DDoS during a ransomware attack, since victims are more willing to pay the ransom, according to the REvil actor [18].

DDoS is sometimes used to distract attention from a ransomware infection [52, 16]. In this context, an attack with the goal to distract from another attack will be defined as a smokescreen [38]. [16] mentions these smokescreens are done by doing sub-saturating DDoS attacks: low-bandwidth and short in duration (less than 5 minutes). This is done to prevent detection by DDoS mitigation systems. During those 5 minutes, IT staff is busy dealing with momentary network outages, whereas the criminals do automated scanning or penetration techniques to map the network and install the ransomware [16].

Besides these specific forms of coordination of ransomware and DDoS, a more fundamental similarity is that both ransomware and DDoS are basically a denial of resource [100, 75]. This indicates that ransomware and DDoS would only be coordinated if they attack different parts of a network, computer or system. For example, it would not make much sense to perform a DDoS attack on a public-facing server if it is already encrypted by ransomware.

(iii) **Coordination of phishing and DDoS:** A third type of coordination is between phishing and DDoS. Several articles describe cases of coordination between phishing and DDoS. Phishing is sometimes used to increase a botnet, which could be used for DDoS [6]. There are two ways phishing leads to an increased botnet. One way is to use credentials to automatically install malware [82]. Another is to send a email containing phishing and malware at the same time. Another possible link is the use of DDoS to either hide a phishing campaign, or make phishing emails seem more genuine by using it as a storyline or context [36, 40, 42].

The role of context in a phishing email was analysed by [27]. Students either got either an email about winning an I-Pad, or a course-related email. They found that 71.3 per cent of the participants who opened the course-related message also clicked on the simulated phishing link and 63.9 per cent submitted credentials. For the Ipad, these were respectively 5.9 and 3 per cent. They conclude that contextualized social engineering threats like course-related emails lead to victims overlooking cues of deception that normally would be caught in non-contextualized messages. The timing of phishing and DDoS was studied by [36]. They found there to be relatively more phishing emails send before and after a DDoS attack, compared to the baseline without DDoS attack. The authors claim this indicates a coordination of DDoS and phishing, although it could not be established whether this coordination was intended.

2.3.4 Campaigns and repeated attacks

It is worth noting that a form of coordination already exists for a long time within these three types of crimes:

(iv) **Multiple DDoS/phishing/ransomware attacks:** DDoS attacks often consists of multiple attacks. [77] analysed the probability of an attack. He found attacks to be relaunched on the same target less than 5 minutes after the end of the previous one is 58 %. 19 % of all attacks are part of a DDoS campaign of at least 5 consecutive attacks. These findings illustrate the effectiveness of coordinating several DDoS attacks, which is defined as repeating attack [77]. This is also common for for many DDoS hacktivist, who work together to create a larger attack [57, 78].

Multiple phishing attacks: Bulk phishing can lead to spear-phishing (more targeted) [43]. An attacker sends the phishing emails first in bulk. When the attacker receives the credentials of the email-account, he or she will

use this email-account to send new specifically targeted phishing emails to the contacts of the account. Since these emails originated from a trusted sender, more people are inclined to click on the link compared to phishing emails send in bulk [9]. Furthermore, phishing emails are often send in campaigns. [45] defined campaigns as sending a similar phishing email several times over a certain time span. Using campaigns is a cost-effective way to attack from the offender's perspective, since the attacker only needs to change the URL where the victim needs to click.

Multiple ransomware attacks: Ransomware could lead to more ransomware because of worm-like capabilities [54, 12, 26]. The ransomware could therefore infect an entire network automatically. This is the reason why WannaCry was so proliferate [26]. Another way different ransomware attacks are linked is because some high-value targets might be of interest to multiple ransomware actors. It happens that companies receive multiple ransomware attacks, encrypting their files multiple time. The only way to decrypt the files is when the ransomware actors cooperate [18].

Although campaigns and repeats could be considered a specific type of coordination, further analysis is outside the scope of this chapter.

2.4 COORDINATE: the Cybercrime cOORDINATion modEl

Internet presents a global ecosystem that offers, among many other things, the tools, e.g., botnets, CaaS, cryptocurrencies, and an anonymous communication infrastructure, that enables the development and execution of attack chains [65, 49]. In this section, we describe how the recent development of tools and infrastructure within that ecosystem facilitates coordinated attacks and help explain the rise of reported coordinated attacks in Section 2.3. Subsequently, we propose COORDINATE, a new model of coordination and testable predictions to help analyse the costs and benefits of coordination for cybercriminals.

2.4.1 Development Tools and Infrastructure in Cybercrime Ecosystem

[8] analysed the cybercrime ecosystem by considering malware, bitcoins and darknet. We extend this research by briefly describing the evolution of underground forums and markets, cryptocurrencies, online anonymity and botnets. In essence, a cybercriminal wants to anonymously communicate with other cybercriminals (through underground forums and markets), anonymously receive

and send money (with cryptocurrencies) and perform anonymously cyberattacks (through online anonymity and botnets).

(i) **Underground forums and markets:** Cybercriminals need to communicate together if they want to collaborate. This might explain the proliferation of online cybercriminal communities on darknet forums [68]. The rise of new and popular communication technologies is tied with the increasing problem of cybercrime [84]. This is because darknet or underground forums promote the trade of attack tools and services, making cyberattacks accessible for actors with low level of technical sophistication [68]. For a detailed examination of underground forums and markets we refer to [63].

(ii) **Cryptocurrency:** Cryptocurrency technically refers to a cryptographic string of numbers and alphabetic symbols, which together give a unique number and is considered a digital currency which can be exchanged for real-life currencies [74]. It is a common way for cybercriminals to stay anonymous and conceal their money footprint [14]. The first darknet market to accept cryptocurrency was Silk Road in 2011. Although the business model of Silk Road was very successful, in 2013 the FBI shut it down. Nevertheless, cryptocurrency enabled to receive money anonymously. Nowadays most Law Enforcement agencies around the world have different methods to attribute crypto wallets to individuals. Therefore, cybercriminals often use mix services to hide money traces [74].

(iii) **Online anonymity:** The Internet community over the world is interested in anonymity. This led to the development of various anonymous networks. The most important are proxies, virtual private network (VPN) and The Onion Router (TOR) [73]. A VPN creates an encrypted connection over a less secure network, usually the internet, to send encrypted traffic [69]. The use of these technologies improves anonymity of internet users, both normal citizens but also criminals who want to hide their online activities [68, 59].

(iv) **Botnets:** Botnets are remotely controlled networks of computers, often with malicious aims [6]. The types and attack patterns of botnets constantly change, due to a large increase since 2016 in IoT devices which have enough processing power to be part of a botnet [90]. Botnets are most commonly used for DDoS attacks, but the infrastructure has also been used to spread phishing and malware [81], like for example the Emotet botnet.

Altogether, these developments led to the rise of:

(a) **Cybercrime-as-a-Service (CaaS):** Cybercrime-as-a-service is the phenomena that cybercriminals not only perform attacks themselves, but also buy or sell the tools and knowledge to other criminals to perform attacks [56]. Most criminal groups have become highly specialized in specific tools and methods to perform a specific part of an attack [33]. According to [31], CaaS leads to commoditization, specialization and cooperation of cybercriminals. Consequently, we can deduct that cybercrime-as-a-service leads to more interdependence between different cybercrimes, because criminals conducting different types of crime can work together to maximize profit.

(b) **Capabilities and resources:** Offenders can expand capabilities by learning from others through darknet forums. The required capabilities are an important distinction between cybercrimes like DDoS, phishing and ransomware. Ransomware is highly technical, phishing is medium difficult (also depending on web-based or email based phishing) and DDoS attacks are less technical [88]. This means that a non-technical actor could not use ransomware for a coordinated attack. One way to circumvent this problem is to buy tools and services from more technical actors, the phenomena CaaS. Nevertheless, not everything can be bought. For example, some actors who sell ransomware do not want to sell to newbies, because they might screw up and therefore get attention of Law Enforcement [26, 52].

(c) **Democratization of cybercrime:** The dissemination of cybercrime has been noted with respect to offenders as well as victims. Several authors noted that the step towards online offending has become easier over time, during the past decades. One does not need to be technically skilled, but with CaaS everyone can buy a phishing kit [31, 95] and start a phishing campaign or buy a DDoS attack and attack one's school [2]. The commoditization of attacks has led to a democratization of offending, according to [37, 62, 35]. A similar development is found with regard to victimisation. One of the consistent findings in traditional crime is that victims tend to be young and male, have a low educational level and are usually relatively poor [10, 21, 47] because it is strongly related to location and going out [46, 60, 86]. With the digitalization of society, however, offending and victimization of cybercrime become much less related to location or being outdoors. Victims of online crime are both males and female, and for some crimes (online banking fraud, identity theft) of all ages or relatively old

[37]. In summation, offenders as well as victims of online crime tend to be more than before a random – or 'normal' - selection of society.

These developments either directly or indirectly influence the costs and benefits of coordinated attacks. Therefore, it also influences an attacker's decision to perform a coordinated attack. Based on the information gathered in this chapter we propose COORDINATE, a model to evaluate the costs and benefits of coordination for cybercriminals.

2.4.2 COORDINATE

From the empirical observations of coordination of DDoS, phishing, ransomware described in Section 2.3 and the evolution of the cybercrime ecosystem we hypothesise four types of coordination based on the costs and benefits of coordination for cybercriminals:

(i) **Direct collaboration:** One or multiple actors coordinate different attacks before performing the attacks. An example is when a ransomware group uses DDoS attacks to put pressure on a victim if he is not paying the ransom during a ransomware attack [18, 3, 61].

(ii) **Indirect collaboration:** One or multiple entities perform an attack and sell the end-product of that attack to other entities. For example: credentials gained from a phishing attack are sold to a ransomware group, who use the credentials to gain access to a system or network and install their ransomware [26, 100, 30].

(iii) **Opportunistic coordination:** One or multiple actors perform an attack. Subsequently, this becomes known to another actor. Subsequently, this actor uses this knowledge to enforce their own attack. For example: the media reports that a company is victim of a ransomware attack. A phishing group using this information as a context in their phishing email, sending them to the victim [79].

(iv) **Random coordination:** It might be that one or multiple offenders coordinate attack at random, and do not know their attack collides in some way with another attack. Then the attack looks like it is coordinated from a victim's perspective, although the offenders do not know this. A example is a bank who faces both phishing emails and DDoS attacks from two different entities, who do not know from each other an attack occurred [36]. Random coordination is outside the scope of the proposed model.

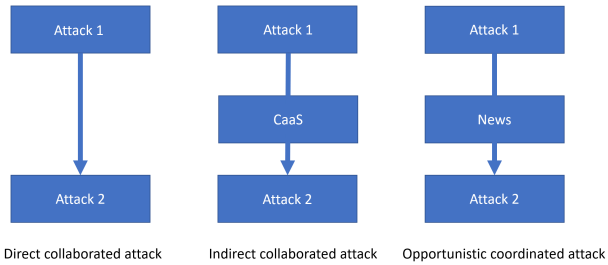


Figure 2.4: The different coordination types examined in this study.

The three relevant types of coordination are depicted in Figure 2.4. Note that it seems that one attack happens after the other, but this is not necessarily what is happening. For example, a DDoS attack could be a smokescreen for installing ransomware at the same time [16, 52]. Nevertheless, the coordination types are applicable to both sequential and parallel coordinated attacks. Here we define a sequential attack as two attacks with no overlap in time and a parallel attack as two attacks with overlap in time.

The various types of coordination lead to different ways of decision-making by an offender compared to no coordination in attack. The literature we found in Section 2.2 mostly focuses on how coordinated attacks could be performed, but not why the attacker would be motivated to do so. From a Rational Choice Perspective [11, 17], financially motivated cybercriminals try to maximize profits while minimizing costs and risks. Based on the use cases and developments presented in the previous sections, we hypothesize the following model, which we call COORDINATE: the CybercrimecOORDINATion modEl.

2.4.2.1 Benefits of coordination for cybercriminals

Performing a coordinated attack compared to a single attack leads to certain benefits. Based on Rational Choice Model of Crime [17], we argue that these benefits need to either increase profit, and/or decrease costs, risks and effort.

(i) **Profitability:** More profit per attack. Every successful attack will generate more profit. It can generate extra profit in two ways. 1) Larger companies or public organizations can be more successfully attacked. Therefore more ransom could be asked during a ransomware attack, or more money could be obtained with phishing or DDoS [76, 29]. 2) Every attack can generate revenue. For example, in a ransomware attack the attackers might gain the ransom, but also selling obtained credentials might directly provide

in extra profits [31, 82]. Higher profit per attack could be most important in direct collaborated attacks, where offenders consciously collaborate, perhaps to go after a 'big fish'. It seems least applicable to opportunistic coordination, because they do not really apply specific targeting [13, 15].

(ii) **Success rate:** Higher probability of success per attack. By putting additional pressure on the victim during a ransomware attack or providing credible context in a phishing email, victims might be more willing to pay ransom or click on the link in the phishing email [18]. Sometimes the attack enables another attack, which means the probability goes from zero per cent (not possible) to a probability higher than zero per cent by coordinating the two attacks.

(iii) **Diffusion of responsibilities:** Coordination leads to diffusion of responsibilities: by performing a small part of the attack, the offender might feel less responsible for the attack [72]. Therefore moral costs are reduced: the feeling of doing something wrong might be less during a coordinated attack. This seems most applicable to indirect collaboration, where the offender selling their services or products do not necessarily know what the other offender is doing with the bought services or products. Diffusion of responsibilities may occur less often with direct collaboration, where an actor is in charge of the entire attack. Decreased moral costs could also occur with opportunistic coordination, since the offender of the second attack does not feel responsible for the first attack.

(iv) **Outsourcing:** Outsourcing the most risky or difficult parts of attack. In coordinated attacks, offenders could decide to perform the parts of an attack which have least risk of being detected or chased by Law Enforcement [31]. For example: they steal credentials or develop ransomware, but someone else deploy the ransomware [26, 12]. Law Enforcement tends to investigate the criminals behind the attack, and not the facilitators and enablers [71, 88]. Therefore, these have less risk of being caught and convicted. Advantages of outsourcing do not occur with direct collaboration, since the offenders have to perform all the aspects of the attack themselves. It most probably happens with indirect collaboration, since many offenders offering their products or services actually offer tools or services to support an attack, but not perform the attack themselves. Finally, opportunistic offenders might only try attacks were they do the less risky attack. For example: they might execute phishing after a ransomware attack. In general, ransomware attacks often attracts more attention than phishing

from Law Enforcement, because impact and severity is often higher. So by phishing after the ransomware, they might receive less attention from Law Enforcement compared to a single phishing attack.

(v) **Shielding:** Repeatedly performing a small part of an attack-type might lead to specialisation [31, 95]. Specialization might lead to better shielding techniques. This does not seem likely for direct collaborated coordinated attacks, because they perform the entire attack chain themselves. On the contrary, better shielding might drive indirect collaboration, where offenders on darknet forums are highly specialised and therefore might have more knowledge how to shield themselves. Likewise, in opportunistic coordinated attacks actors also can not perform the entire attack themselves, and therefore have better shielding compared to actors who are responsible for the entire attack, as in direct collaborated coordinated attacks.

2.4.2.2 Costs of coordination for cybercriminals

Coordinated attacks do not only have advantages, there are also costs:

(i) **Transaction costs**. If the coordinated attack is the result of a collaboration or cooperation of different actors, than this cooperation contains transac-

Table 2.3: Overview proposed hypotheses of relationships between different costs and benefits in COORDINATE. $++$ is a positive relationships, $+$ is a small positive relationship, $+/-$ no relationship, $-$ is a small negative relationship, and $--$ is a negative relationship.

		Direct Collaboration	**Indirect Collaboration**	**Opportunistic Coordination**
Benefits	More profit	$++$	$+$	$+/-$
	Higher probability success	$++$	$++$	$++$
	Decrease moral costs	$+/-$	$++$	$+$
	Outsource most risky parts	$+/-$	$++$	$+$
	Better shielding	$--$	$++$	$+$
Costs	Transaction costs	$++$	$+$	$+/-$
	Timing	$+/-$	$+$	$++$
	Extra effort	$++$	$-$	$--$
	Financial costs	$+$	$++$	$+/-$
	Traces	$++$	$+$	$+/-$

tion costs [95, 99]. From Transaction Cost Economics these costs contain costs of working together, sharing profit, not knowing whether you could trust the other party, etc. [96, 95]. Since direct collaboration consists of the most intensive form of collaboration of all three, it follows that this would have the highest transaction cost, followed by indirect collaboration. Opportunistic coordination does not entail collaboration and therefore no transaction costs.

(ii) **Timing**. For some coordinated attacks timing is important. For example, when phishing for credentials to gain access to a network to install ransomware, the credentials might be invalid after a certain amount of time. Therefore the initial access broker can not wait too long for selling or using the credentials. Timing might be most important for opportunistic coordinated attacks, where they have to react to a another attack in time [36]. For direct collaboration timing might also be important between attacks, but they can decide themselves when the different attacks will be performed. So they are more in control over timing than opportunistic actors. Finally, products and services sold online are probably less time-sensitive than the other two, because it takes time for a vendor to find a buyer. So if timing was important, he would probably be not able to sell it through darknet forums.

(iii) **Extra effort**. Time and energy are required to perform a second attack if done by the same actor. Time spent on the second attack could not be used to do another separate attack, which would have also gained money. This is most important for coordination as a result of direct collaboration, since attackers have to coordinate all the attacks and make sure they have all capacities and resources to perform the attack. For example, if they try to find their own exploits, there is the risk of not finding any. Therefore, it is easier to perform a coordinated attack with products and services bought on darknet forums, and therefore effort should be less for a coordinated attack than uncoordinated attack. This could even more so for opportunistic attacks, they do not need to put any effort in the first attack. So attackers probably do not need to make more effort than if they would perform an uncoordinated attack.

(iv) **Financial costs**. Resources or capabilities needs to be bought, also, if one develops one's own software, than this also directly costs money. These costs are highest for goods and services bought on the darknet market, so indirect collaboration. Financial costs seems to be less so for direct collab-

oration, since attackers only need to buy resources and capabilities they do not have themselves. However, buying resources should be less expensive then end-products. Opportunistic actors do not have to pay anything to perform their coordinated attack, they just react to another attack.

(v) **Traces**. Performing more attacks will lead to more possible traces during an investigation of Law Enforcement. Therefore, performing coordinated attacks could increase the probability of getting caught. This seems most applicable to direct collaboration, since the same group of actors perform the different attacks, and therefore all attacks could be linked back to the group. This seems less applicable for indirect collaboration, because the attacks of criminals are only linked by a purchase over darknet. Linking attacks through darknet markets might be harder than a group with the same modus operandi. Since opportunistic coordinated attack do not have a link with the actors of the first attack, there are no extra traces compared to a single attack.

The hypotheses discussed above are summarised in Table 2.3.We believe these hypotheses need to be tested in further empirical research on coordination of cybercrime.

2.5 CONCLUSIONS

Although coordinated attacks have been described cybersecurity companies and blogs, to our knowledge no scientific research systematically studied coordinated cybercrimes. This chapter set out to identify various ways attacks can be coordinated, describe recent developments w.r.t. coordination/cooperation concepts in cybercrime literature and provide a model of understanding the decision to coordinate attacks or not.

Our first research question: What is the current state of literature on the coordination and coordination of cybercrimes? We addressed this question by analysing the bibliometric mapping of academic literature, we found a cluster of studies which focuses on coordinated cyberattacks from the attackers perspective. They mostly focus on how these crimes can be coordinated, but not on the incentives for the attacker to do so. Therefore, our second research question was: What are the costs and benefits for an offender to decide to perform a coordinated attack or not? We addressed this question by introducing a case study of coordinating DDoS, phishing and ransomware. From the case study, specific vantage points for coordination were identified. Furthermore, through describing the recent developments in the cybercrime ecosystem, we explained

why coordination becomes more feasible for attackers than it did previously. Finally, we deduced a hypothetical model we named the Cybercrime Coordination Model, COORDINATE. From this model we made testable predictions about the importance of certain costs and benefits towards the different types of coordinated attacks.

The results of this study indicate that coordinated attacks result in more harm and are, consequently, more dangerous. We showed that one can already observe attack coordination. If our model is correct, coordinated attacks will be produce more rewards for offenders at lower costs and therefore will occur more often in the future. We are therefore in danger of observing a dynamic system where one crime will lay-out opportunities for new crime that may lead to more and more online crime.

This study was limited by the absence of empirical data on coordinated cybercrimes in order to investigate the severity of such attack events. Despite its exploratory nature, this study offers some insight into the importance of coordinated cybercrimes. We hope this study will be a stepping-stone for other researchers to conduct empirical research on coordinated cybercrimes.

Bibliography

[1] A. Abhishta. *The blind man and the elephant: Measuring economic impacts of ddos attacks*. University of Twente, 2019.

[2] A. Abhishta, M. Junger, R. Joosten, L. Nieuwenhuis and J. Lambert. 'Victim routine influences the number of ddos attacks: Evidence from dutch educational network'. *Proc. of the 40th IEEE Security and Privacy Workshops (SPW), San Fransisco, USA*. IEEE. 2019, pp. 242–247.

[3] L. Abrams. *Another ransomware now uses ddos attacks to force victims to pay*. https://www.bleepingcomputer.com/news/security/another-ransomware-now-uses-ddos-attacks-to-force-victims-to-pay/. Last checked on Aug 19, 2021. 2021. URL: https://www.bleepingcomputer.com/news/security/another-%20ransomware-now-uses-ddos-attacks-to-force-victims-to-pay/.

[4] F. Alserhani, M. Akhlaq, I. Awan and A. Cullen. 'Detection of coordinated attacks using alert correlation model'. *Proc. of the IEEE International Conference on Progress in Informatics and Computing (PIC), Shanghai, China*. Vol. 1. IEEE. 2010, pp. 542–546.

[5] A. Attiah, M. Chatterjee and C. Zou. 'A game theoretic approach to model cyber attack and defense strategies'. *Proc. of the IEEE International Conference on Communications (ICC), Kansas City, USA*. IEEE. 2018, pp. 1–7.

[6] Z. Bederna and T. Szádeczky. 'Effects of botnets–a human-organisational approach'. *Security and Defence Quarterly*, 2021.

[7] P. Brereton, B. Kitchenham, D. Budgen, M. Turner and M. Khalil. 'Lessons from applying the systematic literature review process within the software engineering domain'. *Journal of systems and software* 80.4, 2007, pp. 571–583.

[8] S. Broadhead. 'The contemporary cybercrime ecosystem: A multi-disciplinary overview of the state of affairs and developments'. *Computer Law & Security Review* 34.6, 2018, pp. 1180–1196.

[9] J. Bullee, L. Montoya, M. Junger and P. Hartel. 'Spear phishing in organisations explained'. *Information & Computer Security*, 2017.

[10] J. Bunch, J. Clay-Warner and M. Lei. 'Demographic characteristics and victimization risk: Testing the mediating effects of routine activities'. *Crime & Delinquency* 61.9, 2015, pp. 1181–1205.

[11] L. Cohen and M. Felson. 'Social change and crime rate trends: A routine activity approach'. *American sociological review*, 1979, pp. 588–608.

[12] L. Connolly, M. Lang, P. Taylor and P. Corner. 'The Evolving Threat of Ransomware: From Extortion to Blackmail', 2021.

[13] L. Connolly and D. Wall. 'Hackers are making personalised ransomware to target the most profitable and vulnerable'. *The Conversation* 15, 2019.

[14] L. Connolly and D. Wall. 'The rise of crypto-ransomware in a changing cybercrime landscape: Taxonomising countermeasures'. *Computers & Security* 87, 2019, p. 101568.

[15] L. Connolly, D. Wall, M. Lang and B. Oddson. 'An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability'. *Journal of Cybersecurity* 6.1, 2020, pp. 1–18.

[16] Corero. *The links between ransomware and ddos attacks*. https://www.corero.com/blog/the-links-between-ransom-ransomware-and-ddos-attacks/. Last checked on Jul 22, 2021. 2021. URL: https://www.corero.com/blog/the-links-between-ransom-%20ransomware-and-ddos-attacks/.

[17] D. Cornish and R. Clarke. 'Understanding crime displacement: An application of rational choice theory'. *Criminology* 25.4, 1987, pp. 933–948.

[18] Cyble. *Uncensored interview with revil sodinokibi ransomware operators*. https://blog.cyble.com/2021/07/03/uncensored-interview-with-revil-sodinokibi-ransomware-operators/. Last checked on Jul 22, 2021. 2021. URL: https://blog.cyble.com/2021/07/03/uncensored-interview-with-%20revil-sodinokibi-ransomware-operators/.

[19] A. Darwish, A. E. Zarka and F. Aloul. 'Towards understanding phishing victims' profile'. *Proc. of the IEEE International Conference on Computer Systems and Industrial Informatics (ICCSII), Sharjah, UAE*. IEEE. 2012, pp. 1–5.

[20] D. Effendi, W. Anggraini, A. Jatmiko, H. Rahmayanti, I. Ichsan and M. Rahman. 'Bibliometric analysis of scientific literacy using VOS viewer: Analysis of science education'. 1796.1, 2021, p. 012096.

[21] D. Gottfredson. 'An empirical test of school-based environmental and individual interventions to reduce the risk of delinquent behavior'. *Criminology* 24.4, 1986, pp. 705–731.

[22] T. Grant, I. Burke and R. V. Heerden. 'Comparing models of offensive cyber operations'. *Leading Issues in Cyber Warfare and Security: Cyber Warfare and Security* 2, 2015, p. 35.

[23] Group-IB. *Silence moving into the darkside*. https://www.group-ib.com/resources/threat-research/silence$_$moving-into-the-darkside.pdf. Last checked on Apr 25, 2021. 2018. URL: https://www.group-ib.com/resources/threat-research/silence$%5C_$%20moving-into-the-darkside.pdf.

[24] B. Gupta, A. Tewari, A. Jain and D. Agrawal. 'Fighting against phishing attacks: state of the art and future challenges'. *Neural Computing and Applications* 28.12, 2017, pp. 3629–3654.

[25] R. A. Halaseh and J. Alqatawna. 'Analyzing cybercrimes strategies: The case of phishing attack'. *Proc. of the IEEE Cybersecurity and Cyberforensics Conference (CCC), Amman, Jordan*. IEEE. 2016, pp. 82–88.

[26] N. Hassan. *Ransomware revealed*. Springer, 2019.

[27] F. Hassandoust, H. Singh and J. Williams. 'The Role of Contextualization in Individuals' Vulnerability to Phishing Attempts'. *Australasian Journal of Information Systems* 24, 2020.

[28] H. He, S. Huang, Y. Liu and T. Zhang. 'A tri-level optimization model for power grid defense with the consideration of post-allocated DGs against coordinated cyber-physical attacks'. *International Journal of Electrical Power & Energy Systems* 130, 2021, p. 106903.

[29] M. Hijink. 'Onderhandelen over gijzelsoftware: 'We hadden toch 10 miljoen afgesproken?'' *NRC*, 2021.

[30] M. Hijji and G. Alam. 'A Multivocal Literature Review on Growing Social Engineering Based Cyber-Attacks/Threats During the COVID-19 Pandemic: Challenges and Prospective Solutions'. *IEEE Access* 9, 2021, pp. 7152–7169.

[31] K. Huang, M. Siegel and S. Madnick. 'Systematically understanding the cyber attack business: A survey'. *ACM Computing Surveys* 51.4, 2018, pp. 1–36.

[32] M. Humayun, N. J. A. Alsayat and V. Ponnusamy. 'Internet of things and ransomware: Evolution, mitigation and prevention'. *Egyptian Informatics Journal* 22.1, 2021, pp. 105–117.

[33] T. Hyslip. 'Cybercrime-as-a-Service Operations'. *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, 2020, pp. 815–846.

[34] M. A. Ivanov, B. Kliuchnikova, I. Chugunkov and A. Plaksina. 'Phishing Attacks and Protection Against Them'. *Proc. of the 1st IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), Moscow, Russia*. IEEE. 2021, pp. 425–428.

[35] J. Jansen, M. Junger, L. Montoya, P. Hartel and W. Stol. 'Offenders in a digitized society'. *Cybercrime and the Police*, 2013, pp. 45–59.

[36] M. Junger, A. Abhishta and L. Nieuwenhuis. 'Crime chain: het verband tussen DDoS-aanvallen en Phishing', 2021.

[37] M. Junger, L. Montoya, P. Hartel and M. Heydari. 'Towards the normalization of cybercrime victimization: A routine activities analysis of cybercrime in Europe'. *Proc. of the IEEE International Conference On Cyber Situational Awareness, Data Analytics And Assessment (CyberSA), London, UK*. IEEE. 2017, pp. 1–8.

[38] Kaspersky. *Research reveals hacker tactics: cybercriminals use ddos as smokescreen for other attacks on business*. https://www.kaspersky.com/about/press-releases/2016_research-reveals-hacker-tactics-cybercriminals-use-ddos-as-smokescreen-for-other-attacks-on-business. Last checked on Jul 25, 2021. 2016. URL: https://www.kaspersky.com/about/press-releases/2016$%5C_$research-%20reveals-hacker-tactics-cybercriminals-use-ddos-as-%20smokescreen-for-other-attacks-on-business.

[39] V. Kivilevich. *Ransomware gangs are starting to look like oceans 11*. https://ke-la.com/ransomware-gangs-are-starting-to-look-like-oceans-11/. Last checked on Jul 22, 2021. 2021. URL: https://ke-la.com/ransomware-gangs-are-starting-%20to-look-like-oceans-11/.

[40] KnowBe4. *Whos behind this massive wave of ddos and phishing attacks targeting dutch banks*. https://blog.knowbe4.com/whos-behind-this-massive-wave-of-ddos-and-phishing-attacks-targeting-dutch-banks. Last checked on Jul 25, 2021. 2021. URL: https://blog.knowbe4.com/whos-behind-this-massive-wave-of-%20ddos-and-phishing-attacks-targeting-dutch-banks.

[41] C. Kolias, G. Kambourakis, A. Stavrou and J. Voas. 'DDoS in the IoT: Mirai and other botnets'. *Computer* 50.7, 2017, pp. 80–84.

[42] M. Kotadia. *Ddos makes a phishing e-mail look real*. https://www.zdnet.com/article/ddos-makes-a-phishing-e-mail-look-real/. Last checked on Jul 25, 2021. 2006. URL: https://www.zdnet.com/article/ddos-makes-a-phishing-e-mail-%20look-real/.

[43] Y. Kwak, S. Lee, A. Damiano and A. Vishwanath. 'Why do users not report spear phishing emails?' *Telematics and Informatics* 48, 2020, p. 101343.

[44] H. Lallie, K. Debattista and J. Bal. 'A review of attack graph and attack tree visual syntax in cyber security'. *Computer Science Review* 35, 2020, p. 100219.

[45] E. Lastdrager. *From fishing to phishing*. University of Twente, 2018.

[46] J. Lauritsen, J. Laub and R. Sampson. 'Conventional and delinquent activities: Implications for the prevention of violent victimization among adolescents'. *Violence and victims* 7.2, 1992, pp. 91–108.

[47] J. Lauritsen, R. Sampson and J. Laub. 'The link between offending and victimization among adolescents'. *Criminology* 29.2, 1991, pp. 265–292.

[48] E. R. Leukfeldt, E. Kleemans and W. Stol. 'Cybercriminal networks, social ties and online forums: Social ties versus digital ties within phishing and malware networks'. *The British Journal of Criminology* 57.3, 2017, pp. 704–722.

[49] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich and H. Liu. 'Click trajectories: End-to-end analysis of the spam value chain'. *Proc. of the 32nd IEEE Symposium on Security and Privacy (SP), Oakland, USA*. IEEE. 2011, pp. 431–446.

[50] Z. Li and Q. Liao. 'Game Theory of Data-selling Ransomware'. *Journal of Cyber Security and Mobility*, 2021, pp. 65–96.

[51] Lifars.com. *Gangs Launch DDoS Attacks To Push Victims Into Paying Ransom*. https://lifars.com/2020/11/gangs-launch-ddos-attacks-to-push-victims-into-paying-ransom/. Last checked on Jul 21, 2021. 2020. URL: https://lifars.com/2020/11/gangs-launch-ddos-attacks-to-push-%20victims-into-paying-ransom/.

[52] A. Liska and T. Gallo. *Ransomware: Defending against digital extortion*. O'Reilly Media, 2016.

[53] S. Liu, B. Chen, T. Zourntos, D. Kundur and K. Butler-Purry. 'A coordinated multi-switch attack for cascading failures in smart grid'. *IEEE Transactions on Smart Grid* 5.3, 2014, pp. 1183–1195.

[54] M. Loman. 'How Ransomware Attacks'. *Sophos*, 2019.

[55] T. Mahjabin, Y. Xiao, G. Sun and W. Jiang. 'A survey of distributed denial-of-service attack, prevention, and mitigation techniques'. *International Journal of Distributed Sensor Networks* 13.12, 2017, p. 1550147717741463.

[56] D. Manky. 'Cybercrime as a service: a very modern business'. *Computer Fraud & Security* 2013.6, 2013, pp. 9–13.

[57] S. Mansfield-Devine. 'The evolution of DDoS'. *Computer Fraud & Security* 2014.10, 2014, pp. 15–20.

[58] L. Mathews, A. Joshi and T. Finin. 'Detecting botnets using a collaborative situational-aware idps'. *Proc. of the 2nd IEEE Second International Conference on Information Systems Security and Privacy (ICISSP), Rome, Italy*. 2016.

[59] P. Meland, Y. Bayoumy and G. Sindre. 'The Ransomware-as-a-Service economy within the darknet'. *Computers & Security* 92, 2020, p. 101762.

[60] E. Mustaine and R. Tewksbury. 'Predicting risks of larceny theft victimization: A routine activity analysis using refined lifestyle measures'. *Criminology* 36.4, 1998, pp. 829–858.

[61] S. Newman. *How Ransomware is Teaming Up with DDoS*. `https://www.infosecurity-magazine.com/opinions/ransomware-teaming-ddos/`. Last checked on Jul 21, 2021. 2021. URL: `https://www.infosecurity-magazine.com/opinions/ransomware-%20teaming-ddos/`.

[62] A. Noroozian. *Evaluating Hosting Provider Security Through Abuse Data and the Creation of Metrics*. TU Delft, 2020.

[63] E. Nunes, A. Diab, A. Gunn, E. Marin, V. Mishra, V. Paliath, J. Robertson, J. Shakarian, A. Thart and P. Shakarian. 'Darknet and deepnet mining for proactive cybersecurity threat intelligence'. *Proc. of the 14th IEEE Conference on Intelligence and Security Informatics (ISI), San Antonio, USA*. IEEE. 2016, pp. 7–12.

[64] D. O'Leary. 'What phishing e-mails reveal: An exploratory analysis of phishing attempts using text analysis'. *Journal of Information Systems* 33.3, 2019, pp. 285–307.

[65] A. Oest, Y. Safei, A. Doupé, G. Ahn, B. Wardman and G. Warner. 'Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis'. *Proc. of the 13th APWG Symposium on Electronic Crime Research (eCrime), San Diego, USA*. IEEE. 2018, pp. 1–12.

[66] A. Oest, P. Zhang, B. Wardman, E. Nunes, J. Burgis, A. Zand, K. Thomas, A. Doupé and G. Ahn. 'Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale'. *Proc. of the 29th USENIX Security Symposium (USENIX), Santa Clara, USA*. 2020, pp. 361–377.

[67] H. Oz, A. Aris, A. Levi and A. S. Uluagac. *A survey on ransomware: Evolution, taxonomy, and defense solutions*. arXiv preprint arXiv:2102.06249. 2021.

[68] S. Pastrana, A. Hutchings, A. Caines and P. Buttery. 'Characterizing eve: Analysing cybercrime actors in a large underground forum'. *Proc. of the 21st International Symposium on Research in Attacks, Intrusions, and Defenses (RAID), Heraklion, Greece*. Springer. 2018, pp. 207–227.

[69] A. Pavlicek and F. Sudzina. 'Use of virtual private networks (VPN) and proxy servers: Impact of personality and demographics'. *Proc. of the 13th International Conference on Digital Information Management (ICDIM), Berlin, Germany*. IEEE. 2018, pp. 108–111.

[70] A. Petrillo, A. Pescape and S. Santini. 'A secure adaptive control for cooperative driving of autonomous connected vehicles in the presence of heterogeneous communication delays and cyberattacks'. *IEEE transactions on cybernetics* 51.3, 2020, pp. 1134–1149.

[71] N. Popper. 'Ransomware attacks grow, crippling cities and businesses'. *The New York Times*, 2020.

[72] C. Putman and L. Nieuwenhuis. 'Business model of a botnet'. *Proc. of the 26th IEEE Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Valladolid, Spain*. IEEE. 2018, pp. 441–445.

[73] E. Ramadhani. 'Anonymity communication VPN and Tor: a comparative study'. *Journal of Physics: Conference Series*. Vol. 983. 1. IOP Publishing. 2018, p. 012060.

[74] E. Reddy and A. Minnaar. 'Cryptocurrency: A tool and target for cybercrime'. *Acta Criminologica: African Journal of Criminology & Victimology* 31.3, 2018, pp. 71–92.

[75] B. A. S. Al-rimy, M. A. Maarof and S. Z. M. Shaid. 'A 0-day aware crypto-ransomware early behavioral detection framework'. *Proc. of the 2nd International Conference of Reliable Information and Communication Technology (IRICT), Johor Bahru, Malaysia*. Springer. 2017, pp. 758–766.

[76] E. R. Ritenour. 'Hacking and ransomware: challenges for institutions both large and small'. *American Journal of Roentgenology* 214.4, 2020, pp. 736–737.

[77] J. C. J. Santanna. *DDoS-as-a-Service: investigating booter websites*. University of Twente, 2017.

[78] M. Sauter. '"LOIC Will Tear Us Apart" The Impact of Tool Design and Media Portrayals in the Success of Activist DDOS Attacks'. *American Behavioral Scientist* 57.7, 2013, pp. 983–1007.

[79] Security. *Universiteit Maastricht werd besmet via phishingmail en verouderde software*. https://www.security.nl/posting/642452/Universiteit+Maastricht+werd+besmet+via+phishingmail+en+verouderde+software. Last checked on Jul 21, 2021. 2020. URL: https://www.security.nl/posting/642452/Universiteit+Maastricht%20+werd+besmet+via+phishingmail+en+verouderde+software.

[80] S. Shead. *Symantec data stealing hackers use ddos to distract from attacks*. https://www.zdnet.com/article/symantec-data-stealing-hackers-use-ddos-to-distract-from-attacks/. Last checked on Jul 25, 2021. 2012. URL: https://www.zdnet.com/article/symantec-data-stealing-hackers-%20use-ddos-to-distract-from-attacks/.

[81] T. Sigurdardottir and S. Neubauer. *Emotet from a banking trojan to one of the most advanced botnets*. https://www.cyren.com/blog/articles/emotet-from-a-banking-trojan-to-one-of-the-most-advanced-botnets. Last checked on Jul 22, 2021. 2019. URL: https://www.cyren.com/blog/articles/emotet-from-a-%20banking-trojan-to-one-of-the-most-advanced-botnets.

[82] C. Simoiu, A. Zand, K. Thomas and E. Bursztein. 'Who is targeted by email-based phishing and malware? Measuring factors that differentiate risk'. *Proc. of the 20th ACM Internet Measurement Conference (IMC), Virtual*. 2020, pp. 567–576.

[83] R. Sobers. *Ransomware statistics 2021*. https://www.varonis.com/blog/ransomware-statistics-2021/. Last checked on Jul 19, 2021. 2021. URL: https://www.varonis.com/blog/ransomware-statistics-2021/.

[84] A. Sutanrikulu, S. Czajkowska and J. Grossklags. 'Analysis of Darknet Market Activity as a Country-Specific, Socio-Economic and Technological Phenomenon'. *Proc. of the IEEE APWG Symposium on Electronic Crime Research (eCrime), Virtual*. IEEE. 2020, pp. 1–10.

[85] B. Tonev. *Cyber attack guide ddos attacks*. https://www.scalahosting.com/blog/cyber-attack-guide-ddos-attacks/. Last checked on Jul 22, 2021. 2021. URL: https://www.scalahosting.com/blog/cyber-attack-guide-ddos-%20attacks/.

[86] A. Tseloni, K. Wittebrood, G. Farrell and K. Pease. 'Burglary victimization in England and Wales, the United States and the Netherlands: A cross-national comparative test of routine activities and lifestyle theories'. *British Journal of Criminology* 44.1, 2004, pp. 66–91.

[87] E. Vasilomanolakis, S. Karuppayah, M. Mühlhäuser and M. Fischer. 'Hostage: a mobile honeypot for collaborative defense'. *Proc. of the 7th International Conference on Security of Information and Networks (SINCONF), Glasgow, UK*. 2014, pp. 330–333.

[88] Verizon. *2020 Data breach investigations report*. https://enterprise.verizon.com/resources/reports/2020-data-breach-investigations-report.pdf. Last checked on Jul 19, 2021. 2020. URL: https://enterprise.verizon.com/resources/reports/2020-data-%20breach-investigations-report.pdf.

[89] R. Vishwakarma and A. Jain. 'A survey of DDoS attacking techniques and defence mechanisms in the IoT network'. *Telecommunication Systems* 73.1, 2020, pp. 3–25.

[90] S. Vu, M. Stege, P. El-Habr, J. Bang and N. Dragoni. 'A Survey on Botnets: Incentives, Evolution, Detection and Current Trends'. *Future Internet* 13.8, 2021, p. 198.

[91] D. Walkowski. *What is a distributed denial of service attack*. https://www.f5.com/labs/articles/education/what-is-a-distributed-denial-of-service-attack-. Last checked on Sep 12, 2021. 2019. URL: https://www.f5.com/labs/articles/education/%20what-is-a-distributed-denial-of-service-attack-.

[92] H. Wang, X. Wen, S. Huang, B. Zhou, Q. Wu and N. Liu. 'Generalized attack separation scheme in cyber physical smart grid based on robust interval state estimation'. *International Journal of Electrical Power & Energy Systems* 129, 2021, p. 106741.

[93] P. Wang, X. Wu and X. He. 'Modeling and analyzing cyberattack effects on connected automated vehicular platoons'. *Transportation Research Part C: Emerging Technologies* 115, 2020, p. 102625.

[94] D. Warburton. *2020 phishing and fraud report*. https://www.f5.com/labs/articles/threat-intelligence/2020-phishing-and-fraud-report. Last checked on Sep 12, 2021. 2020. URL: https://www.f5.com/labs/articles/threat-%20intelligence/2020-phishing-and-fraud-report.

[95] R. V. Wegberg, S. Tajalizadehkhoob, K. Soska, U. Akyazi, C. Ganan, B. Klievink, N. Christin and M. V. Eeten. 'Plug and prey? Measuring the commoditization of cybercrime via online anonymous markets'. *Proc. of the 27th USENIX Security Symposium (USENIX), Baltimore, USA*. 2018, pp. 1009–1026.

[96] O. Williamson. 'Transaction cost economics and organization theory'. *Industrial and Corporate Change* 2.2, 1993, pp. 107–156.

[97] S. Xu. 'Collaborative attack vs. collaborative defense'. *Proc. of the 4th International Conference on Collaborative Computing: Networking, Applications and Worksharing (COLLABORATECOM), Orlando, USA*. Springer. 2008, pp. 217–228.

[98] V. Yegneswaran, P. Barford and J. Ullrich. 'Internet intrusions: Global characteristics and prevalence'. *ACM SIGMETRICS Performance Evaluation Review* 31.1, 2003, pp. 138–147.

[99] G. Zhou, J. Zhuge, Y. Fan, K. Du and S. Lu. 'A market in dream: the rapid development of anonymous cybercrime'. *Mobile Networks and Applications* 25.1, 2020, pp. 259–270.

[100] A. Zimba and M. Chishimba. 'On the economic impact of crypto-ransomware attacks: the state of the art on enterprise systems'. *European Journal for Security Research* 4.1, 2019, pp. 3–31.

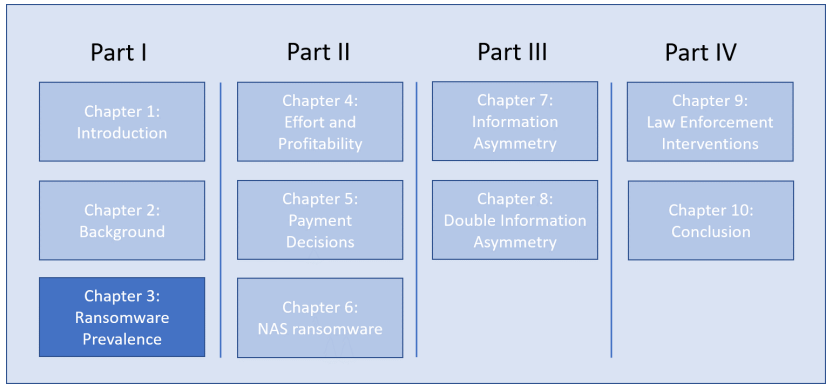
This page is intentionally left blank.

Ransomware actors don't see 'big' and 'small'; they see opportunity. Every vulnerable system is a potential payday.

~ Brian Krebs

Chapter 3

Ransomware Prevalence



Accurate crime measurement is crucial for scientists, policymakers, and the public. Traditional methods, like self-reporting and official statistics, struggle with reliability, validity, and sampling, particularly with the rise of online crime. This study estimates ransomware prevalence in the Netherlands from 2019 to 2022 using a capture-recapture method, combining data from police reports, incident response companies, and leak pages. Results show significant underreporting, with only 41.4% and 40.2% of attacks detected in large and medium businesses, respectively. The estimate for small businesses is less reliable, but it suggests a higher number of unobserved victims compared to larger companies.

3.1 Introduction

Knowing how much crime there is in a country is important for a number of reasons, among them government accountability, informs public awareness and research, and aids in resource allocation [54]. Traditionally there are three main sources of crime statistics: self-report victimization, self-report offending, and official statistics on recorded crime by law enforcement agencies [38]. Measures based on interviews, whether they focus on offending [33, 56] or victimization [5, 13, 21], have problems to solve with respect to the reliability and validity of their instruments, and they have to deal with sampling problems, such as an increase of nonresponse over time [55, 36]. For instance, the Dutch victimization survey has response percentages of around 32% [2]. Furthermore, police reports include only a selection of victims as victims do not always report an incident to the police [57, 59].

While offline crime is not very easy to measure, measurement problems become even more complicated with online crime [19]. The anonymity of the internet makes it hard to identify offenders, and online crimes are more likely to go unnoticed compared to traditional crimes. For example, data theft might not be detected immediately, making it challenging to measure the true extent of the crime. The hidden nature of specific online crimes adds to these measurement challenges, as they are not as physically visible as traditional crimes. Finally, according to [21], one of the main problems of measuring cybercrime is the relative absence of official data. However, this is not true to the same extent for all online crime.

The present study focuses on estimating the prevalence of ransomware. A ransomware attack is an example of online crime, which involves malicious software that encrypts a victim's data, with the attacker demanding a ransom for the decryption key. In recent years, ransomware has become a significant societal concern [8, 4, 15, 16]. This concern comes, among other things, from the high costs to victims and the significant disruptions to daily life, as exemplified by the Colonial Pipeline incident that led to widespread fuel shortages in the United States [4].

Measuring the prevalence of ransomware attacks is crucial for understanding their impact. There are three primary sources that provide data on ransomware attacks: police reports, incident response companies, and leakpages. Police reports provide information on incidents brought to the attention of law enforcement. Incident response companies offer insights from their operations assisting victims in recovering from ransomware attacks. Leakpages are websites where attackers publish data of victims who do not pay the ransom.

By linking individual victims in these datasets, its combination provides a way to measure ransomware prevalence, taking into account that every dataset in itself might be biased, as described previously. Using this combination we apply capture-recapture methodology, or multiple system estimation (MSE), to compute estimates of the total number of ransomware attacks for large, average, and small businesses [61]. Accordingly, our main research question is:

> **How many ransomware attacks are there in the Netherlands in 2019 - 2022?**

Multiple systems estimation (MSE) is a methodology used in official statistics, particularly with population censuses and administrative data sources. MSE, also known as capture-recapture, is widely used to estimate the size of populations that cannot be completely observed [14]. This method links multiple data sources, or 'lists,' to estimate the number of unobserved cases. By definition, the number of cases that is missed by all lists is unknown. By analyzing the overlap between these lists, it is possible to estimate this number, and once we have this estimate, we can infer the total number of incidents.

The outline of this chapter is as follows: in §2 we consider the background literature on traditional crime rate estimation methods and potentially new data sources based on the ransomware crime script. In §3 we present our data and the methodology. Afterwards, §4 presents the results on the amount of ransomware attacks in the Netherlands. In Section §5 we compare our results with the Dutch Victimization Survey of the Statistics Netherlands [6]. Subsequently, we discuss our findings and conclude in §6 and §7, respectively.

3.2 Background

Having basic information on crime is essential for nation-states. Citizens of developed countries usually have at least some concerns about crime levels in their community [11, 24, 49]. Knowledge about the amount of crime and its characteristics matters to citizens and policymakers. Accordingly, adequate crime statistics are important. A commission of the UK government [54] listed five major reasons why a nation needs crime statistics at a national level:

1. **Government accountability:** To provide reliable quantitative measurements of criminal activity and trends that enable parliament to fulfill its democratic function of holding the government accountable for this aspect of the state of the nation.

2. **Public awareness and research:** To keep the public, media, academia, and relevant special interest groups informed about the state of crime in the country, and to provide (access to) data that informs wider debates and non-governmental research agendas.

3. **Resource allocation:** To inform relevant aspects of short-term resource allocation, both within government and for external related bodies, e.g., for policing and Victim Support.

4. **Performance and accountability:** To inform performance management and accountability at the national level for agencies such as the police.

5. **Strategic policy development:** To provide an evidence base for longer-term government strategic and policy developments [54].

A common measurement tool is victimization (and offender) self-report surveys. Victimization surveys provide a valuable perspective on the level of crime as experienced by the population, capturing incidents that are not reported to or recorded by the police. Victimization (and offender) surveys have been conducted in the Netherlands since 1980 by the Statistics Netherlands (CBS), offering a long-term view of crime trends [28, 2]. By sampling private households and asking individuals aged 15 years and older about their experiences with various crimes, victimization surveys can uncover hidden crime figures, especially for offenses that victims may choose not to report to the police.

Since 2017, Statistics Netherlands (CBS) introduced a victimization survey specifically focused on online crime that focused on businesses: the Dutch Cybersecurity Monitor [7]. Data is collected through the annual ICT survey, involving around 20,000 randomly selected companies and 22,000 self-employed individuals. Specific questions about ransomware have been included since 2021. In 2022, Statistics Netherlands reported that 15% of Dutch residents were victims of online crime, with 80% of them not reporting incidents to the police [7]. In 2021, 6,300 ransomware attacks were reported, including 4,000 incidents among self-employed individuals and 2,300 targeting businesses. By 2022, this increased to 8,310 attacks, with 6,000 involving self-employed individuals. Larger companies were disproportionately affected, with 4% of businesses with 250+ employees reporting attacks in 2021, compared to 0.3% of self-employed individuals. This trend continued in 2022, when larger companies were still more affected by ransomware than smaller ones.

Business victimization surveys have the advantage, like victimization surveys of individuals, of measuring crime that is not necessarily reported to the police (see below). However, alongside advantages, business victimization surveys

also have problems and issues. The sampling process is complex. For example, who to interview from a large company, how to achieve representation from all economic sectors and companies of different sizes, are issues that need to be satisfactorily resolved [22, 23]. Non-response is a problem with only around 50% of companies participating in the English/Welsh Commercial Victimisation Survey [23, 29]. Also, business victimization surveys are based on information from a single respondent, and the percentage of victimized companies who responded with "don't know" or "no answer" is high (30.8%) [31]. Furthermore, operationalizing the various concepts that make up 'online crime' is not straightforward. There is some overlap with different categories of online crime [31] and respondents may not be aware of the types of online crime and terminology used in the surveys [30].

Another traditional source of crime statistics are police reports. Police reports contain recorded incidents reported to or discovered by law enforcement. In the Netherlands, these records have been systematically collected since 1950, providing a long-term dataset for crime trend analysis [59]. They also provide legally verified information on crimes, making them a reliable source for serious offenses.

Nevertheless, police reports are limited by underreporting, as was mentioned above. This has been shown in surveys of individuals [57, 59] and of businesses [23, 17, 31]. This matters as underreporting is related to crime characteristics such as whether the perpetrator was a known person [52], the type and impact of the incident [31], and fear of reputational damage [1].

Many crimes, especially online crime, go unreported because victims may feel that law enforcement cannot help, or because the crime is not recognized as serious enough to report [42]. Furthermore, few victims report online crime to the police, compared to offline crime [32, 57], although this may be an effect of the type of crime and not a difference between online and offline crime. For example, [57] found a willingness to report of 8-10% of victims of online fraud and [42] found a willingness to report of 2-5% of victims of a particular ransomware variant. Additionally, not all reported crimes are officially recorded due to investigative priorities or legal policies [59]. All these aspects of commercial victimization surveys introduce selection biases into the police data. Furthermore, changes in laws, public awareness campaigns, and administrative practices can influence the consistency and comparability of police data over time. Thus, while useful, police reports are not representative of the mix of crimes experienced by victims [53].

The modus operandi of ransomware may provide potential new data sources to measure the prevalence of ransomware attacks. The modus operandi can be

described using a crime script, which breaks down the steps involved in executing an attack [9, 27]. Crime scripts might reveal potential new data sources to measure ransomware incidents. The ransomware crime script [44, 37] includes (1) developing infrastructure and malware, (2) buying ransomware malware from other malicious actors, defined as Ransomware-as-a-Service (RaaS), (3) gaining access via methods like phishing or brute force attacks, (4) moving laterally within the network, (5) exfiltrating sensitive data for extra extortion, (6) encrypting files, (7) communicating with victims for ransom negotiation, (8) deciding on ransom payment, (9) applying blackmail, and (10) laundering ransom and providing decryption keys [25, 39, 26, 48, 35, 20, 51, 34, 50, 47].

This crime script suggests additional methods for measuring ransomware incidents beyond traditional approaches, such as using leak pages where victims are exposed for non-payment, and data from incident response companies that assist with recovery, negotiations, and ransom management. Other potential sources, like negotiation pages, bitcoin payment records, and the market for initial access brokers, are beyond the scope of this chapter.

Incident response companies offer valuable insights into ransomware attacks that are often not reported to law enforcement [41, 60]. These companies assist victims in recovering from attacks, negotiating with attackers, and managing ransom payments. However, their data tends to overrepresent larger organizations, as only companies with sufficient financial resources can typically afford these services, leading to a bias in the dataset.

Leak pages, where ransomware groups publish the names or data of victims who refuse to pay the ransom, provide another source of unreported incidents. Monitoring these sites can reveal additional ransomware cases. However, this data is also biased. Not all victims are exposed; attackers may withhold data if a ransom was paid, or may focus on high-profile targets to boost their reputation [40]. Some attackers also lack the resources to publish all cases. As a result, leak pages tend to overrepresent larger companies, further skewing the distribution of reported victims [43].

In the present study, we integrate data from police reports, incident response companies, and leak pages to develop a comprehensive picture of ransomware incidents. By cross-referencing victim names, we can identify which victims appear across multiple datasets and which are unique to a single source. This approach enables us to estimate the number of unobserved ransomware attacks, producing independent estimates that we will compare with the victimization survey of the Statistics Netherlands, the Cybersecurity Monitor, in the discussion section [7].

3.3 Methodology

3.3.1 Data

From the study, the population size was based on observations from three datasets between 1 January 2019 and 31 December 2022.

1. **Police Reports (P):** Official reports of ransomware attacks targeting Dutch companies were filed with Dutch Law Enforcement. For a detailed report about the data collection process, we refer to [44, 41]. From the 525 attacks, we excluded attacks on individuals and attempted attacks. We included 434 incidents in this study.
2. **Incident Response Companies (I):** Data from an Incident Response company based in the Netherlands, specialized in helping victims recover from ransomware attacks. From the 99 attacks, 30 incidents were outside the Netherlands and therefore left out of the analysis, since we do not know whether they reported to the Police. Since we need to match cases with the other two data sources, this makes it unfeasible to use this data. We included 69 incidents in this study.
3. **Leakpages (L):** Websites where attackers publish stolen data or victim names if the ransom is unpaid. From the 9200 attacks, 9139 attacks were outside the Netherlands and therefore not used in this study. The leakpage dataset was from ecrime.ch and provided to the researchers [10]. We included 61 attacks in this study.

This study aimed to estimate the prevalence of unreported ransomware attacks across different company sizes in the Netherlands, analyzing data from police reports (P), leak page data (L), and incident response data (I), categorized by small (K), medium (M), and large (G) companies. Companies between 1-50 employees are categorized as small, between 51-250 employees as medium, and 251+ employees as large.

A summary of the data is presented in Table 3.1. Observations were linked by considering company size and victim company name across observations. We considered the probability that two different victims have the same company name and size and are attacked at the same time period to be acceptably small. This procedure led to 477 unique observations.

3.3.2 Analysis

To estimate the hidden number of ransomware attacks, we employ a method for the estimation of the size of a population known as multiple systems estimation.

Table 3.1: Dataset used for this study. The categories are one-hot encoded and categorized by data source (P, I, L) and company size (S).

P	I	L	S	Frequency
1	0	0	L	30
1	1	0	L	8
0	0	1	L	8
1	0	1	L	8
0	1	1	L	1
1	1	1	L	2
1	0	0	M	48
0	1	0	M	6
1	1	0	M	13
0	0	1	M	7
1	0	1	M	12
1	1	1	M	2
1	0	0	S	293
0	1	0	S	12
1	1	0	S	4
0	0	1	S	15
1	0	1	S	1
0	1	1	S	2
1	1	1	S	5

We follow the explanation that was provided earlier in Coumans et al. (2017), for the estimation of homeless. This estimation method has its roots in biology, where it was originally developed to estimate the size of hidden populations, such as animal species in the wild. Over time, it has been adapted for broader use in fields like epidemiology and social sciences and is particularly effective for estimating elusive populations, such as drug users and individuals experiencing homelessness.

The methodology is widely recognized in statistical applications. Examples include public health research [14], homelessness studies [12], official statistics [58], and investigations into human slavery [12].

Capture-recapture techniques using linked administrative datasets provide an efficient and cost-effective solution for estimating population sizes, such as the prevalence of ransomware attacks. A key advantage of this method is its ability to address incomplete data, a common issue with ransomware registers. However, its application depends on several assumptions. For two linked datasets, the inclusion in one dataset must be independent of inclusion in the other. When linking more than two datasets, this strict assumption can be replaced by the requirement that higher-order interactions are absent. Additionally, reliable linking of datasets requires sufficient identifying information, which must comply with privacy regulations.

To illustrate, consider two datasets, A and B. Linking them results in counts of cases unique to A, unique to B, and common to both A and B. These counts form a contingency table, with the unobserved cell representing cases missing from both datasets. Estimating this missing cell allows the total population size to be calculated by summing observed and estimated counts.

Log-linear models are used to estimate the unobserved cell. These models describe the logarithm of cell frequencies in terms of main effects and interaction terms. For a 2x2 contingency table of datasets A and B, the saturated log-linear model is given by:

$$\log m_{ab} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_{ab}^{AB},$$

where m_{ab} represents the expected frequency, λ is the intercept, λ_a^A and λ_b^B are the main effects, and λ_{ab}^{AB} is the interaction term. Since the unobserved cell is missing, this saturated model cannot be directly estimated.

Instead, the independence model assumes no interaction between A and B, and is written as:

$$\log m_{ab} = \lambda + \lambda_a^A + \lambda_b^B.$$

This model allows for estimation under the assumption that the datasets are independent. While this assumption is often unrealistic, refinements can improve its application.

One refinement includes covariates, such as company size, which account for heterogeneity in inclusion probabilities. With a covariate X, the table expands to three dimensions, and the model can be expressed as:

$$\log m_{abx} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_x^X + \lambda_{ax}^{AX} + \lambda_{bx}^{BX}.$$

This approach relaxes the strict independence assumption, replacing it with the condition that independence holds within levels of the covariate.

A second refinement adds a third dataset C, enabling pairwise dependencies to be modeled without requiring complete independence. For three datasets, the log-linear model is:

$$\log m_{abc} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_{ab}^{AB} + \lambda_{ac}^{AC} + \lambda_{bc}^{BC}.$$

In shorthand, this model is denoted as $[AB][AC][BC]$. Note that while pairwise dependencies are allowed, higher-order interactions are excluded. The inclusion of additional datasets and covariates provides more flexibility, relaxing the assumptions of independence and homogeneity.

For model selection, the standard approach involves evaluating model fit using criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These measures penalize models that are overly complex, helping to prevent overfitting. The model with the lowest AIC or BIC is preferred [3].

3.4 Results

A model search is carried out using the BIC, and this leads to the model $[PI][PS][ILS]$. I.e. there is an interaction between P and I, between P and S and between I, L and S. So, controlling for the other variables, in this model there is no direct relation between P and L. The estimated frequencies with 95% confidence intervals for the unobserved cases are presented below:

Table 3.2: Estimated ransomware incidents under model $[PI][PS][ILS]$

	Observed Cases	Unobserved Estimated	Total	Observed (%)	CI 2.5	CI 97.5
L	57	80.7	137.7	41.4%	55.1	122.6
M	88	130.7	218.7	40.2%	98.3	190.1
S	332	2373.4	2705.4	12.3%	1272.6	7057.2

For Large and Middle size companies the estimates of unobserved attacks are quite reliable with points estimates 80.7 (CI 55.1 – 122.6) and 130.7 (CI

Table 3.3: Model search using three levels of Size

Model	Logl	pars	AIC	BIC	Large	Middle	Small
1. $[PI][PS][ILS]$	-770.3	16	1572.6	1638.2	81	131	2,373
2. 1. + PL	-768.4	17	1570.9	1640.7	169	274	11,978
3. 2. + PIS	-768.5	18	1573.0	1646.8	87	116	4,725
4. 1. - PI	-774.5	15	1579.0	1640.5	72	111	1,182
5. 4. -PS	-780.9	14	1589.8	1647.2	100	157	751
6. 4. -ILS	-784.3	14	1596.8	1654.0	82	138	1,204

98.3 – 190.1), but for Small companies the number of unobserved attacks is not reliable, with estimate 2,373.4 (CI 1,272.6 – 7,057.2). Given the large number of observed cases for Small companies, which is 332, we can only conclude that for Small companies the number of ransomware attacks is larger than for Middle and Large companies.

The estimated total number of ransomware attacks for Large and Medium companies is 137.7 and 218.7, respectively, with significant underreporting in both categories. Observed cases totaled 145, while unobserved cases were estimated at 211.4, making the overall total 356.4 attacks. This indicates that 40.7% of ransomware attacks on Large and Medium companies are reported, while 59.3% go unreported. For Large companies, 41.4% of attacks are observed (57 incidents) and 58.6% unobserved (80.7 incidents), while for Medium companies, 40.2% of attacks are observed (88 incidents) and 59.8% are unobserved (130.7 incidents).

We study the model search procedure, in order to have more confidence in this outcome. See Table 3.3. Model $[PI][PS][ILS]$ has the smallest BIC of 1,638.2. It has 16 parameters. Adding the term PL to the model leads to a higher BIC of 1,640.7, but lowers the AIC. For this model the estimate for Small companies increases considerably, and becomes unrealistically large. For other models adding or deleting terms lead to suboptimal AIC and BIC values.

If we consider model 2 in more detail, by fitting models on the table where we left out the counts for small companies, we find estimates 81 for Large and 130 for Middle Sized companies. We conclude that the estimates for Model 2 found in Table 3.3 are due to the inclusion of the Small companies, that lead to instability of all estimates. We conclude that we can safely use the estimates in Table 3.3. In

summary, our analysis indicates that a significant number of ransomware attacks remain unobserved through conventional reporting methods.

3.5 Comparing with Cybersecurity Monitor

In this section, we compare our estimates with a victimization survey from Statistics Netherlands in 2021 and 2022, the Cybersecurity Monitor [7] (see Table 3.4). Our models estimate that large companies experienced 138 ransomware attacks, while medium-sized companies faced 218 attacks between 2019 and 2022. Combining these estimated number of total ransomware attacks with the number of companies in the Netherlands in 2021 for different company sizes, extrapolated from the Cybersecurity Monitor [7], we calculate the ransomware attack risk for large companies at 5.3% and for medium-sized companies at 2.2% between 2019 and 2022. These figures translate to an average annual risk of 1.3% for large companies and 0.6% for medium companies of becoming a ransomware victim. Although there may be some uncertainty in these estimates due to fluctuations in the number of companies between 2019 and 2022, we believe they reflect the correct order of magnitude. In comparison, the Cybersecurity Monitor reported ransomware attack rates of 4.0% for large companies in 2021 and of 2.3% for medium-sized companies, dropping to 2.3% and 1.4%, respectively, in 2022 [7].

Year	Small Companies	Medium Companies	Large Companies
Ransomware Attack Probability (%)			
Study: 2019-2022	0.2	2.2	5.3
CBS: 2021	2.0	2.3	4.0
CBS: 2022	0.5	1.4	2.3
Yearly Average Ransomware Attack Probability (%)			
Study: 2019-2022	0.1	0.6	1.3
CBS: 2021-2022	1.3	1.9	3.2
Reported to Police and/or Cybersecurity Company Aggregated (%)			
Study (+leakpage) 2019-2022	12.3	40.2	41.4
CBS Police 2021-2022	24.9	43.4	48.4
CBS IR Company 2021-2022	36.9	53.8	58.7

Table 3.4: Ransomware Attacks and Reporting Percentages by Company Size according to the present study and Cybersecurity Monitor of CBS (Statistics Netherlands) [7]

Our estimates appear to be relatively lower than those from the Cybersecurity Monitor, which could be due to several factors. First, our analysis focuses on direct victims, excluding indirect victims affected through interdependence of companies. The Statistics Netherlands dataset may include both direct and indirect victims, inflating their numbers. Second, our data does not account for attempted ransomware attacks, which are likely underreported to the police, incident response companies, and leakpages, but may be included in victimization surveys. Lastly, calculation limitations could lead to discrepancies in outcomes; for instance, the exact number of companies per size category is only available for 2021, and we had to extrapolate data for other years. Furthermore, only the percentage of ransomware attacks for 2021 and 2022 are available from CBS.

Despite these limitations, our estimates for the risk of ransomware attacks fall within the confidence intervals (CI) of our study (Table 3.4). Specifically, the CBS estimate for large companies (4.0% in 2021 and 2.3% in 2022) aligns with our CI of 2.1% to 4.7%. For medium-sized companies, CBS estimates (2.3% in 2021 and 1.4% in 2022) fall within our CI of 1.0% to 1.9%. For small companies, CBS estimates (2.0% in 2021 and 0.5% in 2022) are consistent with our CI of 0.8% to 4.6%. This alignment suggests that both CBS and our estimates provide reliable estimates of risk of ransomware attacks, demonstrating the robustness of our findings.

3.6 Discussion

The present study estimates the total number of ransomware attacks on businesses in the Netherlands between 2019 and 2022. According to our estimates, 138 large companies, 219 medium companies, and 2706 small companies suffered from a ransomware attack, suffered from a ransomware attack. While the estimates for large and medium companies are reliable, those for small companies carry high uncertainty due to wide confidence intervals. As a result, we present the findings for large, medium, and small companies separately, acknowledging the limitations for small companies. Based on our estimates, we calculated that there is an annual risk of 1.3% for large companies and 0.6% for medium companies of suffering a ransomware attack. This is in line with previous figures of the Cybersecurity Monitor published by Statistics Netherlands in 2021 and 2022 [7].

Our analysis shows significant underreporting of incidents to the police across all company sizes. For large companies, about 41.4% of attacks are observed, while 58.6% go unreported. Similarly, 40.2% of medium-sized company attacks are captured, leaving 59.8% unobserved. However, it should be noted that about

40% of attacks reported to the police, incident response company and/or leakpage, is considerably more than police reporting of online crime in general, like online fraud. Previous research found police reporting rates for online fraud of 11.5% in the UK [46], 14% in the US [45], 13.4% in Portugal [18], and in the Netherlands, percentages ranging from 11.8% [32] to 13 and 14% [57].

One reason for higher reporting rates in our findings compared to prior research, might be the more severe impact of ransomware attacks on medium and large companies [44]. Serious online crimes are generally reported more often, as supported by prior research [41, 42]. For instance, Deadbolt ransomware, which primarily targets individuals and small businesses, had low reporting rates of 2.8% to 5.1% [42]. Smaller companies may choose not to report due to lower perceived financial loss or other factors. In contrast, larger companies are more likely to report ransomware attacks, potentially due to operational impacts or insurance requirements [41].

The estimated percentage of ransomware attacks observed (or reported) in our study aligns with the Cybersecurity Monitor's reporting figures (see Table 3.4). According to the Cybersecurity Monitor, 37% of companies with two or more employees sought help from cybersecurity firms after an attack, while only 18% reported the incident to the police, with reporting rates decreasing for smaller businesses. These percentages are close to the 40% observed in our dataset from the three data sources. This is noteworthy given the limitations of our data, such as relying on only one incident response (IR) company, while the Cybersecurity Monitor includes victims who used any cybersecurity or IR service. Despite these limitations, the consistency between the datasets highlights the robustness of our findings.

Finally, our study has several other limitations that affect the generalizability of our findings. Firstly, the willingness of victims to report ransomware attacks to the police may vary across countries due to cultural and moral differences. Since this study focused only on the Netherlands, the estimates may differ when using data from other countries. The representation of victims on leak pages might also vary internationally, influenced by differing tendencies to pay ransoms. Additionally, our study is based on data from a single incident response company, which may not be representative of the broader industry. Finally, as mentioned before, we do not include data on individuals who become victim of ransomware, attempted ransomware and indirect victims. These numbers would provide a more reliable estimation of the victimization of ransomware.

Despite these limitations, we believe our results are significant for several reasons. Firstly, our methodology allows us to extract valuable information from multiple data sources and understand the interaction between these sources.

Secondly, while the exact figures may vary, we expect the general trend of higher underreporting rates among small companies to hold true across different contexts. This is likely due to small companies being less represented in various data sources compared to medium and large companies. However, this hypothesis needs to be tested in follow-up research.

3.7 Conclusion

This study highlights the importance of using multiple data sources to measure the full scope of ransomware attacks. To answer our main research question: **How many ransomware attacks occurred in the Netherlands between 2019 and 2022?**, we applied the capture-recapture methodology. Our analyses indicate that, for large companies, 57 (41.4%) ransomware attacks were reported, with 80.7 (58.6%) of the attacks unobserved. For medium-sized companies, 88 (40.2%) ransomware attacks were reported, with 130.7 (59.8%) of the attacks unobserved. Overall, 137.7 large companies, 218.7 medium companies, and 2705.4 small companies suffered from a ransomware attack. We noted that the estimate small companies is unreliable. The average annual risk of a ransomware attack is 1.3% for large companies and 0.6% for mid-sized companies.

Our results align closely with the Statistics Netherlands Cybersecurity Monitor [7]. This has several implications: First, the results are robust, as we obtain similar estimates using independent methods. Second, our approach may be more cost-efficient than a large-scale victimization survey, making it preferable for exploratory research or to reduce costs.

Future research should focus on small businesses, where uncertainty in our estimates remains high due to wide confidence intervals. The uncertainty could be reduced if more of the attacks reported to the police were also detected by Incident Response Companies and on Leakpages, increasing the overlap between sources. However, it is unclear how this can be achieved. Small companies often lack the resources to address cybersecurity threats and may underreport attacks due to perceived insignificance, resource limitations, or unawareness of reporting mechanisms. There is also a belief that police may not take small companies as seriously as larger ones, resulting in fewer police reports. Many small businesses cannot afford incident response services, further reducing detection. Offenders may also avoid posting small firms on leak pages to maintain their reputation. This underreporting suggests many ransomware incidents go undetected, highlighting the need for additional datasets of ransomware targeting small businesses. However, estimates for medium and large companies are

encouraging, as higher-than-expected reporting rates implies a more accurate picture of ransomware than previously assumed.

This page is intentionally left blank.

Bibliography

[1] A. Abhishta, R. van Rijswijk-Deij and L. J. Nieuwenhuis. 'Measuring the impact of a successful DDoS attack on the customer behaviour of managed DNS service providers'. *ACM SIGCOMM Computer Communication Review* 48.5, 2019, pp. 70–76.

[2] M. Akkermans, R. Kloosterman, E. Moons, C. Reep and M. T.-v. d. Aa. *Veiligheidsmonitor 2021 (the safety monitor 2021)*. Retrieved from `https://www.cbs.nl/-/media/pdf/2022/09/veiligheidsmonitor.pdf`. Centraal Bureau voor de Statistiek, 2022, p. 99.

[3] D. R. Anderson and K. P. Burnham. 'Avoiding pitfalls when using information-theoretic methods'. *The Journal of Wildlife Management*, 2002, pp. 912–918.

[4] J. Blatchly. 'The Impact of Ransomware–A Comparison of Worldwide Governmental Policies and Recommendations for Future Directives'. PhD thesis. Utica University, 2023.

[5] D. Cantor and J. P. Lynch. 'Self-report surveys as measures of crime and criminal victimization'. *Criminal Justice & Behavior* 4, 2000, pp. 85–138.

[6] Centraal Bureau voor de Statistiek. *Aantal bedrijven naar grootteklasse*. Tech. rep. 2024. URL: `https://www.cbs.nl/nl-nl/cijfers/detail/81588NED`.

[7] Centraal Bureau voor de Statistiek. *Online veiligheid en criminaliteit 2022*. `https://www.cbs.nl/nl-nl/publicatie/2023/19/online-veiligheid-en-criminaliteit-2022`. Accessed: 2024-06-19. 2023.

[8] A. Y. Connolly and H. Borrion. 'Reducing ransomware crime: analysis of victims' payment decisions'. *Computers & Security* 119, 2022, p. 102760.

[9] D. B. Cornish. 'The procedural analysis of offending and its relevance for situational prevention'. *Crime Prevention Studies* 3.1, 1994, pp. 151–196.

[10] C. Cosin. *Ecrime*. Retrieved March 1, 2023. 2022. URL: https://ecrime.ch/.

[11] T. A. M. Crawford and K. Evans. 'Crime prevention and community safety'. *Oxford Handbook of Criminology*. Ed. by A. Leibling, S. Maruna and L. McAra. 6th ed. Oxford, UK: Oxford University Press, 2016.

[12] M. J. L. F. Cruyff, J. van Dijk and P. G. M. van der Heijden. 'The challenge of counting victims of human trafficking'. *Chance* 30, 2017, pp. 41–49.

[13] L. E. Daigle, J. A. Snyder and B. S. Fisher. 'Measuring victimization: Issues and new directions'. *The Handbook of Measurement Issues in Criminology and Criminal Justice*. Ed. by B. M. Huebner and T. S. Bynum. West Sussex, UK: John Wiley & Sons, Inc., 2016, pp. 249–276.

[14] I. W. G. for Disease Monitoring and Forecasting. 'Capture–recapture and multiple record systems estimation. Part I. History and theoretical development'. *American Journal of Epidemiology* 142, 1995, pp. 1059–1068.

[15] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2021*. Tech. rep. Retrieved August 31, 2022. Luxembourg, 2021. URL: https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021.

[16] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2023*. Tech. rep. Retrieved August 31, 2023. Luxembourg, 2023. URL: https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-assessment-iocta-2023.

[17] J. Flatley. *Crime against businesses: findings from the 2022 Commercial Victimisation Survey*. Tech. rep. London, UK, 2023. URL: https://www.gov.uk/government/statistics/crime-against-businesses-findings-from-the-2023-commercial-victimisation-survey/.

[18] C. Fonseca, S. Moreira and I. Guedes. 'Online consumer fraud victimization and reporting: A quantitative study of the predictors and motives'. *Victims & Offenders* 17.5, 2022, pp. 756–780.

[19] J. Gibbon, T. Marjanov, A. Hutchings and J. Aston. 'Measuring the Unmeasurable: Estimating True Population of Hidden Online Communities'. *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2024, pp. 56–66.

[20] J. Hernandez-Castro, A. Cartwright and E. Cartwright. 'An economic analysis of ransomware and its welfare consequences'. *Royal Society Open Science* 7.3, 2020, p. 190023.

[21] T. J. Holt. 'Cybercrime'. *The Handbook of Measurement Issues in Criminology and Criminal Justice*. Ed. by B. M. Huebner and T. S. Bynum. West Sussex, UK: John Wiley & Sons, Inc., 2016, pp. 29–48.

[22] M. Hopkins. 'Business, victimisation and victimology: Reflections on contemporary patterns of commercial victimisation and the concept of businesses as 'ideal victims''. *International Review of Victimology* 22.2, 2016, pp. 161–178.

[23] M. Hopkins. 'The crime drop and the changing face of commercial victimization: Reflections on the 'commercial crime drop' in the UK and the implications for future research'. *Criminology & Criminal Justice* 16.4, 2016, pp. 410–430.

[24] M. Hough and J. V. Robert. 'Public Opinion and Criminal Justice: The British Crime Survey and Beyond'. *Surveying Crime in the 21st Century: Commemorating the 25th Anniversary of the British Crime Survey*. Ed. by J. Hough and M. Maxfield. Vol. 22. Monsey, NY: Criminal Justice Press, 2007, pp. 197–220.

[25] D. Y. Huang, M. M. Aliapoulios, V. G. Li, L. Invernizzi, K. McRoberts, E. Bursztein, J. Levin, K. Levchenko, A. C. Snoeren and D. McCoy. 'Tracking Ransomware End-to-end'. *39th IEEE Symposium on Security and Privacy (S & P)*. 2018, pp. 618–631.

[26] K. Huang, M. Siegel and S. Madnic. 'Systematically Understanding the Cyber Attack Business: A Survey'. *ACM Computing Surveys* 51.4, 2019, Article 70. URL: https://doi.org/10.1145/3199674.

[27] A. Hutchings and T. J. Holt. 'A crime script analysis of the online stolen data market'. *British Journal of Criminology* 55.3, 2015, pp. 596–614.

[28] H. Huys and J. Rooduijn. 'A new survey on justice and security'. *Netherlands Official Statistics* 9, 1994, pp. 47–51.

[29] IPSOS. *Crime against businesses: findings from the 2023 Commercial Victimisation Survey*. Tech. rep. Technical Report 2023. London, UK: IPSOS, 2023. URL: https://assets-uk.ipsos.com/pa/cvs/2023/cvstechnicalreport.pdf.

[30] M. Junger and P. Hartel. 'Crime Survey & Cybercrime: The State-of-the-art'. *Measuring Cybercrime in the Time of Covid-19: The Role of Crime and Criminal Justice Statistics*. Ed. by M. F. Aebi, S. Caneppele and L. Molnar. Strassbourg, France (online): Eleven Publisher, 2022, pp. 75–88.

[31] S. Kemp, D. Buil-Gil, F. Miró-Llinares and N. Lord. 'When do businesses report cybercrime? Findings from a UK study'. *Criminology and Criminal Justice*, 2021.

[32] L. Koning, M. Junger and B. P. Veldkamp. 'Reporting fraud victimization to the police: factors that affect why victims do not report'. *Psychology, Crime and Law*, 2023. in press.

[33] M. D. Krohn, T. P. Thornberry, C. L. Gibson and J. M. Baldwin. 'The development and impact of self-report measures of crime and delinquency'. *Journal of Quantitative Criminology* 26.4, 2010, pp. 509–525.

[34] P. Leo, Ö. Işik and F. Muhly. 'The Ransomware Dilemma'. *MIT Sloan Management Review*, 2022.

[35] Z. Li and Q. Liao. 'Game Theory of Data-selling Ransomware'. *J. Cyber Secur. Mobil.* 10.1, 2021, pp. 65–96.

[36] A. Luiten, J. Hox and E. de Leeuw. 'Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys'. *Journal of Official Statistics* 36.3, 2020, pp. 469–487.

[37] S. R. Matthijsse, M. S. van 't Hoff-de Goede and E. R. Leukfeldt. 'Your files have been encrypted: A crime script analysis of ransomware attacks'. *Trends in Organized Crime*, 2023, pp. 1–27.

[38] M. G. Maxfield and E. R. Babbie. *Research Methods for Criminal Justice and Criminology*. CengageBrain.com, 2010.

[39] P. H. Meland, Y. F. F. Bayoumy and G. Sindre. 'The Ransomware-as-a-Service Economy Within the Darknet'. *Computers & Security* 92, 2020, p. 101762.

[40] T. Meurs, E. Cartwright, A. Cartwright, M. Junger and A. Abhishta. 'Deception in Double Extortion Ransomware Attacks: An Analysis of Profitability and Credibility'. *Computers & Security* 138, 2024, p. 103670.

[41] T. Meurs, E. Cartwright, A. Cartwright, M. Junger, R. Hoheisel, E. Tews and A. Abhishta. 'Ransomware Economics: A Two-Step Approach to Model Ransom Paid'. *2023 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2023, pp. 1–13.

[42] T. Meurs, R. Hoheisel, M. Junger and A. Abhishta. 'NAS Ransomware: Ransomware Targeting NAS Devices'. *Proceedings of the Human Factors in Cybercrime Conference 2024*. in press. 2024.

[43] T. Meurs, R. Hoheisel, M. Junger, A. Abhishta and D. McCoy. 'What To Do Against Ransomware? Evaluating Law Enforcement Interventions'. *2024 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2024, pp. 1–13.

[44] T. Meurs, M. Junger, E. Tews and A. Abhishta. 'Ransomware: How Attacker's Effort, Victim Characteristics and Context Influence Ransom Requested, Payment and Financial Loss'. *2022 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2022, pp. 1–13.

[45] R. E. Morgan. *Financial Fraud in the United States, 2017*. Tech. rep. NCJ 255817. Washington DC: Bureau of Justice Statistics, 2021. URL: https://bjs.ojp.gov/redirect-legacy/content/pub/pdf/ffus17.pdf.

[46] ONS. *Nature of Fraud and Computer Misuse in England and Wales: Year Ending March 2022*. Tech. rep. 2022. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/natureoffraudandcomputermisuseinenglandandwales/yearendingmarch2022.

[47] K. Oosthoek, J. Cable and G. Smaragdakis. 'A Tale of Two Markets: Investigating the Ransomware Payments Economy'. *arXiv preprint* arXiv:2205.05028, 2022.

[48] H. Oz, A. Aris, A. Levi and A. S. Uluagac. 'A Survey on Ransomware: Evolution, Taxonomy, and Defense Solutions'. *ACM Computing Surveys (CSUR)*, 2021.

[49] R. Pain. 'Place, Social Relations and the Fear of Crime: A Review'. *Progress in Human Geography* 24.3, 2000, pp. 365–387.

[50] B. Payne and E. Mienie. 'Multiple-extortion ransomware: The case for active cyber threat intelligence'. *ECCWS 2021 20th European Conference on Cyber Warfare and Security*. Vol. 331. Academic Conferences Inter Ltd, 2021.

[51] C. P. Research. *Leaks of Conti Ransomware Group Paint Picture of a Surprisingly Normal Tech Start-Up*. Tech. rep. 2022. URL: https://research.checkpoint.com/2022/leaks-of-conti-ransomware-group-paint-picture-of-a-surprisingly-normal-tech-start-up-sort-of/.

[52] I. Shahbazov, Z. Afandiyev and A. Balayeva. 'Some Determinants of Crime Reporting Among Economic and Financial Crime Victims: The Case of Azerbaijan'. *Journal of White Collar and Corporate Crime* 4.1, 2023, pp. 24–37.

[53] W. G. Skogan. 'Reporting Crimes to the Police: The Status of World Research'. *Journal of Research in Crime and Delinquency* 21.2, 1984, pp. 113–137.

[54] A. Smith. *Crime Statistics: An Independent Review*. Tech. rep. London, UK: Home Office, 2006. URL: http://rds.homeoffice.gov.uk/rds/pdfs06/crime-statistics-independent-review-06.pdf.

[55] I. A. L. Stoop. *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague, The Netherlands: Social and Cultural Planning Office, 2005.

[56] T. P. Thornberry and M. D. Krohn. 'The Self-Report Method for Measuring Delinquency and Crime'. *Measurement and Analysis of Crime and Justice*. Ed. by D. Duffee. Vol. 4. Washington, DC: National Institute of Justice, 2000, pp. 33–84.

[57] S. G. Van de Weijer, R. Leukfeldt and W. Bernasco. 'Determinants of Reporting Cybercrime: A Comparison Between Identity Theft, Consumer Fraud, and Hacking'. *European Journal of Criminology* 16.4, 2019, pp. 486–508.

[58] P. G. M. Van der Heijden, M. Cruyff, P. A. Smith, P. Bycroft, P. Graham and N. Matheson-Dunning. 'Multiple System Estimation Using Covariates Having Missing Values and Measurement Error: Estimating the Size of the Māori Population in New Zealand'. *Journal of the Royal Statistical Society: Series A* 185, 2022, pp. 156–177.

[59] K. Wittebrood and M. Junger. 'Trends in Violent Crime: A Comparison Between Police Statistics and Victimization Surveys'. *Social Indicators Research* 59.2, 2002, pp. 153–173.

[60] D. W. Woods, R. Böhme, J. Wolff and D. Schwarcz. 'Lessons Lost: Incident Response in the Age of Cyber Insurance and Breach Attorneys'. *32nd USENIX Security Symposium (USENIX Security 23)*. 2023, pp. 2259–2273.

[61] L.-C. Zhang and J. Dunne. 'Trimmed Dual System Estimation'. *Capture–Recapture Methods for the Social and Medical Sciences*. Ed. by D. Böhning, P. Van der Heijden and J. Bunge. Boca Raton: CRC Press, 2018, pp. 229–235.

This page is intentionally left blank.

Part II

Profitability of Ransomware

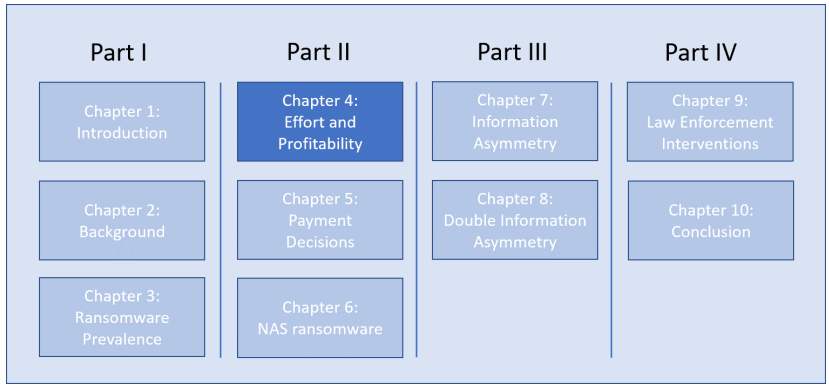
This page is intentionally left blank.

Someone's sitting in the shade today because someone planted a malicious piece of ransomware a long time ago

~ Adapted from Warren Buffet

Chapter 4

Ransomware Offenders and Victims



So far only a few empirical studies have analysed the financial impact of ransomware attacks. This study aims to understand the expected financial gains for attackers and financial losses of victims after a ransomware attack. To do so, we build a dataset based on 453 ransomware attack investigation reports in the Netherlands reported to the Dutch Police between 2019 and 2022. Using rational choice model of crime (RCM) and crime scripting we hypothesise that the effort of an attacker, victim characteristics and context variables influence not only the ransom requested by an attacker but also the financial losses reported by victims. We use generalised linear models to evaluate and quantify this influence.

4.1 Introduction

Europols annual Internet Organised Crime Threat Assessment Report mentions ransomware as top priority [22]. Ransomware (ransom software) is a subset of malware designed to restrict access to a network, system or data until a requested ransom amount from the attacker is paid [55]. Financially motivated attackers see large sums of ransom paid for victims to decrypt and retrieve their systems and files during a ransomware attack. The paid ransom is often only a small part of the financial loss for the victim after a criminal attack [60, 1]. Since the IT infrastructure is down, business continuity is often a problem. Therefore, downtime could be an important factor for financial loss. Furthermore, recovery costs, like buying new hardware and software and hiring specialists to clean and recover the systems, could also be an important contributor to financial loss.

Usually, the aim of a ransomware attack is to obtain a ransom, however, using stolen data from ransomware, attackers can also accomplish various other goals [15]. [15] also describes how stolen data can be used to blackmail the victim: (1) by incrimination, for example by reporting the victim to data protection authorities, (2) by threatening with reputational damage/lost revenue by exposure of sensitive data on the dark web, leading to loss of trust of customers and additional victimisation, (3) by threatening with exposing intellectual property, and (4) by fear of humiliation, for instance by exposing embarrassing information about customers or employees [15]. This data can also be used to derive information to support new attacks, e.g., selling email addresses for phishing campaigns [61].

The rational choice model (RCM) of crime [18, 27] assumes that attackers and victims are rational actors, who weigh the costs of their actions against the benefits in order to make a rational choice. It should be noted that RCM defines the weighing of costs and benefits as rational and this assumption helped in understanding behavioural decisions by malicious actors in different types of crimes, like car-theft [13] and burglary [64]. Using RCM, we hypothesise that increase in effort put in by the attacker in an attack increases their ransom demands. At the same time, victims who are not prepared for a ransomware attack (e.g. do not have appropriate back ups) are more likely to pay the ransom.

In this study our goal is to empirically determine the factors that explain the expected financial gains for ransomware attackers and financial losses of victims after a ransomware attack. Therefore, we state main research question as follows: *what are the factors that contribute to the ransom requested by attackers and financial loss of victims?* To answer this question, we focus on three subquestions:

1. Which factors influence the amount of ransom requested by attackers for the decryption key?

2. Which factors influence the likelihood that victims will pay the ransom?

3. Which factors influence the financial losses of victims reported to the police after an attack?

We analyse 453 Dutch Police Investigation reports of ransomware between January 2019 and July 2022 to collect information on the effort invested by attacker, characteristics of the victim (e.g., yearly revenue, industry sector) and the contextual information regarding the attack (e.g., year and season). To systematically include the factors that contribute to attacker's effort we propose a crime script for a generic ransomware attack. Using generalised linear models (GLM) we test the impact of factors related to rational choice model of crime on the demanded ransom and the likelihood of victims to pay ransom. Our key contributions are:

1. We annotate and analyse 453 Dutch Police Investigation reports describing different ransomware attacks.

2. We show that the amount of effort put by the attacker and yearly revenue of the victim influence the amount of ransom requested by the attacker;

3. We find that along with cost & attacker's effort related variables, the payment of the ransom is determined by the victims being able to recover the encrypted data with backups after an attack;

4. We evaluate the factors that influence the financial loss reported by the victim. We find that factors such as ransom paid, the yearly revenue of the victim and use of RaaS (Ransomware-as-a-Service) by an attacker are statistically significant factors in determining the financial loss reported by a victim after an attack.

The structure of this chapter is as follows: We discuss past literature related to use of cyber crime theories and evaluation of ransomware attacks in §9.2. We introduce the proposed crime script and state our hypotheses in §4.3. Then, we explain composition of our dataset and methodology for analysis in §4.4. Finally, after showcasing our results in §4.5, we discuss our conclusions and future work in §4.6 and §4.7.

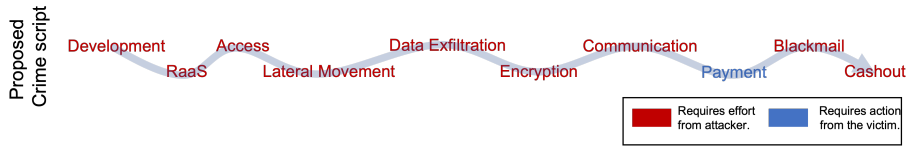


Figure 4.1: The steps of the crime script of a ransomware attack used in this study to structure the data.

4.2 Background Work

Previous work on ransomware has focused mostly on the technical aspects of ransomware [6, 55, 54]. Technical aspects include forensic analysis [42], network detection of command-and-control communication during ransomware attacks [6] and reverse engineering [62]. Several countermeasures have been proposed [42, 62]. E.g., [6] mentions taking advantage of weak encryption techniques used by attacker and improving user awareness to prevent phishing attacks.

A current trend is to study ransomware from other scientific fields, like crime science and economics [55, 39]. Crime science research on ransomware has focused mostly on qualitative impact on victims [16, 7]. [16] surveyed 50 organizations in the UK and North America and studied the factors contributing to the severity of an attack, measured by asking how severe an attack was: low, medium or very severe. The authors did not find a difference in severity between ransomware attacks through phishing versus using exploits as an initial access vector. However, the targeting of victims resulted in more severe attacks than the opportunistic choice of victims.

Economic research on ransomware takes a more theoretical approach [34, 9, 49, 5]. [34] assumes that attackers want to maximize profit and therefore request a ransom which is the trade-off between the probability of a victim paying and maximizing the ransom and therefore profit. One of their results is estimating a demand function of buying the decryption key, where a percentage of the victims would pay a certain price or ransom for returning their files. The authors argue that attackers could maximize profits by estimating the demand function as realistically as possible and subsequently set a ransom which maximizes profits. In this case, it would be beneficial for the attacker to research the victim to estimate the willingness to pay. Their seems to be anecdotal evidence that this happens in practice [59]. However, the authors do not mention which

specific factors explain a high or low willingness to pay, except for the ransom requested [34].

Other studies used game-theoretical models to understand willingness to pay [49, 5]. [49] found data exfiltration to be an important determinant for willingness to pay. Attackers extort the victim by releasing sensitive information online if they do not pay. This gives the victim an incentive to pay the ransom to prevent publication, even if they could recover encrypted files from a backup. [36] explain why victims would pay or not: a cost-benefit analysis of victims between the financial costs of not paying, which is related to downtime costs, and ethical concerns of paying criminals. The author mentions that the most important factor for victims to pay is having recoverable backups or not [36].

In sum, both crime science, as well as economic research, emphasized that attackers' behaviour can be described with a rational choice model of crime (RCM) [18, 27]. The costs and benefits calculations are made by attackers to determine the ransom to request to maximize profit.

Besides RCM, crime science also proposed opportunities as a factor that guides attackers' behaviour [27]. Opportunities are characteristics of the targets. Target or victim characteristics influence costs and benefits calculations. Target or victim characteristics influence costs and benefits calculations. For example, in line with routine activity theory [28, 48], we assume that wealthier victims, that is victims with a high yearly revenue and a large staff, constitute more attractive victims as they are likely to be able to pay a higher ransom. Also, victims with a cyber insurance are relatively attractive as they may not care to pay. The main asset in order to avoid paying for victims is to have a backup that can be restored easily. What may help reduce the damage is hiring an incident response company to avoid paying or pay less. Engaging in lengthy negotiations may also help reducing the ransom that has to be paid. Having one's infrastructure in the cloud also helps to reduce the final ransom. Taken together, victim characteristics could influence attackers' behaviour.

Next to attackers' effort and the context of the attacks might also influence the ransomware attacks [31, 41, 23]. Besides focusing on wealthy victims in order to be able to request large ransoms, other aspects may play a role. As companies become more and more dependent on their digital assets for their business continuity this may lead to an increasing vulnerability, which may lead to a trend of requesting larger ransoms over the years and accordingly, a 'willingness' or need to pay larger ransoms over the years [31]. Second, cybercriminals might be more willing to attack in different seasons [23]. For example, [41] found that seasonality influenced fraud against businesses. This might also occur within the ransomware landscape.

In conclusion, the literature on ransomware attacks has been mostly based on theory, relatively small samples or more qualitative descriptions of ransomware attacks. There is little quantitative empirical research on the risk factors/determinants ransom requested by attackers and financial loss reported by victims after a ransomware attack [15, 50]. This is the focus of the present study. To structure the data and analyse the concepts of opportunity and effort, we propose a crime script of ransomware in the next section.

4.3 Proposed Crime Script and Hypotheses

As described in the previous section, we hypothesize that ransomware attacks are the result of crude cost-benefit calculation by attackers and their assessment of where the good opportunities lie. To understand the costs, risks and opportunities, it is important to consider the ransomware crime script. From the field of crime science, crime scripting is a way to systematically study the procedures, actions and decisions when performing a crime [17, 40, 57]. Similarly, in computer science several authors described a kill chain in which the various steps on performing an attack were described [3, 15]. Previous research on ransomware described taxonomies that sometimes included a series of steps [6, 55, 20] as well as different actors and roles [50]. In the present study we propose to describe ransomware as a simplified step-wise process: ransomware is a complex crime involving many steps, often involving a group that probably comprises several members and sometimes also involves collaboration with other groups. Based on insights from previous research [31, 43, 52, 14], we propose the following global ransomware script (see Figure 4.1):

1) Development: To start with, it is important to organise the infrastructure and develop the malware beforehand [37]. The infrastructure is needed to deliver the malware and to obfuscate network traces from the system of the victim to the attacker [43, 37].

2) RaaS: RaaS and collaboration with other groups. When individuals or groups lack expertise they can make use of Ransomware-as-a-Service (RaaS). In practice, this means hiring the ransomware from other cyber criminals [50, 38]. The term affiliate has been used to describe the actor hiring the ransomware. RaaS enables affiliates with relatively low-technical skills to use advanced ransomware and this makes the attack much easier to launch [6]. RaaS was described as a way to 'democratize crime' [50]. The advantage for the affiliates is, obviously, that it becomes much easier to execute ransomware attacks: all actors involved in an attack could specialize in a specific part of the attack. For example, obtain-

ing credentials from a victim's network or developing malware [38]. However, extra effort may lie in coordinating their work with the RaaS developer and the possibility to have to share a part of the profit. Each actor, ransomware developer or affiliate, do what they can do well, and do not need to do the other party's work. Accordingly, we believe it is a reasonable hypothesis that, overall, RaaS requires less total work for the involved actors.

3) Access: Gain access to a victim's computer or network and maintain that access. To gain access to the victim's system, attackers need to distribute the ransomware. Reference [55] described how this is usually done. Mostly, attackers send a phishing email that contains a malicious file or a link (33%) or they sand spam (8%). Other options are malicious apps, to infect mobile phones (13%); drive-by-download e.g., malicious advertisements (10%); exploit kits (15%) or a Remote Desktop Protocol (RDP: 8%). Vulnerabilities in the victim's platform such as in operating systems, browsers, or software can also be used by ransomware attackers as infection vectors (10%).

4) Lateral movement: It is moving to other computers on the network with the goal to get an impression of the files and gain control over the entire network.

5) Data exfiltration: Although many groups state they exfiltrate data, probably to put pressure on the victim, most ransomware groups do not actually do this. Data exfiltration is a big risk for the victim, as, for instance, their data may end up on the dark web on a 'sucker's list', be sold, or become visible on the open web for everyone to see [49, 15]. Data exfiltration is considered to take more effort than no data exfiltration.

6) Encryption: Performing encryption of the victims files is of course, key to the entire process.

7) Communication: The attackers need to communicate and possibly negotiate with the victim. They also need to provide payment credentials and determine the size on the requested ransom. To this end the attackers can send a ransom note to the victim: they want to first have contact with the victim before informing them what ransom requested is. This gives them the opportunity to change the ransom, depending on victim characteristics like yearly revenue [34, 59]. A personalized ransom note is considered to be more effort for the attackers than a standard ransom note for all victims. Furthermore, the mode of communication also influences the attackers effort: some attackers communicate with their victim through e-mail, others use a self-made *TORchat* application. Using a self-made TOR application requires more work on the attackers' side. Although

within the RaaS ecosystem affiliates often do not need to make the *TORchat* themselves, we would hypothesize that the overall effort increases. Developers of the ransomware might ask for larger ransoms requested by affiliates due to their extra effort put into building the TOR-chat.

8) Payment: At this stage the victim needs to think about paying or not paying. If the victim does not want to pay, for instance because he has a good backup, the attack could stop here. But victims often do not have a useable backup. According to [47] restoring backups is often difficult: 85% fail during restoration attempt. Consequently, at this stage the victim usually starts communicating with the attackers about the ransom. The victim may be willing to pay, but think the ransom is too high, sometimes he is not allowed to pay the ransom, such as some public organizations. To that end, the victim might engage an incident response company that helps negotiating and the payment of the ransom. The ransom after negotiations may depend not only on the requested ransom, but also on the negotiating skill of the victim and/or the incident response company that the victim hired. The experience of the Dutch police is that attackers have an incentive not to take too much time to negotiate: longer negotiations may lead to a lower final ransom and they want to have their money quickly. Asking a ransom that is unrealistically high may increase the negotiation time [58].

9) Blackmail: Different additional extortion methods can be used to put additional pressure on the victim: perform DDoS attacks on the victims website and/or calling or e-mailing clients or employees of the victim's company [56]. It is important to note that the publication of data on a leakpage is also a type of blackmail, but in this study is categorized as *5) Data exfiltration*.

10) Cash-out: Getting the money, laundering it through different mixers or money mules [53]. Additionally, provide the decryption keys to the victim and possibly helping the victims with decryption of their files.

We emphasize that this crime script is a rough description of a ransomware attack and serves the purpose of this study. Further research might generalize this crime script to include more different types of ransomware attacks which are outside of the scope of this chapter.

The crime script presented above is a brief overview of the steps of an complete version of a ransomware attack: not all attacks include all steps. Some ransomware groups are known to perform some of the steps described above. For example, eCh0raix is strain which targets solely Network Attached Storage (NAS) devices [35]. The group(s) behind this strain are known to be RaaS, so

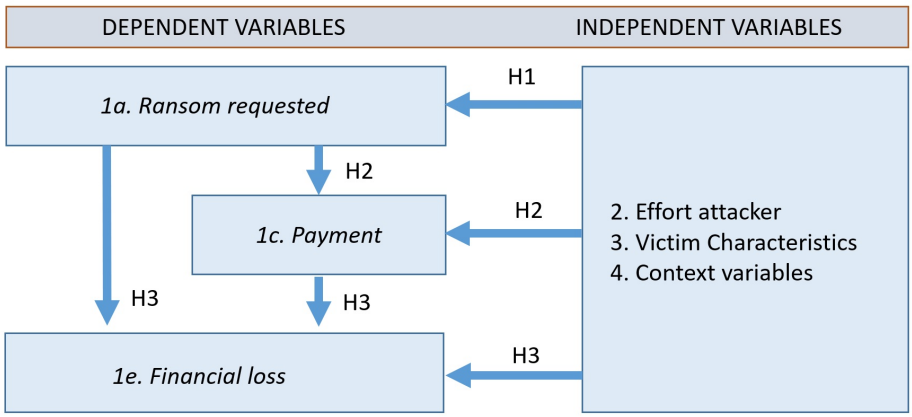


Figure 4.2: Hypothesis within this study. Effort of attacker, victim characteristics, context variables determine ransom requested (H1). Combined they influence whether a victim pays (H2). Financial loss of a victim is determined by ransom requested, paid, effort of attacker, victim characteristics, and context variables (H3).

affiliates can buy the ransomware in exchange for a share of the profits. Furthermore, these attacks are characterized by only encrypting the NAS device and then leaving a ransom note with a fixed ransom for all victims and bitcoin address in exchange for the decrypter keys. So, from the crime script, only steps 1), 2), 3) and 9) are performed. Similarly, the group(s) behind the ransomware strain Conti [19], is known for being RaaS, and perform almost all steps of the mentioned crime script, except for step 9) Blackmail. As we will illustrate in this study, eCh0raix requests smaller amounts of ransom than Conti, as expected from our reasoning above.

Based on the Rational Choice Model of crime (RCM) and the crime script, it is assumed that increasing the costs of an attack must be balanced by larger rewards and/or easier opportunities, and/or smaller risks, otherwise, attackers will not be interested in investing more time and effort. We therefore hypothesize, that when more effort is put into the attack, the result should be a larger ransom requested and larger financial loss for victims. Specifically, we hypothesize (see Figure 4.2):

1. The ransom requested (RR) is the result of a costs-benefits calculation by the attackers, considering opportunities and context (H1). It is expected that more attackers effort leads to larger RR.

2. The decision to pay the ransom is the result of the RR and the costs and benefits of the victim (H2). It is expected that victims who have back-ups and attacks where data has been exfiltrated, leads to larger probability of paying.

3. The losses by the victim are the result of RR, payment and attackers' effort, victim characteristics and context variables (H3). It is expected that large RR, effort by attackers, large companies and payment lead to larger financial loss after a ransomware attack.

4.4 Data and Methodology

Between 1 January 2019 and 1 July 2022 453 ransomware attacks were reported to the Dutch Police. Attacks were collected by searching the police file systems for the keyword 'ransomware'. Subsequently we collected and coded the data using the variables shown in Table 4.1 below. We show the step-by-step methodology in Figure 4.3. Of these 453 police investigation reports, 13 were attempted ransomware attacks where no files were encrypted, in 6 cases no ransomware was found and 81 ransomware attacks on individuals were outside the scope of this study. Combined, investigation reports for 353 attack remain in our dataset and were used for further analysis. Figure 4.4 shows the monthly distribution of reports. [32, 47] state that the number of ransomware attacks have increased substantially in the last few years. In our data only 3 of the reported attacks

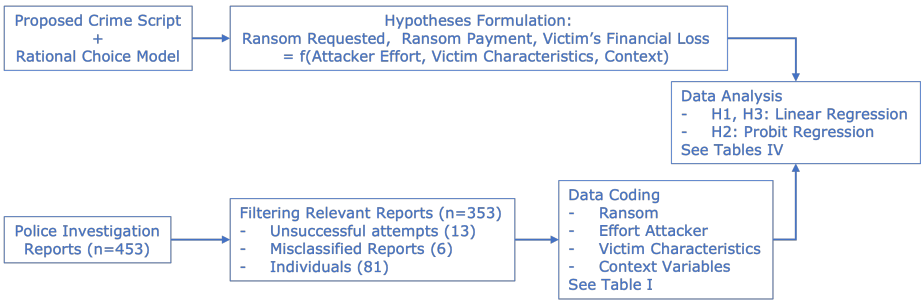


Figure 4.3: Methodology used to analyse police investigation reports.

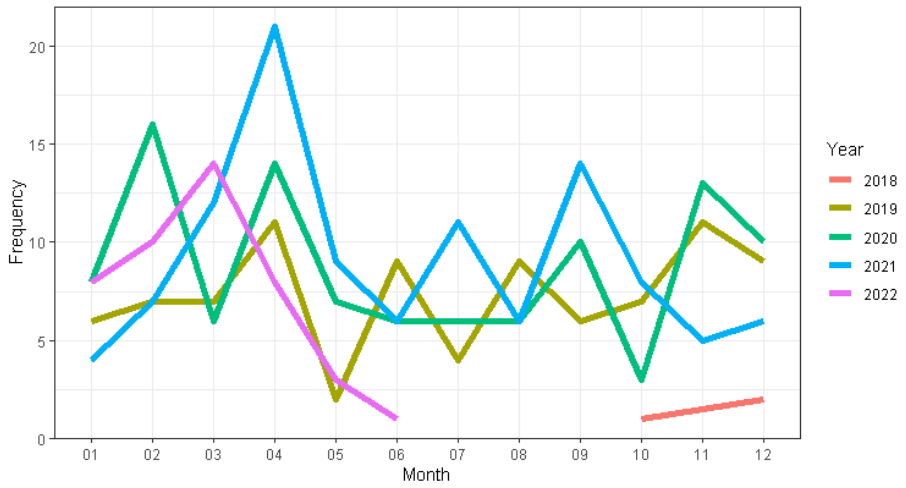


Figure 4.4: Frequency of ransomware monthly attacks based on date of encryption in reports. 3 reported attacks were from 2018, while nearly a 100 attacks are from 2019, 2020 and 2021 each. 44 attacks are reported since beginning of 2022.

were from 2018, while nearly 100 attacks are from 2019, 2020 and 2021 each. Since we use reports made after January 1st, 2019, we do not see a substantial change in the number of reported attacks in these three years.

Next, we describe the variables coded in police investigation reports. The three **dependent variables** in our study (see Table 4.1) are:

1a. *Ransom requested:* The ransom attackers request for the decryption of the victims files, is a good estimation of how much financial gain they hope to make with the attack [34]. This is the ransom requested in the beginning of the ransomware process, before the negotiations (in euro). The reason is twofold: 1) the ransom after negotiations also depends on the negotiating skill of the victim and/or the incident response company that the victim hired and 2) attackers have an incentive to not take too much time to negotiate, because they want to have their money quickly. Asking a ransom far from the financial gain they hope to make, might increase the negotiation time [58].

Table 4.1: Variables used in this study and in different regression analysis. In the first column the variables are depicted: 1) dependent variables, 2) is criminal effort, 3) are victim characteristics, and 4) is context. In the second column the units or categories of a variable. In the third column the amount of missing observations per variable. Finally, the last three columns depict which variables are used for the regression analysis on ransom requested (Y1=RR), payment (Y2=Pay) and financial loss (Y3=FL).

Variable	Unit / categories	Missing values (%)	Y1= RR	Y2=Pay	Y3=FL
1a. Ransom requested start negotiations	Euro, Log 10 transformed	196 (55.5%)	X		
1b. Ransom requested end negotiations	Euro, Log 10 transformed	194 (55.0%)		X	X
1c. Payment	Yes = 1 / no = 0	22 (6.2%)		X	
1d. Ransom paid	Euro, Log 10 transformed	26 (7.4%)			X
1e. Financial loss	Euro, Log 10 transformed	184 (52.1%)			X
2a. Data exfiltration	Yes = 1 / no = 0	229 (64.9%)	X	X	X
2b. Targeted ransomnote	Yes = 1 / no = 0	1 (0.3%)	X	X	X
2c. RaaS	Yes = 1 / no = 0	181 (51.3%)	X	X	X
2d. Strain	Lockbit, Dharma, Conti, Phobos, Sodinokibi, ech0raix, Others	87 (24.6%)	X	X	X
2e. NAS	Yes = 1 / no = 0	1 (0.3%)	X	X	X
2f. Access	Phishing, exploits, different	1 (0.3%)	X	X	X
2g. Blackmail	Attacker contacts employees, customers, other type of pressure	2 (0.6%)	X	X	X
2h. Communication victim-attacker	E-mail, TOR-chat, different	64 (18.1%)	X	X	X
3a. Yearly revenue victim	Euro, Log 10 transformed	25 (7.1%)	X	X	X
3b. Staff at victim's company	Log10 transformed	11 (3.1%)			
3c. Sector	Sectors described by Dutch Chamber of Commerce	1 (0.3%)	X	X	X
3d. Insurance	Yes = 1 / no = 0	28 (7.9%)	X	X	X
3e. Backup	No = 0, yes+no recovery = 1, yes+partial recovery = 2, yes+full recovery = 3	11 (3.1%)	X	X	X
3f. IR company	Yes = 1 / no = 0	244 (69.1%)		X	X
3g. Days negotiating	Days, Log10 transformed	45 (12.7%)		X	X
3h. Repeat victimization	Yes+ransomware = 2 , yes+ other cybercrime = 1, no = 0	314 (89.0%)			
3i. Cloud	No = 0, yes = 1, partially = 2, mitigating = 3	22 (6.2%)	X	X	X
4a. Year	2018/2019/2020/2021/2022	4 (1.1%)	X	X	X
4b. Season	Summer/Autumn/Winter/Spring	4 (1.1%)	X	X	X
4c. Time encryption	Date, time (DDMMYYYYhhmm)	4 (1.1%)			
4d. Time data exfiltration	Date, time (DDMMYYYYhhmm)	325 (92.1%)			
4e. Time access	Date, time (DDMMYYYYhhmm)	264 (74.8%)			

1c. *Payment:* This variable is whether victims would pay or not (categories: yes = 1 / no = 0). In our data set, 21% of victims paid. This is different from the willingness to pay [10]. Some victims might be willing to pay, but think the ransom is too high or they are not allowed to pay the ransom, like public organizations.

1e. *Financial loss:* This is the total financial loss reported by the victim (in euro). Some victims specified different aspects of the costs, e.g., repair costs, reputation costs, liability, and payment of ransom. Nevertheless, most victims only gave a rough estimate of the total costs.

These three dependent variables, are log-transformed. This transforms the non-linear distribution to get an approximately normal distributed variable, as is common in social-empirical studies [11]. The logarithm base 10 is chosen to increase the readability of figures.
The **independent variables** in this study (See Table 5.1) are:

1b. *Ransom requested end negotiations:* To understand if ransom requested influences payment, it is important to consider the amount of ransom which was requested after negotiations (in euro), since this is the amount the victim needs to pay.

1d. *Ransom Paid:* To study the factors influencing financial loss, the ransom paid to the attackers has been used as a dependent variable (in euro). This was constructed as a function multiplying payment (*1c.Payment*) and final ransom (*1b.Ransom requested end negotiations*).

2. Effort attacker. To measure effort information was collected on several variables.

2a. *Data exfiltration:* Exfiltrated of data measured whether data from the victim were exfiltrated (categories: yes = 1 / no = 0). Although many groups state they exfiltrate data, probably to put pressure on the victim, most ransomware groups do not. We reported a confirmed data exfiltration when analysis of the network logs has been performed and unusual large amount of data uploading was found or when the victims data has been published on a leak page and the data is identified of being from the victim. Data exfiltration is considered more effort than no data exfiltration.

2b. *Targeted ransom note:* We noted whether the criminals wanted to first have contact with the victim before informing them what ransom they requested, which we define as targeted ransom note (categories: yes = 1 / no

= 0). Yes means that first contact with the attackers was required to obtain information about the ransom. No means that the ransom was stated on the ransom note. A personalized ransom note is considered more effort than a standard ransom note for all victims. To our knowledge did a personalized ransom note not yet lead to identification of the attackers.

2c. *RaaS:* Collaboration with other criminals, measures whether the attackers made use of Ransomware-as-a-service (RaaS) [9] or whether they collaborated with other groups to perform the attack (categories: yes = 1 / no = 0). RaaS is considered to be more effort than groups who do not perform RaaS.

2d. *Strain:* The name of the ransomware strain found on the victims encrypted files. Often this is the extension. The attacks from a specific strain that executed the attack were included if more than 5 attacks were observed, the rest was aggregated to the variable 'Others'. This is due the sensitivity of the data. We assume that groups behind strains vary in the amount of effort used in attacks, and therefore might also vary in the required ransom. This variable accounts for all variance due to factors not labelled in this study but are different between strains or groups.

2e. *NAS:* Network Attached Storage measures whether attackers targeted a Network Attached Storage device (NAS, categories: yes = 1 / no = 0).

2f. *Access:* What type of access was used to infiltrate the victims network (categories: exploit/phishing/different).

2g. *Blackmail:* Whether attackers contacted the victim or clients of the victim to exert additional pressure on the victim to pay (categories: yes = 1 / no = 0).

2h. *Communication victim-attacker:* Whether victim and attacker communicated through e-mail, a self-made *TORchat* application, or differently (categories: TOR/e-mail/different). A self-made *TORchat* application is considered more effort than e-mail.

3. Victim characteristics. To measure opportunity, information was collected on several other variables:

3a. *Yearly revenue:* Yearly revenue victim in euro.

3b. *Staff:* Staff working at victim's company.

3c. *Sector:* Economic sector of the victim's company, as categorized by the Dutch Chamber of Commerce.

3d. *Insurance:* Whether the victim has insurance which covers ransomware attacks (categories: yes = 1 / no = 0).

3e. *Backup:* Whether there were backups and the state of the backups (categories: no = 0, yes but not possible to recover of data = 1, yes but could partially recover data = 2, yes and could fully recover data = 3).

3f. *IR company:* If an Incident Response company helped the victim recover from the attack or/and negotiate with the attackers to get the decrypter (categories: yes = 1 / no = 0).

3g. *Days negotiating :* Amount of days negotiating in logarithm.

3f. *Repeat victimization:* Whether the victim has experienced a ransomware attack before, or another type of cybercrime (categories: yes+ransomware = 2, yes+other cybercrime = 1, no = 0).

3i. *Cloud:* Whether the victim has their IT infrastructure in the cloud (categories: no = 0, yes = 1, partially = 2, mitigating = 3).

4. Context variables. To measure the context of the attack, information was collected on the following variables:

4a. *Year:* Year of encryption*

4b. *Season:* Categories: Summer, autumn, winter, spring.

4c. *Time encryption:* Full date and time of encryption.

4d. *Time data exfiltration:* Full date and time of the stealing and exfiltrating of data of the victim.

4e. *Time access:* Full date and time when the first malicious activity on the target network was recorded.

To impute the missing observations we use Multiple Imputation Chained Equations (MICE) method [8, 44, 33], which is a more reliable than list wise deletion or simple imputation methods [29]. For a good explanation of how

*Note that this can be before 1st of January 2019, since encryption occurs before reporting it to the police. We filter based on the date it was reported to the police.

MICE works, we refer to [29]. The MICE method still gives reliable estimates with 60% missing variables [44]. We omit variables with more than 70% missing observations from our analysis: repeat victimization (3h), the time between access and encryption (4d), and time between data exfiltration and encryption (4e).

Analysis were conducted using R version 4.0.2, packages *MICE*, *ggplot* and *dplyr*. To test the hypothesis (see Figure 4.2) a subset of the variables were used, as depicted in the final three columns of Table 5.1:

- H1: The factors influencing ransom requested. The variables in the 'Y1=RR' column were used as independent variables to perform linear regression analysis on the variable *1a. Ransom requested start negotiations*.
- H2: The factors influencing payment. The variables in the 'Y2=Pay' column were used as independent variables in probit regression analysis on the variable *1c. Payment*.
- H3: The factors influencing ransom requested. The variables in the 'Y3=FL' column were used as independent variables to perform linear regression analysis on the variable *1e. Financial loss*.

For all three models (Y1=RR, Y2=Pay and Y3=FL), backward stepwise selection was performed to find the best fitting model, using the *step* function in R. Stepwise selection is a method to find the best performing model by iteratively adding and removing predictors [63].

We model the ransom requested a the start of negotiations (Y1) and financial loss (Y3) with a Generalized Linear Model (GLM) with family Gaussian. Payment (Y2) is modeled as a Generalized Linear Model with family Probit. The specific choice for using Gaussian GLM is due to the dependent variable constituting a specific amount of money for (Y1) en (Y3). The Probit GLM is used for Y2 because it has a binary outcome variable. Furthermore, as described in [45] our observations might possibly also have interdependence of events and non-equal mean and variance of the dependent variable. A general model for GLM is defined as follows[25]:

$$Y_i = \beta_i x_i + ... + \delta_i \tag{4.1}$$

where i refers to the different observations, β_i are the estimated coefficients for x_i, x_i are the independent variables collected for the observations as described in Table 5.1 and δ_i is the residual. After the GLM, we group the dummy's of the different nominal variables and perform a Likelihood-ratio test [24] to determine

Table 4.2: Descriptive statistics of victim companies of different sectors. Mean and median revenue are in million euros, insured, no backup, and paid are percentages. Financial Loss and ransom is in thousand euros.

	Sector	Number of attacks	Mean Revenue (Meuro)	Median Revenue (Meuro)	(%) Insured	(%) No Backup	Financial Loss (euro)	(%) Ransom Paid	Ransom Requested (euro)
1	Construction	53	562.84	2.43	10.2	35.3	256,410	27.5	182,840
2	Healthcare	21	37.62	2.33	10.5	42.9	77,690	26.3	23,770
3	Trade	113	133.96	2.84	4.9	38.9	737,610	25.5	1,106,800
4	ICT	60	120.59	3.81	13	30.8	232,580	30.9	1,343,190
5	MAS	12	376.36	0.63	0	18.2	12,500	9.1	13,700
6	Media	20	142.54	3.30	0	52.9	344,800	15.8	11,640
7	Education	14	101.43	19.44	0	14.3	49,800	21.4	555,660
8	Government	10	60.17	18.45	10	20	393,330	0	820,350
9	Leisure	20	6.61	1.08	15	55	27,000	15	81,020
10	Transport	29	389.05	6.00	7.4	34.6	838,85	30.8	529,540

the effect of the different variables. A p-value of 0.05 or lower supports the hypothesis that the variable is a significant predictor for the dependent variable with significance level $\alpha = 0.05$.

4.5 Data Analysis and Results

We analyse the data and interpret the results using the following 4 steps: First, we give a general overview of the data with the help of descriptive statistics. Second, we present our analysis for the three hypotheses. Third, we identify and discuss factors that contribute to the ransom payment. Last, we examine the financial loss reported by companies after a ransomware attack.

4.5.1 Descriptive Statistics

Descriptive statistics for a subset of independent variables are shown in Tables 4.2 and 4.3 below. Table 4.2 gives an overview of the different victim characteristics, grouped by sector. Companies within the industry sector *Trade* experienced most ransomware attacks (113 attacks). MAS (Milieu en Agrarische Sector, agriculture) was the fewest with 10 attacks. In the construction sector companies with the largest revenue faced ransomware attacks: 562 million euro. Leisure the least: 6,61 million euros was the average yearly revenue for companies who faced a ransomware attack. However, if we consider the median, education, and

Table 4.3: Descriptive statistics of the different ransomware strains. Mean and median revenue in million euros, insured, no backup, and paid are percentages. Damage and ransom are in euros.

	Strain	Number of attacks	Mean Revenue (Meuro)	Median Revenue (Meuro)	(%) Insured	(%) No Backup	Financial Loss (euro)	(%) Ransom Paid	Ransom Requested (euro)
1	Conti	19	437.43	30.71	28.6	15.8	4,726,280	16.7	6,598,380
2	Dharma	14	7.68	3.98	0	35.7	29,010	50	18,760
3	Others	143	327.26	3.74	4.7	28.6	298,610	27.7	542,330
4	eCh0raix	10	1.27	0.6	11.1	50	2,620	20	750
5	Lockbit	16	23.7	4.57	18.8	43.8	184,380	20	98,980
6	Phobos	32	261.85	1.07	3.2	51.7	167,560	30	21,190
7	Sodinokibi	32	56.43	3.56	16.7	21.9	170,070	18.5	658,010

government had the largest yearly revenue. Companies in the Leisure sector were most often insured with 15 %, and not reported in MAS, Media, and education with 0%. Finally, the average ransom was largest in the ICT sector, with 1.3 million euros, and lowest for the media, with 11,000 euros on average.

In Table 4.3 we present an overview of different attacking strains. Most attacks were performed by group(s) behind Phobos and Sodinokibi (32 times). However, the attacks associated with Conti targets the companies with the largest mean and median revenue: respectively 437 million and 31 million euro. The strain 'Others', contains all other groups. Compared to the groups mentioned here, they target relatively large companies with 327 million euro on average, or 3.7 million euro median.

The final three columns of Tables 4.2 and 4.3 illustrate the dependent variables: financial loss, payment and ransom requested. It is noteworthy that in Table 4.2 the trade and ICT sector have the largest ransom requested, both averages are larger than 1 million euros. Payment is largest in ICT, and transport. Financial loss was largest in trade. This might be due to downtime costs: for companies who sell products or offer services the downtime costs might be highest. Compared to MAS (agriculture) or construction, where work probably could continue without the immediate use of computers.

Considering the different strains in Table 4.3, we find the largest ransom requested by Conti, also the highest financial loss. Dharma has the largest amount payed, perhaps because of the low amount of ransom requested compared to other groups. Ech0raix is the group that targets the smallest companies with

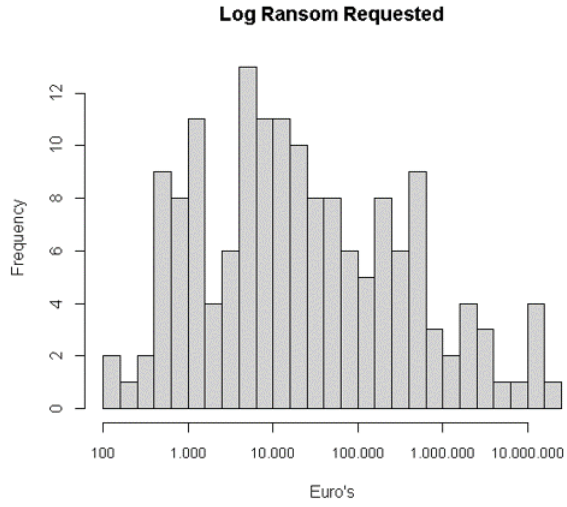


Figure 4.5: Distribution of log ransom requested before negotiations.

1.27 million euro yearly revenue on average, demanding 750 euro and reported financial loss of 2,620 euros. These results seem to be in line with the relationship which was described in the Introduction: more effort, as defined by the crime script, should lead to larger ransom requested by attackers and financial loss by victims.

4.5.2 Hypothesis testing

Next, we use regression analysis to test the three hypotheses introduces in Section 4.3. We use a linear regression model to test **H1** and **H3** and use a probit regression to test **H2**. We discuss the details of our hypothesis testing methodology in Section 4.4. The likelihood ratios for each of the variables tested for the three hypotheses are show in Table 4.4.

Based on the GLM likelihood ratios for **H1** we find that variables that capture attacker's effort such as 'Data exfiltration', use of 'RaaS', 'Blackmail' and active 'Communication between attacker and victim' are all significant factors in predicting the requested ransom. The median ransom requested when the communication was made using *TORchat* was 21K euros, whereas the median ransom requested when other communication channels were used was nearly

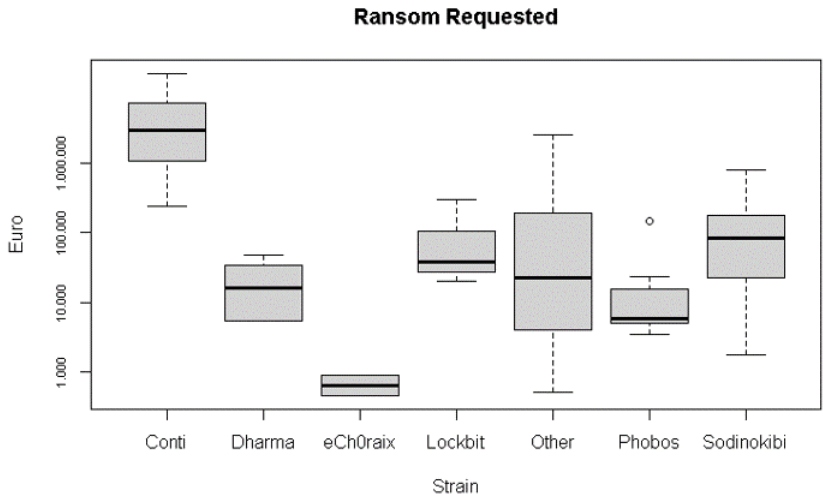


Figure 4.6: Boxplot of log ransom requested for each ransomware strain.

3.5K euros. Also, factors that increase the perceived benefits for attacks such as 'yearly revenue' of the victim firm, 'industry sector' and 'Insurance' were also statistically significant in predicting the ransom requested. Figures 4.5 and 4.6 illustrate respectively the distribution of log ransom requested and the ransom requested when malware strain from a particular group was used.

While analysing the factors that influence the likelihood of victims to pay ransom (**H2**), we find that cost and attacker's effort related variables such as 'ransom requested end negotiations', 'Data exfiltration', 'Targeted ransomnote', 'Blackmail' and 'Days negotiating' were statistically significant predictors. Interestingly, victim characteristic related variable capturing its 'Backup' status was also significant, 28% of the victims with no back ups pay ransom, where as 48% of those whose backups became unrecoverable after ransomware infection pay the ransom. 16% of the victims who has partial recovery of the backups paid ransom, while only 6% of the victims with fully recoverable backups paid the ransom.

In analysis of **H3** we evaluate the factors that explained the loss reported by victim firms. Paid ransom formed a significant part of the reported losses. We again find that attacker effort related factors such as 'Data exfiltration', 'Targeted

Table 4.4: Results of regression analysis

Regression			**Regression**		
Variables	Likelihood Ratio	Df	Variables	Likelihood Ratio	Df
Y1=RR			3d. Insurance	11.00	1
2a. Data exfiltration	96.47**	1	3e. Backup	20.72**	3
2b. Targeted ransomnote	0.57	1	3f. IR company	12.00	1
2c. RaaS	32.99**	1	3g. Days negotiating	55.34**	1
2d. Strain	49.81	6	3i. Cloud	567.00	1
2e. NAS	0.9	1	4a. Year	5.46	4
2f. Access	19.61	3	4b. Season	167.00	3
2g. Blackmail	35.31*	1	**Y3=FL**		
2h. Communication victim-attack	114.76**	3	1b. Ransom requested end negotiations	552.00	1
3a. Yearly revenue victim	37.82*	1	1d. Ransom paid	34.01**	1
3c. Sector	162.42*	9	2a. Data exfiltration	6.10*	1
3d. Insurance	76.29**	1	2b. Targeted ransomnote	3.15*	1
3e. Backup	40.93	3	2c. RaaS	5.46*	1
3i. Cloud	0.02	1	2d. Strain	2.14	6
4a. Year	37.05	4	2e. NAS	289.00	1
4b. Season	21.16	3	2f. Access	2.15	3
Y2=Pay			2g. Blackmail	3.63*	1
1b. Ransom requested end negotiations	9.74**	1	2h. Communication victim-attack	2.29	3
2a. Data exfiltration	8.83**	1	3a. Yearly revenue victim	40.14**	1
2b. Targeted ransomnote	4.27*	1	3c. Sector	6.44	9
2c. RaaS	40.00	1	3d. Insurance	88.00	1
2d. Strain	5.21	6	3e. Backup	7.93*	3
2e. NAS	507.00	1	3f. IR company	271.00	1
2f. Access	3.41	3	3g. Days negotiating	44.00	1
2g. Blackmail	19.03**	1	3i. Cloud	1.08	1
2h. Communication victim-attack	3.84	3	4a. Year	4.83	4
3a. Yearly revenue victim	0.00	1	4b. Season	13.22**	3
3c. Sector	13.47	9			

Note. All data is rounded to 2nd significant figure.
*$p \leq .05$. **$p \leq .01$.

ransomnote' and use of 'RaaS' again significantly affected the amount of reported losses. The median financial losses reported were the lowest when no or full backup was available. This shows that victims that did not already have a backup as part of their resilience strategy, perceived the financial impact of such attacks to be low. While, one with full and recoverable backups were able to hit the ground running without suffering huge losses.

Financial loss of victims (see Figure 4.7) is influenced by the yearly revenue of the victim, the amount of ransom paid, whether the attacking group is known to be RaaS and whether an Incident Response company helped the victim to recover from the ransomware attack.

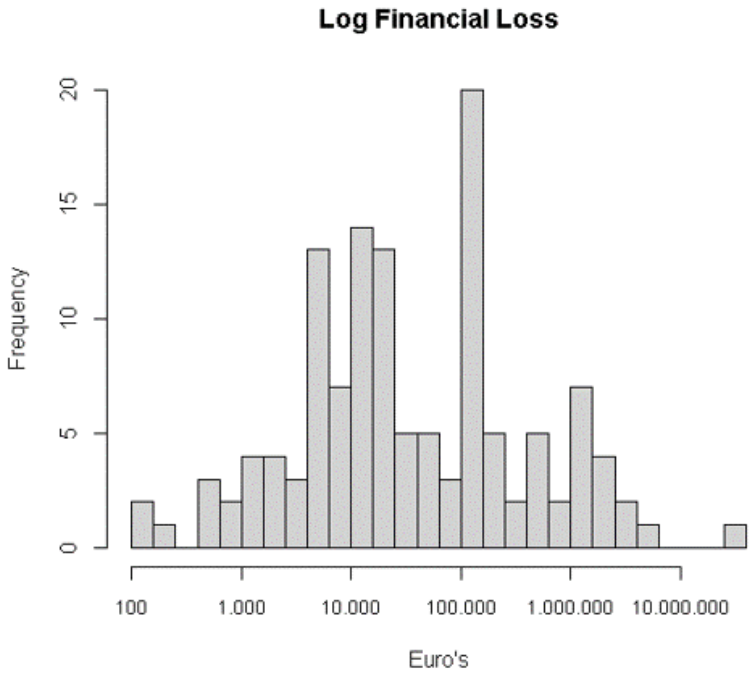


Figure 4.7: Distribution of log financial loss for victims after a ransomware attack.

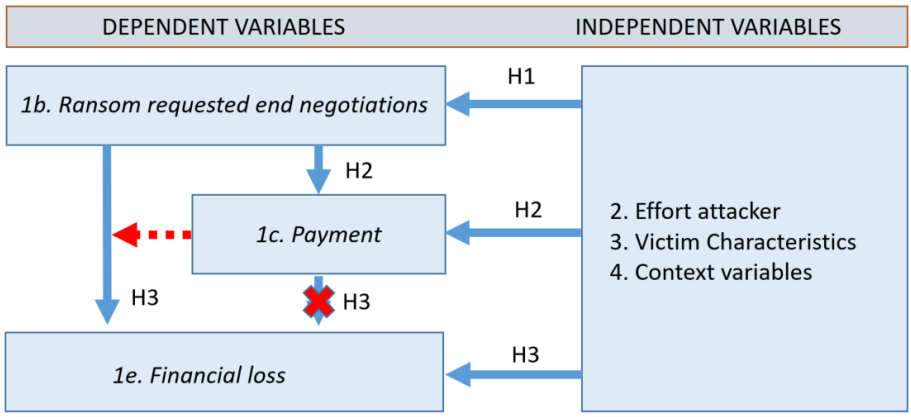


Figure 4.8: Compared to Figure 4.2, our results support the hypothesis that attackers' effort, victim characteristics and context variables influence ransom requested, payment and financial loss. Furthermore, variable '1c. Payment', that has an interaction effect, along with '1a. Ransom requested' is also an important factor for determining the financial loss for victims after a ransomware attack.

Finally, we analysed the direct effect of ransom requested on payment (1), ransom requested, and payment on financial loss (2). Performing a probit regression of ransom requested on payment led to insignificant results ($\beta = -0.1284$, p $= 0.33$). Regression ransom requested and payment on financial loss, we found ransom requested predicts financial loss ($\beta = 0.82$, p$<$0.001), but payment variable itself had no effect on financial loss ($\beta = 0.13$, p $= 0.38$). This result might also be due to how financial loss is operationalized. Victims report to the police financial loss often a couple of weeks after the attack started. At that point, downtime (or other) costs might be not that much different between victims who paid or did not pay.

4.6 Discussion and Conclusion

This study set out to examine the relationship between ransom requested, payment, and financial loss. We examined 353 ransomware attacks reported to the Dutch Police. Based on the RCM and ransomware crime script, we argued that attackers' effort, victim characteristics and context variables are important factors to understand the ransom requested by attackers, whether victims paid the ransom and the financial loss reported by the victims after the attack.

For the ransom requested, we found that data exfiltration, RaaS groups, insurance and mode of communication are important predictors for ransom requested. These results support the hypothesis that attackers' effort and victim characteristics are important factors for the ransom requested by attackers. Furthermore, yearly revenue, blackmail and sector had a possible effect on ransom requested. Finally, these results show that effort could be quantified considering a crime script of ransomware.

For payment, the ransom requested after negotiations, data exfiltration, targeted ransomnote, blackmail, backup, days of negotiating best predict whether victims pay the ransom. We find these results in agreement with the rational choice model of crime. Our results do not indicate a difference between victim characteristics like yearly revenue and sector for the probability of paying. Victim behaviour could be more important: did they have a backup and how long did they negotiate? The effort of the attacker also influenced the decision of the victim to pay: Data exfiltration, targeted ransom note and blackmail are positively related to the probability of payment.

For financial loss we found ransom paid, data exfiltration, RaaS, yearly revenue of the victim, backup and season to be important factors contributing to financial loss as reported by the victim to the police. It is important to note that none of the victims was able to indicate the costs of reputation damage and liability or costs in the long term to the police, because they needed to disclose a (realistic) financial loss when reporting the attack to the Police. Nevertheless, these results seem to support the hypothesise that financial loss is determined by attackers' effort, victim characteristics, context and the amount of ransom paid. Interestingly, whether the victims did or did not pay the ransom, did not contribute to the reported financial loss.

Finally, considering the direct relationships between dependent variables (see Figure 4.8), a direct relationship between ransom requested and financial loss was found. Furthermore, a direct effect of ransom requested after negotiations on payment and payment to financial loss was also found statistically significant.

4.7 Limitations and Further Work

There are different limitations of this study:

1. We collected data based on ransomware reports filed by companies and individuals to the Dutch Police. The nature of this data makes it a challenge to generalize the results to other countries and victims who do not report the ransomware attack to the Dutch Police. These challenges could be

tackled by collecting data from multiple Law Enforcement agencies around the world and incident response companies, for example. These companies also help victims who do not report the attack to the police, making it possible to estimate selection bias due to the willingness to report.

2. The crime script of ransomware as described in this chapter sets out to understand and structure the collected data. However, it is possible to improve this crime script by including more ransomware attacks and from different countries.

3. The regression models could be biased due to the large number of missing observations. Although the MICE method is, to our understanding, a good way to impute the missing data, the models could be improved by decreasing the amount of missing observations. One way to achieve this is by training police officers to ask for more and specific information when the victim reports a ransomware attack.

4. Due to the sensitivity of the data, only one annotator could code the data. This might result in several types of biases [2]. This problem was important when coding the categorical variables. For example, a company that sells buses, should it belong to trade or to transport? We tried to limit the severity of this limitation by anonymously discussing these issues with experts outside the project and writing down the choices to improve consistency. In this specific case we decided that selling buses belongs to trade, since that is the main objective of the company. Further research could address this issue by asking permission for multiple researchers to get access to the sensitive data from the start of the research project.

To understand how sample bias might have affected this study, we compare sample size of ransomware attacks and percentage payments with previous literature. Considering the sample size of ransomware attacks in other studies, [21] examined 623 ransomware incidents in the EU, United Kingdom and United States between May 2021 and June 2022. [46] examined 101 ransomware attacks in 2020 in 81 countries. Comparing these two studies with the present study, we examined a relatively large sample size: around 100 cases within one year in the Netherlands. Furthermore, comparing the sample of this study with research from the industry [51], it seems the sample might contain more cases from individuals and small and medium enterprises, since they perhaps cannot afford incident response services after a ransomware attack.

Considering the percentage payments,[12] indicates 85% of the victims pays, but this is based on 13 observations and was in 2016. [14] surveyed 41 companies in the UK between 2014 and 2018, of which 8 companies (19.5%) paid the ransom. Victims were sampled from UK Police data. The percentages in these two studies are based on small samples. Finding of [14] aligns with our study. As described earlier, it might be that victims who pay are less inclined to go to the Police to report their attack. Perhaps because the Police expresses the strong view to never negotiate or pay (ransomware) criminals. Payment rates from the industry seem to confirm this. According to [51], Kaspersky found that in 2020 52% of ransomware victims paid. It would be interesting to reproduce this study with data from incident response companies and to survey companies when they would go to the Police. As is, this was mostly studied considering other cybercrimes [30, 65] but not ransomware. Taken together, these results indicate that not all companies report to the police and that victims that pay are less willing to report to the police.

One other interesting finding in the present study is that 6 ransomware strains account for almost 50% of the cases. This seems to align with previous offline crime research: there has been a concentration of offending and offenders in time and space [26]. However, other ransomware research did not found such a strong concentration [21, 46].

In conclusion, this study is the first attempt to do a large-scale empirical study. Despite its limitations, the relatively large sample size [12, 14, 51] made it possible to study the effort of the attacker, victim characteristics and context variables in depth and their influence on the ransom requested, the payment of the ransom and the financial loss reported by the victim.

Furthermore, this study might support interventions by Law Enforcement and policy makers. Law Enforcement could intervene on the factors which influence the ransom requested, to reduce the amount of money attackers make with ransomware attacks. Policy makers could conduct targeted prevention campaigns to companies in specific sectors and large companies, as these characteristics seem to indicate larger ransom requested and therefore more profitable for attackers. These campaigns could be increased during specific seasons, as this was an indicator for the financial loss of victims. Victims who are under attack could be warned and be prepared for potential blackmail strategies and publication of confidential data on leakpages. Finally, prevention campaigns could focus on prevention: make sure that potential victims have reliable backups, which are not accessible through the network by attackers. Backups decrease the probability of paying and therefore decreases the financial gains of ransomware attacks.

4.8 Ethics

We follow the principles from Menlo Report [4] to justify the ethical considerations made in this study:

Respect for persons: Privacy of victims was taken into considerations when writing this chapter. By not considering individual cases and only aggregating to strains and sector of victims, we feel confident the privacy of victims is respected.

Beneficence: Information of the police investigations was only available to researcher who had a proper police screening. For the other researchers involved in this project only aggregated results were available. Although this conflicts with the scientific principles of transparency and reproduceability, this seemed the only way to conduct a large-scale empirical ransomware study. Furthermore, results presented in this chapter should exclude personal identifiable information.

Justice: Selection of ransomware attacks was only on the keyword 'ransomware' in the police systems. In this way, all ransomware attacks got an equal chance to be part of the study. No extra effort was put into attacks which got a lot of media attention.

Respect for Law and Public Interest: An important factor we took into consideration was the information position regarding specific groups and/or strains or the way the Dutch Police operates. These were excluded from the chapter.

4.9 Acknowledgements

We would like to extend our sincere gratitude to the Dutch Police. In particular, we would like to thank Emma Ratia, Theo van der Plas and Cees van Tent for making the project possible. Furthermore, we thank the Cybercrime Unit East Netherlands and the Ransomware Taskforce for their expertise. Please note that the views expressed in this work are ours alone and do not necessarily reflect those of the Dutch Police. Finally, we would like to thank our shepherd, Laurin Weissinger, and the anonymous reviewers for their suggestions to improve this chapter.

Bibliography

[1] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. V. Eeten, M. Levi and S. Savage. 'Measuring the cost of cybercrime'. *The Economics of Information Security and Privacy*. Berlin, Heidelberg: Springer, 2013, pp. 265–300.

[2] R. Artstein and M. Poesio. 'Bias decreases in proportion to the number of annotators'. *Proceedings of FG-MoL 2005: The 10th conference on Formal Grammar and The 9th Meeting on*. Vol. 139. 2009.

[3] P. Bahrami, A. Nikkhah, T. Dehghantanha, T. Dargahi, R. M. Parizi, K. C. Choo and H. H. S. Javadi. 'Cyber Kill Chain-Based Taxonomy of Advanced Persistent Threat Actors: Analogy of Tactics, Techniques, and Procedures'. *Journal of Information Processing Systems* 15.4, 2019, pp. 865–889.

[4] M. Bailey, D. Dittrich, E. Kenneally and D. Maughan. 'The menlo report'. *IEEE Security & Privacy* 10.2, 2012, pp. 71–75.

[5] R. P. Baksi. 'Pay or Not Pay? A Game-Theoretical Analysis of Ransomware Interactions Considering a Defender's Deception Architecture'. *52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S)*. IEEE. 2022, pp. 53–54.

[6] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak and M. K. Khan. 'Ransomware: Recent advances, analysis, challenges and future research directions'. *Computers & Security* 111, 2021, p. 102490.

[7] H. Borrion and L. Y. Connolly. *Your money or your business: Decision-making processes in ransomware attacks*. 2020.

[8] S. V. Buuren and K. Groothuis-Oudshoorn. 'mice: Multivariate imputation by chained equations in R'. *Journal of Statistical Software* 45, 2011, pp. 1–67.

[9] A. Cartwright and E. Cartwright. 'Ransomware and reputation'. *Games* 10.2, 2019, p. 26.

[10] A. Cartwright, E. Cartwright, L. Xue and J. Hernandez-Castro. 'An investigation of individual willingness to pay ransomware'. *Journal of Financial Crime*, 2022. Ahead-of-print.

[11] F. Changyong, W. Hongyue, L. Naiji, C. Tian, H. Hua and L. Ying. 'Log-transformation and its implications for data analysis'. *Shanghai Archives of Psychiatry* 26.2, 2014, p. 105.

[12] K. Choi, T. Scott and D. LeClair. 'Ransomware against police: diagnosis of risk factors via application of cyber-routing activities theory'. *International Journal Forensic Science Pathol* 4, 2016, pp. 253–258.

[13] R. Clarke and P. Harris. 'A rational choice perspective on the targets of automobile theft'. *Criminal Behaviour and Mental Health* 2.1, 1992, pp. 25–42.

[14] A. Connolly and H. Borrion. 'Reducing Ransomware Crime: Analysis of Victims' Payment Decisions'. *Computers & Security*, 2022, p. 102760.

[15] L. Y. Connolly, M. Lang, P. Taylor and P. Corner. *The Evolving Threat of Ransomware: From Extortion to Blackmail*. 2021.

[16] L. Y. Connolly, D. Wall, M. Lang and B. Oddson. 'An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability'. *Journal of Cybersecurity* 6.1, 2020, tyaa023.

[17] D. Cornish. 'The procedural analysis of offending and its relevance for situational prevention'. *Crime Prevention Studies* 3.1, 1994, pp. 151–196.

[18] D. Cornish and R. Clarke. 'Crime specialisation, crime displacement and rational choice theory'. *Criminal behavior and the justice system*. Berlin, Heidelberg: Springer, 1989, pp. 103–117.

[19] T. Cymru. *Analyzing ransomware negotiations with CONTI: An in-depth analysis*. 2022.

[20] J. DiMaggio. *Ransom Mafia: Analysis of the World's First Ransomware Cartel*. Analyst1, 7 April. 2021.

[21] ENISA. *Ransomware: Publicly Reported Incidents are only the tip of the iceberg*. Available from: https://www.enisa.europa.eu/news/ransomware-publicly-reported-incidents-are-only-the-tip-of-the-iceberg [accessed November 2022]. 2022.

[22] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2021*. Publications Office of the European Union, Luxembourg. Retrieved August 31, 2022, from https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021. 2021.

[23] G. Falk. 'The influence of the seasons on the crime rate'. *J. Crim. L. Criminology & Police Science* 43, 1952, p. 199.

[24] J. Fan, C. Zhang and J. Zhang. 'Generalized likelihood ratio statistics and Wilks phenomenon'. *The Annals of Statistics* 29.1, 2001, pp. 153–193.

[25] J. J. Faraway. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC, 2016.

[26] G. Farrell. 'Crime concentration theory'. *Crime prevention and community Safety* 17.4, 2015, pp. 233–248.

[27] M. Felson and R. V. Clarke. *Opportunity makes the thief*. Police Research Series, Paper 98. Pages 1-36. 1998.

[28] M. Felson and L. E. Cohen. 'Human ecology and crime: A routine activity approach'. *Human Ecology* 8.4, 1980, pp. 389–406.

[29] J. V. Ginkel, M. Linting, R. Rippe and A. van der Voort. 'Rebutting existing misconceptions about multiple imputation as a method for handling missing data'. *Journal of Personality Assessment* 102.3, 2020, pp. 297–308.

[30] A. Graham, T. C. Kulig and F. T. Cullen. 'Willingness to report crime to the police: Traditional crime, cybercrime, and procedural justice'. *Policing: An International Journal*, 2019.

[31] N. A. Hassan. *Ransomware revealed: a beginner's guide to protecting and recovering from ransomware attacks*. Apress, 2019.

[32] A. Haymore. *We Wait, Because We Know You. Inside the Ransomware Negotiation Economics*. Retrieved from https://research.nccgroup.com/2021/11/12/we-wait-because-we-know-you-inside-the-ransomware-negotiation-economics/. 2021.

[33] K. Heidt. 'Comparison of imputation methods for mixed data missing at random', 2019.

[34] J. Hernandez-Castro, A. Cartwright and E. Cartwright. 'An economic analysis of ransomware and its welfare consequences'. *Royal Society Open Science* 7.3, 2020, p. 190023.

[35] S. Hilt and F. Merces. *Backing Your Backup*. 2022.

[36] T. Hofmann. 'How organisations can ethically negotiate ransomware payments'. *Network Security* 2020.10, 2020, pp. 13–17.

[37] D. Huang, M. Aliapoulios, V. Li, L. Invernizzi, K. McRoberts, E. Bursztein, J. Levin, K. Levchenko, A. Snoeren and D. McCoy. 'Tracking Ransomware End-to-end'. *39th IEEE Symposium on Security and Privacy, S & P*. 2018, pp. 618–631.

[38] K. Huang, M. Siegel and S. Madnick. 'Systematically Understanding the Cyber Attack Business: A Survey'. *ACM Comput. Surv.* 51.4, 2019, Article 70.

[39] M. Humayun, N. Z. Jhanjhi, A. Alsayat and V. Ponnusamy. 'Internet of things and ransomware: Evolution, mitigation and prevention'. *Egyptian Informatics Journal* 22.1, 2021, pp. 105–117.

[40] A. Hutchings and T. J. Holt. 'A crime script analysis of the online stolen data market'. *British Journal of Criminology* 55.3, 2015, pp. 596–614.

[41] M. Junger, V. Wang and M. Schlömer. 'Fraud against businesses both online and offline: Crime scripts, business characteristics, efforts, and benefits'. *Crime Science* 9.1, 2020, pp. 1–15.

[42] I. Kara and M. Aydos. 'The rise of ransomware: Forensic analysis for windows based ransomware attacks'. *Expert Systems with Applications* 190, 2022, p. 116198.

[43] M. Keshavarzi and H. Ghaffary. 'I2CE3: A dedicated and separated attack chain for ransomware offenses as the most infamous cyber extortion'. *Computer Science Review* 36, 2020, p. 100233.

[44] E. Kontopantelis, I. White, M. Sperrin and I. Buchan. 'Outcome-sensitive multiple imputation: a simulation study'. *BMC Medical Research Methodology* 17.1, 2017, pp. 1–13.

[45] M. Korczynski, S. Tajalizadehkhoob, A. Noroozian, M. Wullink, C. Hesselman and M. V. Eeten. 'Reputation metrics design to improve intermediary incentives for security of TLDs'. *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE. 2017, pp. 579–594.

[46] P. Langlois. *2020 Data Breach Investigations Report*. 2020.

[47] P. Leo, Ö. Işik and F. Muhly. 'The Ransomware Dilemma'. *MIT Sloan Management Review*, 2022.

[48] E. R. Leukfeldt and M. Yar. 'Applying routine activity theory to cybercrime: A theoretical and empirical analysis'. *Deviant Behavior* 37.3, 2016, pp. 263–280.

[49] Z. Li and Q. Liao. 'Game theory of data-selling ransomware'. *Journal of Cyber Security and Mobility*, 2021, pp. 65–96.

[50] P. H. Meland, Y. F. F. Bayoumy and G. Sindre. 'The Ransomware-as-a-Service economy within the darknet'. *Computers & Security* 92, 2020, p. 101762.

[51] D. Ndichu. *Kaspersky: over half of ransomware victims paid off attackers in 2020*. Kaspersky, 4 April. Available at: https://gulfbusiness.com/kaspersky-over-half-of-ransomware-victims-paid-off-attackers-in-2020/ [Accessed November 2022]. 2021.

[52] M. Olaimat, M. A. Maarof and B. Al-rimy. 'Ransomware anti-analysis and evasion techniques: A survey and research directions'. *2021 3rd International Cyber Resilience Conference (CRC)*. IEEE. 2021, pp. 1–6.

[53] K. Oosthoek, J. Cable and G. Smaragdakis. 'A Tale of Two Markets: Investigating the Ransomware Payments Economy'. *arXiv preprint arXiv:2205.05028*, 2022.

[54] O. Owolafe and A. Thompson. 'Analysis of Crypto-Ransomware Using Network Traffic'. *Journal of Information Security and Cybercrimes Research* 5.1, 2022, pp. 72–79.

[55] H. Oz, A. Aris, A. Levi and A. S. Uluagac. 'A survey on ransomware: Evolution, taxonomy, and defense solutions'. *ACM Computing Surveys (CSUR)*, 2021.

[56] B. Payne and E. Mienie. 'Multiple-Extortion Ransomware: The Case for Active Cyber Threat Intelligence'. *ECCWS 2021 20th European Conference on Cyber Warfare and Security*. Academic Conferences Inter Ltd, 2021, p. 331.

[57] A. Rege, Z. Obradovic, N. Asadi, B. Singer and N. Masceri. 'A temporal assessment of cyber intrusion chains using multidisciplinary frameworks and methodologies'. *2017 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (Cyber SA)*. IEEE. 2017, pp. 1–7.

[58] C. P. Research. *Behind the curtains of the ransomware economy - the victims and the Cybercriminals*. Retrieved July 12, 2022, from https://research.checkpoint.com/2022/behind-the-curtains-of-the-ransomware-economy-the-victims-and-the-cybercriminals/. 2022.

[59] C. P. Research. *Leaks of Conti Ransomware Group Paint Picture of a Surprisingly Normal Tech Start-Up*. Retrieved August 31, 2022, from https://research.checkpoint.com/2022/leaks-of-conti-ransomware-group-paint-picture-of-a-surprisingly-normal-tech-start-up-sort-of/. 2022.

[60] C. Simoiu, J. Bonneau, C. Gates and S. Goel. '"I was told to buy a software or lose my computer. I ignored it": A study of ransomware'. *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. 2019, pp. 155–174.

[61] G. Stringhini. *Adversarial Behaviours Knowledge Area*. Cyber Security Body of Knowledge, The National Cyber Security Centre. 2019.

[62] B. Urooj, M. A. Shah, C. Maple, M. K. Abbasi and S. Riasat. 'Malware detection: a framework for reverse engineered android applications through machine learning algorithms'. *IEEE Access*, 2022.

[63] H. Väliaho and T. Pekkonen. *A Procedure for Stepwise Regression Analysis*. De Gruyter, 2022.

[64] D. Walsh. 'Victim selection procedures among economic criminals: The rational choice perspective'. *The reasoning criminal*. Routledge, 2017, pp. 39–52.

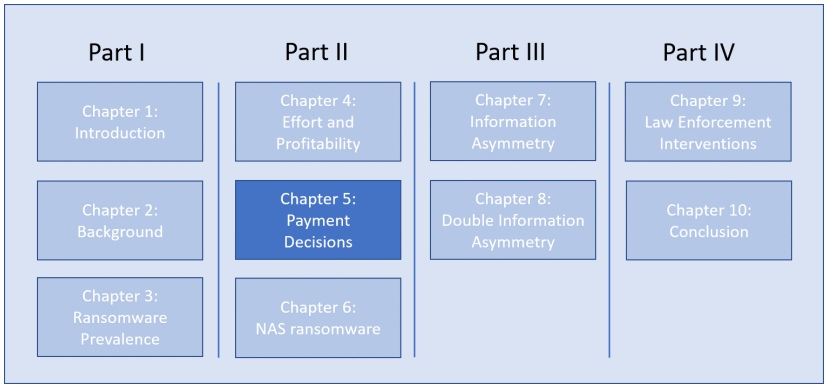
[65] S. V. de Weijer, R. Leukfeldt and W. Bernasco. 'Determinants of reporting cybercrime: A comparison between identity theft, consumer fraud, and hacking'. *European Journal of Criminology* 16.4, 2019, pp. 486–508.

Speaking about myself, it would be high time to stop. There will be enough money for more than one hundred years. But there is never a lot of money – there is always not enough money.

~ Alleged affiliate of ransomware group REvil

Chapter 5

Ransomware Payment Decisions



This study aims to address this existing gap by conducting an empirical investigation that focuses on the ransom paid by victims. Extending on past research, we analyse 382 ransomware attacks reported to the Dutch Police and/or handled by an Incident Response (IR) company. One challenge of modeling ransom payments is the large proportion of victims who did not pay, which leads to zero-inflation. We tackled this problem by employing a hurdle model, which effectively deals with zero-inflation by capturing ransom paid as a two-step decision-making process: first, victims decide whether to comply with the ransom demands, and if they choose to do so, they then need to determine the acceptable ransom amount.

5.1 Introduction

Over the past few years, crypto-ransomware has emerged as a major concern for society, as reflected in Europol's consistent recognition of crypto-ransomware as a top priority [15, 14]. In the United States alone, [3] estimates that 576 organizations fell victim to crypto-ransomware attacks in 2021 with 159.4 billion USD in downtime costs. Crypto-ransomware, ransomware for short, is a type of malware that encrypts files, allows victims to regain access upon paying a ransom to the attackers. The surge in ransomware attacks can be attributed to its profitability [10, 5]. Furthermore, [20] highlights that criminals are incentivized to target victims who highly value their data, leading to increased social welfare costs. To address the surge in ransomware attacks and mitigate the impact, efforts must be made to reduce the profitability of ransomware attacks.

To reduce ransomware profitability, [16] proposes three defensive strategies: lowering the value of ransom payments, increasing the costs of ransomware attacks, and decreasing the willingness of victims to pay. The present study focuses on the profitability of ransomware attacks by focusing on the victims' decision-making to pay a certain ransom amount.

Criminals profit from ransomware attacks primarily because victims choose to pay the ransom [18, 7]. Various sources provide insights into the size of ransomware payments in recent years. An empirical study of ransomware attacks reported to the Dutch Police from 2019 to 2022 indicated an average ransom demand of 720,256 euros, with 21% of victims actually paying the ransom, resulting in financial losses of 433,191 euros [27], as victims have estimated when they reported to the police.

A company tracking ransomware payments on the Bitcoin blockchain, estimated the total ransomware payments to be 765 Million USD in 2020, 766 Million USD in 2021, and 457 Million USD in 2022 [9]. They attribute the drop in ransomware revenue in 2022 not to fewer attacks but to victims' reduced willingness to pay ransomware attackers. Interestingly, more than 50% of the revenue is concentrated among the top 5-7 strains, while the total number of strains is estimated to be around 10,000 [9].

[29] constructed a data set using ransom notes uploaded by victims, which indicated a cumulative ransom payment of 101,297,569 USD between 2017 and 2022. Additionally, [34] monitored 41,424 victims from 2012 to 2021, revealing a combined ransom payment of 176 Million USD.

Combined, these figures provide an approximation of the profitability of ransomware attacks. However, it is crucial to highlight that the social welfare costs are significantly higher, as the ransom paid merely represents a fraction

of the victims' recovery costs [3, 27, 7] and there are non-monetary costs like psychological costs, social costs and impact on customers and service users [30].

A complimentary way for criminals to profit from ransomware attacks emerges through data exfiltration. Present-day ransomware attacks frequently involves stealing data to pressure victims into paying, with threats of public exposure on leak pages [27, 24]. Additionally, criminals may choose to sell the stolen data to rival business competitors or other malicious actors for potential use in subsequent attacks [21, 25].

It is important to understand victims' willingness to pay to decrease ransomware attacks' profitability. A survey by [20] estimated a willingness to pay around 150 British pounds among 149 individuals in the UK. However, this study's limitations are twofold: it focused on individuals, disregarding potential differences with businesses' decision-making, and the survey lacked real-life applicability as it asked participants to speculate on hypothetical scenarios rather than reporting real-life situations. In a follow-up study, [6] found that a proportion of individuals appear to reject paying any ransom.

An alternative approach is provided by [10]. They conducted interviews with 41 ransomware victims from SMEs, large companies, and public organizations in the UK. Among them, 8 victims (20%) opted to pay, mainly to avoid bankruptcy. Conversely, 22 victims (67%) had no intention of paying. [10] proposes a two-step decision-making process: first, determining affordability, and second, evaluating the advantages and disadvantages of paying versus alternative data retrieval methods to minimize disruption and further financial losses.

The primary aim of this study is to apply a quantitative model to capture the decisions made by businesses. Given that many businesses do not pay the ransom we adopt a hurdle model approach to capture factors that influence payment of a ransom and factors that influence the amount of ransom paid. This statistical approach can be seen to capture the two-step decision making process proposed by [10]. We state the main research question as follows: *What factors determine the ransom paid during ransomware attacks?* To answer this question, we focus on three sub-questions:

RQ 1: Which factors determine whether victims will pay or will not pay the ransom?

RQ 2: In case the victims decide to pay, which factors determine the ransom amount victims will pay?

RQ 3: Do different factors influence the ransom payment decision and the amount ransom paid?

We analyse 382 ransomware attacks reported to the Dutch Police between 1 January 2019 and 1 January 2023 and incidents handled by an Incident Response (IR) company between 21 February 2020 and 1 January 2023. To deal with zero-inflation of ransom payments and effectively capture the two-step decision-making process regarding ransom payments of victims we employ a hurdle model. Our key contributions are:

1. Extending on [27], we annotate 525 ransomware attacks reported to the Dutch Police and 116 to an IR company, therefore controlling for possible low willingness to report to the police. For our regression analysis we analyse 382 ransomware attacks.

2. We demonstrate that modeling the ransom paid to criminals could be modeled as a two-step process: whether victims choose to pay and determining the amount of ransom paid. Furthermore, we identify distinct factors influencing the first and second step.

3. More specifically, having insurance results in ransoms that are 2.7 times larger, data exfiltration increases the ransom 4.4 times, and each 1% increase in a victim's yearly revenue causes a 0.12% rise in the ransom paid.

4. We propose a method to construct a demand curve based on empirical data of ransom payments.

The outline of this paper is as follows: In §5.2, we discuss existing literature concerning the profitability of ransomware attacks and state seven hypotheses to answer our research questions. Subsequently, in §5.3, we present our data and the methodology. Afterwards, §5.4 presents the results obtained from our research. To conclude, we discuss our findings and outline future work in §5.5 and §5.6, respectively.

5.2 Related Work and Hypotheses

Ransomware is a financially motivated crime, with cybercriminals seeking to maximize profit by controlling the size of the ransom [20, 10, 18]. Given the relative ease of attacking victims and the low risk of capture, the ransom amount becomes a critical variable they criminals can manipulate. Therefore we make it the focal point of our analysis [20].

The criminals' potential profit heavily relies on the willingness of victims to pay the ransom, influenced by various factors such as the importance of files to the victim, the availability of recent backups, liquid funds, relative trust in the

criminals, and willingness to negotiate with them. From the criminals' perspective, their focus lies on determining the maximum amount each victim is willing to pay for file recovery, also known as the willingness to pay (WTP) [20].

Heterogeneity in businesses maximum willingness to pay a ransom, incentivizes criminals to adopt price discrimination strategies, as cited in previous works [20, 7, 18]. Price discrimination increases profits by encouraging more victims to pay, as the ransom can be lowered for those with a lower WTP while keeping higher prices for others [27, 18].

If criminals do not use price discrimination between victims, which is defined as uniform pricing, criminals impose an identical ransom amount to all victims. Uniform pricing is a characteristic of certain ransomware strains like Deadbolt [26] and old ransomware strains like CryptoLocker [7].

In contrast, second-degree and third-degree price discrimination are classic price discrimination methods. Second-degree price discrimination involves offering victims diverse package options, allowing them to pay solely for the decrypter, preventing data publication, or obtaining a comprehensive security report from the criminal, or any combination of these options [20, 24, 18]. Third-degree price discrimination directly distinguishes different victim types. Criminals using third-degree price discrimination may analyze victims' company details, including yearly revenue from public sources or obtained insurance policy documents during the attack.

Uniform pricing, second-degree and third-degree price discrimination are all pricing methods observed in real-life ransomware attacks and lead to different types of dynamics between criminals and victims [27, 25, 7, 18]. With uniform pricing, the ransom note states the ransom amount and bitcoin address in the ransom note, though this approach is becoming less common [27, 19]. In contrast, second- and third-degree price strategies typically involve negotiation through email or TOR-chat and offering ransoms based on factors like the number of servers encrypted or the services provided by the criminal, such as data decryption, prevention of data publication, or even a security report how the criminal infected the company and which security measures to take [27, 18]. Typically, an adversary initiates the negotiation by specifying an initial ransom, and the victim has the option to counter with a request for a lower price, commonly known as a discount. The negotiation progresses with both parties engaging in reciprocal offers to reach an agreement [18].

Ransomware criminals seem successful in implementing price discrimination strategies. Empirical studies (e.g., [27, 18]) have identified factors influencing the ransom requested and the WTP in ransomware attacks, such as data exfiltration, Ransomware-as-a-Service (RaaS), blackmail, victim's yearly revenue,

sector, and insurance. Notably, an overlap in these factors suggests that criminals may have effectively identified variables for which victims are willing to pay, indicating successful price discrimination strategies.

A significant subset of victims consists of those who refuse to pay the ransom, as emphasized by [10, 6, 5]. For these individuals, the WTP is effectively zero. As a result, it is reasonable to distinguish between those inclined to pay the ransom and those who are not, before modeling the ransom amount they would pay. Considering the decision to pay the ransom as such a two-step procedure can be even more valuable if distinct factors influence the first and second step. For example, having off-line backups might influence the decision to pay, but not the ransom amount if the victim wants to pay. This leads to our first hypothesis:

Hypothesis 1: *The factors influencing the decision to pay are different from the factors influencing the ransom amount paid.*

One of the critical decision that victims face is how to mitigate the ransomware attack, especially since most victims have little experience with ransomware attacks. This lack of expertise creates tension and uncertainty, making it more likely for victims to seek specialized guidance. Especially when the situation is critical and recovery seems difficult. Therefore, many victims might consult an Incident Response (IR) company [38]. Based on this behaviour, we hypothesize that victims who turn to IR companies are more inclined to consider making ransom payments and may also be willing to pay larger ransom amounts.

Hypothesis 2: *Victims' decision to go to the IR company are more inclined to consider payment (H2.1) and pay larger ransom amounts compared to victims who decided not to go to the IR company (H2.2).*

Two strategies employed by companies to reduce the impact of ransomware attacks and decrease the willingness to pay (WTP) are cyber insurance and recoverable offline-backups [16].

Companies benefit from insurance coverage during ransomware incidents in multiple ways [16, 38, 28]. Firstly, insurance providers may have experience in assessing the situation and determining whether the company can recover without paying the ransom. Secondly, if payment is necessary, insurance companies may assist in negotiating and reducing the ransom amount. Thirdly, they might compensate the ransom if paid, and finally, they facilitate the company's recovery process by for example hiring an IR company [28, 38].

Furthermore, there is ongoing debate on whether cyber insurance leads companies to reduce investments in preventive security measures, as insurance coverage may alleviate the financial consequences of an attack, resulting in moral

hazard. For a more elaborate analysis on the relationship between insurance and ransomware we refer to [38, 28, 36].

It is important to note that although cyber insurance may ease the financial burden on victims, it does not address the underlying incentives for ransomware attacks. On the contrary, from the attacker's perspective, cyber insurance might actually encourage more victims to pay the ransom. This could make ransomware attacks more profitable and, unfortunately, more attractive for cybercriminals.

Hypothesis 3: *Victims with cyber insurance are more inclined to consider payment (H3.1) and pay larger ransom amounts compared to victims with no cyber insurance (H3.2).*

Backups represent a valuable strategy for companies to mitigate the impact of ransomware attacks [16, 27]. In the event of file encryption, backups offer a means of restoring data. However, there are three complications. Firstly, attackers actively seek out and delete backups to discourage victims from relying on them and encourage ransom payment. Secondly, difficulties may arise in the recovery process, even if backups remain unaffected by criminals. Research by [37] shows that both cloud-based and colocation backup methods may incur a larger fraction of costs compared to paying the ransom. Additionally, many companies lack awareness of the time required for backup recovery, which might result in considering ransom payment to speed up the process. Thirdly, currently most criminals encrypt and exfiltrate files, threatening publication if no payment is made, imposing costs regardless of backups [24, 21]. Nonetheless, despite these challenges, we expect that having accessible offline backups will likely lead to a reduced number of companies paying the ransom, without affecting the ransom amounts [10].

Hypothesis 4 : *The presence of recoverable backups leads victims to be less inclined to consider payment (H4.1), while not influencing the ransom amount paid, compared to victims lacking recoverable backups (H4.2).*

As mentioned previously, data exfiltration is another incentive for victims to consider paying a ransom and larger ransom amount [24, 21, 27]. Companies want to prevent undesirable outcomes linked to the publication of data and the damage it could cause to their reputation.

Hypothesis 5: *Data exfiltration leads to victims more inclined to consider payment (H5.1) and pay larger ransom amounts compared to victims where no data is exfiltrated (H5.2).*

Table 5.1: Variables used in this chapter and percentage missing values.

Variables	Unit / categories	Missing Values (%)
1a. Ransom requested end negotiations	Euro, Log 10 transformed	228/481 (47%)
1b. Payment	Yes = 1 / No = 0	33/481 (7%)
1c. Ransom paid	Euro, Log 10 transformed	61/481 (13%)
2a. Time negotiating	Hours	70/481 (15%)
2b. Insurance	Yes = 1 / No = 0	50/481 (10%)
2c. Backups	No = 0, Yes + no recovery = 1, Yes + partial recovery = 2, Yes + full recovery = 3	46/481 (10%)
2d. Data exfiltration	Yes = 1 / No = 0	60/481 (13%)
2e. Yearly revenue victim	Euro, Log 10 transformed	11/481 (2%)
2f. Sector victim	Sectors described by Dutch Chamber of Commerce	20/481 (4%)
3a. Data set	IR company = 2, IR company + police = 1 Police = 0	0/481 (0%)
3b. Year encryption	2019, 2020, 2021, 2022	18/481 (4%)

Another relevant factor might be the victim's yearly revenue. The victim's yearly revenue can impact the WTP due to two reasons. Firstly, it influences their ransom payment capacity [10]. Secondly, criminals might use it for second or third-degree price discrimination as described above[27, 18, 7]. Hence, the victim's yearly revenue affects the decision to pay and the ransom amount paid.

Hypothesis 6: *Yearly revenue of the victim influences the decision to pay (H6.1) and the ransom paid (H6.2).*

Finally, some sources describe that the ransom revenues for criminals have increased from 2019 to 2021, but decreased in 2022 [9, 27]. There are claims that insurers and businesses are reacting to the increased attacks in 2020 and 2021 and so we might expect ransoms to be falling [11]. Consequently, the final hypothesis of this study is:

Hypothesis 7: *The frequency of ransom paid is not different over the years (H7.1), but the amount ransom paid is (H7.2).*

5.3 Data and Methodology

A strength of this study is that we use two data sets. The first is an extension of data set and methodology previously used by [27] and consists of 525 ransomware attacks reported to the Dutch Police between 1 January 2019 and 1 January 2023. The second data set are 116 incidents reported by an incident response

Table 5.2: Sector Size in Netherlands According to CBS [8].

Sector Name	Description of Companies in Sector	Dutch Sector Size (%)
Trade	Involves the buying and selling of goods and services.	29.9
Healthcare	Provides medical services, including hospitals and clinics.	29.3
Government	Covers public administration, defense, and social services.	12.1
Education	Includes schools, colleges, and educational services.	11.8
Construction	Concerns the building of infrastructure and buildings.	8.6
Transport	Involves the movement of goods and people.	8.5
Leisure	Includes recreation, entertainment, and tourism.	6.8
ICT	Focuses on Information and Communication Technology services.	5.2
Media	Covers broadcasting, publishing, and other forms of media dissemination.	5.2
Agriculture	Involves farming, forestry, and fishing.	2.1

company (IR company) active in the Netherlands between 20 February 2020 and 1 January 2023. Using another source of data in addition to police reports could help account for situations where people may be less willing to report to the police [33, 35].

To compile the Dutch Police data set, a search was conducted in the police systems employing the keyword *'ransomware'*, which was further analyzed with the authors manually classifying the incidents involving crypto-ransomware attacks. On the other hand, the incidents recorded by the IR company were specifically disclosed to the members of the team for the purpose of our project.

From both data sets we exclude attempted ransomware attacks and attacks reported by individuals, resulting in 418 ransomware attacks in the data set of the Dutch Police and 97 ransomware attacks in the IR company data set. Removing duplicates between data sets, we have a combined data set of 481 unique successful ransomware attacks on companies, see Table 5.1.

The Dutch Police data set is limited to cases within the Netherlands due to jurisdictional limitations. In contrast, the IR company data set comprises cases from various countries where the company was actively involved. Among the 97 attacks in the IR company data set, 42 were recorded outside the Netherlands. Given the geographical proximity of these countries to the Netherlands, it was deemed reasonable to include them in the study, as we anticipate no systematic differences from the other cases in the IR company data set.

Given the presence of both data sets, we have the opportunity to examine the willingness of victims to report ransomware attacks to the police. Within the 97 IR company cases, three distinct categories emerge: 1) 21 cases (22%) did not

report to the police, 2) 34 cases (35%) reported the incident to the police, and 3) 42 cases (43%) occurred in foreign countries, and their reporting status to local authorities remains unknown. Consequently, when focusing solely on cases within the Netherlands, 34 out of 55 victims (62%) reported the ransomware attacks to the Dutch Police. A visual representation of these data sets is presented in Figure 5.1. There seems to be an increase of ransomware attacks in 2020 and 2021 compared to 2019 and 2020. Note that the rise in ransomware attacks in the IR company data set might be due to extra incidents in foreign countries.

Next, we describe the variables coded in this chapter. The **dependent variables** in our study (see Table 5.1) are:

1. Ransom Paid: This section examines variables which help construct the outcome variable *1c. Ransom paid*, which is the main focus of our study.

1a. *Ransom requested end of negotiations:* This variable represents the final offer made by the criminal during negotiations and was measured in euros. It was used to construct *1c. Ransom paid*.

1b. *Payment:* This binary variable indicates whether victims paid the ransom or not. It is categorized as follows: yes = 1, no = 0.

1c. *Ransom paid:* This variable is the primary focus of our study, measured in euros. This variable is calculated by multiplying the payment (*1c.Payment*) with the final ransom (*1b.Ransom requested end of negotiations*). If the ransom requested after the end of negotiations was unknown and there was no payment, then the ransom paid was recorded as 0.

The **independent variables** in this chapter (See Table 5.1) are:

2. Victim Characteristics: This section examines various characteristics of the victim that could serve as significant indicators for the ransom amount paid. These characteristics align closely with our research hypotheses.

2a. *Time Negotiating:* The number of hours devoted to negotiations. If no negotiations occurred, the value for time negotiating was recorded as 0.

2b. *Insurance:* A binary variable indicating whether the victim has insurance coverage that includes ransomware attacks. Categories are defined as follows: yes = 1, no = 0.

2c. *Backups:* This categorical variable represents the presence of backups and their state in the event of a ransomware attack. It is categorized as follows:

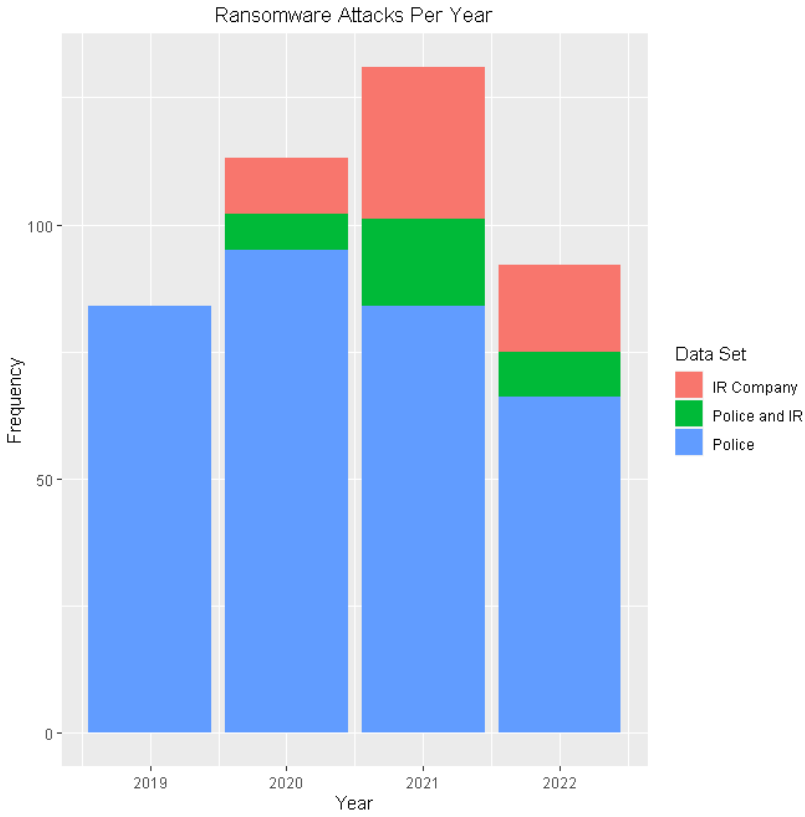


Figure 5.1: Ransomware attacks per year reported to the Police, to the IR company or to both.

no = 0, yes but not possible to recover data = 1, yes but could partially recover data = 2, yes and could fully recover data = 3.

2d. *Data Exfiltration:* This binary variable indicates whether data from the victim was exfiltrated during the ransomware attack. It is categorized as follows: yes = 1, no = 0. Note that although many ransomware groups claim to exfiltrate data as a means of pressuring victims, most groups do not actually carry out this action [24]. Data exfiltration is documented as confirmed if in-depth analysis of network logs reveals significant and abnormal data uploading activity. Additionally, if the victim's data is found to be published on a leak page and verified as belonging to the victim, it is also categorized as data exfiltration.

2e. *Yearly Revenue Victim:* This variable represents the annual revenue of the victim's company, measured in euros and log-transformed, due to very skewed data [27]. The data was obtained from various public sources, including ZoomInfo and DnB [1]. It is worth noting that these sources are also utilized by criminals to access the yearly revenue information of their targets. While there may be inaccuracies in the data retrieved from these sources, its usage by criminals provides a relevant basis for examining potential price discrimination strategies.

2f. *Sector Victim:* This categorical variable identifies the economic sector to which the victim's company belongs, based on the categories employed by the Dutch Chamber of Commerce [8]. See Table 5.2.

3. Contextual Variables: This section highlights the inclusion of metadata which might influence the ransom paid.

3a. *Data set:* A categorical variable indicating the origin of the attack data, categorized as police data set, IR company data set, or both.

3b. *Year encryption:* A categorical variable indicating the year when encryption of victim's files occurred, limited to 2019, 2020, 2021, or 2022.

As ransom paid is the primary focus of our study, we have employed a listwise deletion approach for the regression analysis, removing all cases where the amount of ransom payment was unobserved, resulting in 430 ransomware cases for descriptive analysis. Similarly, applying listwise deletion for the other variables (as depicted in Table 5.1), resulted in 382 observation for the regression analysis. Although using a different sample size for descriptive and regression

analysis might make it harder to compare results, we want our analysis to be as close to the real-life data as possible. Likewise, the listwise deletion approach could introduce potential bias if the ransom payment data is not missing-at-random [31]. However, the method aligns with our research objectives, since we only want to analyse observed ransom payments and the amount of missing observations is relatively low, less than 10%.

Analysis were conducted using Rstudio and R version 4.3.1, with packages *pscl*, *ggplot* and *dplyr*.

We adopt a two-step approach to model the ransom paid in our study, utilizing a hurdle model as proposed by [23]. The hurdle model is suitable for capturing the decision-making process of ransom payment, with the "hurdle" representing the likelihood of a victim paying the ransom, and only after overcoming this hurdle, positive ransom payments are observed. This framework combines two components: the first models the probability of attaining a ransom paid or no payment, while the second part models the ransom amount given that the ransom payment is non-zero. Hurdle models give extra insight by capturing factors influencing zeroes and factors influencing positive amounts [12, 17]. The advantage of using a hurdle model is that it could handle excess zeros efficiently [17]

In our analysis, we employ a hurdle model with a negative binomial distribution. This distribution allows us to model the ransom paid while relaxing the assumption of equal mean and variance as would be the case using a Poisson Distribution. Furthermore, we use a Log Link function to model the logarithm of ransom paid. Log-transforming variables with monetary scales is common in social-empirical studies to transform a non-linear distributed variable to an approximately normal distributed variable [27, 34]. The probability of a victim making no payment can be represented as follows:

$$P(Y_i = 0) = \frac{1}{1 + e^{-\lambda}} \tag{5.1}$$

Where Y_i is the ransom amount Y paid by victim i and the parameter λ is used to predict the count of zero ransom payments. The probability of a non-zero ransom amount, conditional on payment of ransom amount $y > 0$ is:

$$P(Y_i = y) = \frac{\Gamma(y + r_i)}{y! \cdot \Gamma(r_i)} \left(\frac{r_i}{r_i + \mu_i}\right)^{r_i} \left(\frac{\mu_i}{r_i + \mu_i}\right)^{y} \tag{5.2}$$

With r_i is the dispersion parameter for victim i and μ_i is the mean parameter for victim i.

We model the expected ransom amount when a victim pays a ransom as:

$$E(Y_i|Y_i > 0) = e^{\beta_0+\beta_1 x_i} \tag{5.3}$$

With x_i the relevant covariate for victim i, and β_0 and β_1 regression coefficients. Using the probability distributions (5.1) and (5.2) we could extract the total expected ransom amount for both victims who pay and who do not.

$$E(Y_i) = P(Y_i > 0) \times E(Y_i|Y_i > 0) \tag{5.4}$$

Equation (5.4) allows us to construct a regression model that accounts for multiple regressors, enabling us to evaluate their effect on ransom paid. With this regression model, we seek to validate the previously stated hypotheses. We set the significance level to $\alpha = 0.05$, and a p-value below this threshold supports the hypothesis that the variable is significant.

5.4 Results

5.4.1 Descriptive Analysis

Table 5.3: Sum and Average Ransom Paid For Different variables. N=430.

	Categories	# Paid	# Not Paid	Sum Ransom Paid (euro)	Average Ransom Paid (euro)
Year	2019	16 (19%)	68 (81%)	312,053	19,503
	2020	39 (35%)	74 (65%)	27,629,373	708,445
	2021	37 (28%)	94 (72%)	11,848,461	320,229
	2022	25 (27%)	67 (73%)	10,637,366	425,495
Insurance	No	75 (24%)	232 (76%)	9,976,185	133,016
	Yes	33 (44%)	42 (56%)	23,367,453	708,105
Backups	No	28 (27%)	76 (73%)	1,417,017	50,608
	Yes, but not recoverable	45 (58%)	33 (42%)	16,251,606	361,147
	Yes, but partially recoverable	24 (28%)	63 (72%)	11,706,451	487,769
	Yes, and fully recoverable	13 (11%)	109 (89%)	19,671,327	1,513,179
Data exfiltration	No	82 (25%)	250 (75%)	7,331,363	89,407
	Yes	35 (40%)	53 (60%)	43,095,889	1,231,311
Data set	Police	71 (21%)	263 (79%)	21,087,499	301,250
	Police and IR company	18 (52%)	16 (48%)	9,460,831	556,520
	IR company	32 (52%)	30 (48%)	19,878,922	662,631
Total		121 (28%)	309 (72%)	50,427,252	431,002

Table 5.4: Descriptive statistics of victim companies of different sectors. Mean and median revenue are in Million euros, insured, no backup, and paid are percentages. Average ransom paid is in euro and cumulative ransom paid is in Million euros. Bottom row demonstrates unweighted column average. N=430.

	Sector	Number attacks	Number attacks (%)	CBS Sector Size (%)	Mean Revenue (Meuro)	Median Revenue (Meuro)	Insured (%)	No Backup (%)	Paid (%)	Average Ransom Paid (euro)	Cumulative Ransom Paid (Meuro)
1	Trade	140	32.6	29.9	301.91	4.07	19.4	46.8	30.7	112,793	15.79
2	Construction	77	17.9	8.6	382.30	4.47	28.8	48.0	28.6	46,676	3.59
3	ICT	63	14.7	5.2	397.08	3.81	19.7	46.6	28.6	268,039	16.89
4	Healthcare	29	6.7	29.3	37.44	3.61	19.2	37.9	32.1	94,784	2.75
5	Leisure	29	6.7	6.8	7.55	1.24	22.2	59.3	24.1	31,934	0.93
6	Transport	27	6.3	8.5	490.40	5.82	7.7	64.0	33.3	102,690	2.77
7	Media	25	5.8	5.2	424.02	3.64	16.7	47.8	20.0	274,409	6.86
8	Education	14	3.3	11.8	107.40	16.87	0.0	28.6	21.4	22,138	0.31
9	Agriculture	14	3.3	2.1	387.61	0.83	14.3	53.8	15.4	12,389	0.17
10	Government	12	2.8	12.1	58.60	21.27	16.7	41.7	8.3	34,146	0.41
	Average	**43**	-	-	**269.43**	**6.66**	**16.5**	**47.7**	**24.3**	**104,100**	**5.05**

In this subsection we first examine the cumulative, average, and frequency of ransom payments in relation to the various variables outlined in Section 5.3. Subsequently, we will conduct a detailed analysis of the characteristics specific to victim companies across different sectors. For an overview of our results, please refer to Tables 5.3 and 5.4.

Among the 430 victims, 121 victims decided to proceed with ransom payment, approximately 28%. Regarding the total ransom payments made, the combined sum in our data set amounts to 50,427,252 euros. For those who chose to pay, the average ransom amount was 431,002 euros, with a median of 35,000 euros. See Figure 5.2 for the distribution of ransom paid. The distribution seems to be lognormal distributed, but not uniform, which might contradict that criminals in this dataset use uniform pricing.

In 2020 a substantial sum of approximately 28 Million euros ransom was paid, marking an increase compared to the preceding year when the ransom payment amounted to 312,053 euros. This rise in ransom payments can be attributed not only to an increase in the number of attacks (84 in 2019 and 113 in 2020) but also to a higher average ransom paid per attack. However, in subsequent years, namely 2021 and 2022, the ransom payments decreased to around 12 Million and 11 Million euros, respectively. A Kruskal-Wallis test, which is a non-parametric test with the null hypothesis that in all years the ransom paid is the same, results in KW=8.825, df=3, p-value=0.03. This implies that at least in one year the ransom paid is different compared to other years.

Regarding insurance, it is observed that having insurance coverage correlates with a higher likelihood of payment, with 44% of victims opting to pay when insured, as opposed to 24% when uninsured. Additionally, the average amount paid is also greater when the victim has insurance, 708,105 euros, compared to 133,016 euros for those without insurance. Consequently, the total amount of ransom paid is significantly higher for insured victims, reaching approximately 23 Million euros, in contrast to around 10 Million euros for uninsured victims. The Kruskal-Wallis test with null hypothesis that ransom paid with and without

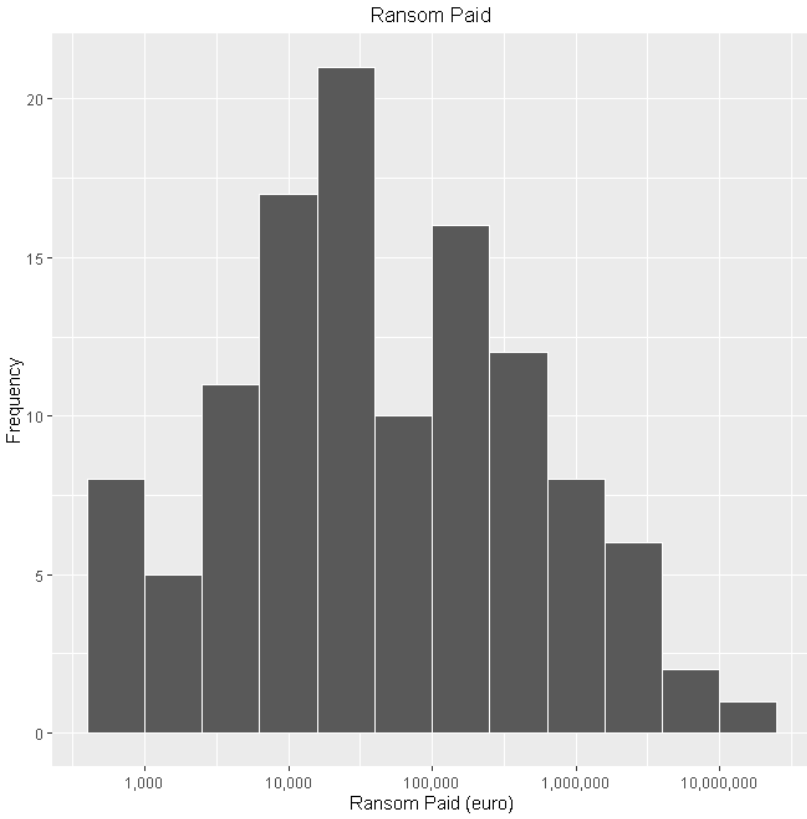


Figure 5.2: Distribution of ransom paid.

insurance is equal, results in KW=20.12, df=1, p-value<0.001. This indicates that having insurance leads to more ransom paid.

Regarding backups, it seems that having recoverable backups leads to a lower probability of payment, observed in only 11% of cases. However, the average ransom paid per attack and the total ransom paid are higher compared to scenarios with other backup conditions. It is noteworthy that victims who lack backups generally pay lower ransoms than those who have backups that cannot be restored, with both the average ransom per attack and the cumulative amounts being lower. One plausible explanation could be that businesses holding data considered valuable enough for ransom payments are generally more likely to employ backup systems, compared to those with less valuable data. The Kruskal-Wallis test with null hypothesis that all backups measures lead to same ransom paid, results in KW=49.65, df=3, p-value<0.001. This indicates that having backups leads to more ransom paid.

In relation to data exfiltration, cases involving exfiltration of data result in a higher probability of payment, as observed in 40% of such incidents, compared to 25% when no data exfiltration occurs. Additionally, the average amount paid is substantially larger, approximately 1.2 Million euros when data is exfiltrated, as opposed to 89,407 euros when no data exfiltration is confirmed. Consequently, the total ransom paid is considerably higher in cases where data exfiltration takes place, reaching approximately 43 Million euros, in contrast to approximately 7 Million euros in attacks without data exfiltration. The Kruskal-Wallis test with null hypothesis that ransom paid with and without data exfiltration is equal, results in KW=15.38, df=1, p-value<0.001. This indicates that data exfiltration leads to more ransom paid.

In terms of the data set used, a higher proportion of ransom payments occur in the data set of the incident response (IR) company, accounting for 52% of cases. This percentage aligns with the combined number of payments made to both the IR company and the police. In comparison, the data set from the police shows a lower ransom payment rate of 21%.

The average annual revenue of victim companies was 269.43 Million euros (sd = 1,802 Million euros). The median revenue was 6.66 Million euros, and the geometric mean was 3.03 Million euros.

The average time spent negotiating was approximately 37 hours (sd = 92 hours). However, when considering only cases where the victim engaged in negotiations, the average negotiating time was 111 hours (sd = 131 hours). Notably, the average negotiating time was lower when a ransom was paid, approximately 25 hours (sd = 78 hours), compared to cases where no payment was made, 72 hours (sd = 118 hours).

In terms of the number of attacks across sectors, the Trade sector stands out with 140 attacks (32.56%), which seems proportionate to its CBS sector size of 29.9%, as shown in Table 5.4. Construction and ICT sectors follow with 77 and 63 attacks, respectively, which is particularly significant given their smaller sector sizes of 8.6% and 5.2% according to CBS data [8].

When examining the characteristics of victim companies in different sectors in Table 5.4, it is noteworthy that healthcare and leisure sectors have higher-than-average percentage of insured companies (22.2% and 19.2%) compared to an average of 16.5%. Victim companies in the Leisure sector have more non-recoverable backups in place than average with 59.3%. Healthcare has higher proportion of companies with recoverable backup measures in place 37.9%. Despite their good backup practices, companies in the healthcare sector pay more than average the ransom with 32.1%, Leisure is close with the average with 24.1%. However, their average and cumulative ransom payments are lower compared to other sectors. These numbers illustrate that different sectors have their own unique challenges when it comes to dealing with ransomware attacks.

On the higher end of the revenue spectrum, the Transport, Media, and ICT sectors have the largest average revenues of 490.40, 424.02, and 397.08 million euros, respectively. Transport companies are less frequently insured (7.7%) and have fewer backup systems (64%), yet pay ransoms at a considerably higher rate of 33.3% compared to the average of 24.3%. The ICT sector, despite an average rate of backup implementation (46.6%), have the highest average ransom payments of 268,039 euros on average, contributing to the largest cumulative ransom of 16.89 million euros across all sectors.

In conclusion, the ICT sector seems the most lucrative target for ransomware groups, since they pay the largest ransom per attack on average. One explanation is that ICT companies often provide critical infrastructure or services to numerous clients. Consequently, if such companies experience downtime due to a ransomware attack, it can have a cascading impact on a large number of clients, thus providing ransomware groups with greater leverage to demand larger ransoms, which aligns with the trends observed in our data set.

5.4.2 Demand Curve of Ransomware

A useful concept for understanding ransom payments is a demand curve [7]. A demand curve, for any ransom amount, depicts the proportion of victims willing to pay that amount. Constructing a demand curve based on empirical observations is challenging [22], due to endogeneity: ransomware criminals might adjust the ransom amount based on the victims' response to negotiation. Consequently, the observed data represents a mixture of both the victims willingness

to pay (demand side) and the criminals willingness to negotiate up or down the ransom amount, and make a deal (supply side).

To explore the willingness of victims to pay the ransom we, therefore, need to make assumptions. To motivate our assumptions consider the following stylized thought experiment: (i) During negotiations with a victim, the criminals incrementally change the ransom request until they discern maximum willingness to pay. (ii) The criminal then decides whether to accept the highest amount the victim will pay, or walk away because the ransom amount is too low (and may harm reputation in future negotiations). This means that if a victim paid the ransom then the criminals were able to fully exploit the victims willingness to pay. In reality, this will under-estimate willingness to pay because victims may have been willing to pay a higher ransom than the criminals requested. It also means that if a victim did not pay then we can assume they would have paid an amount just below the last amount requested. In reality, this will over-estimate willingness to pay.

Assumption 1: (a) Victims who paid a ransom of x euros would have paid any ransom less than x but not a ransom above x. (b) Victims who did not pay a ransom request of x euros would not pay any ransom larger than x but would pay a ransom below x.

Applying Assumption 1(a) we can derive an estimated demand curve for those companies that paid a ransom. For instance, considering the lowest amount paid, which is approximately 500 euros, we infer that 100% of the victims who paid were willing to pay this price. Similarly, observing that around 50% of victims paid more than 35,000 euros, we deduce that 50% of them were willing to pay this amount. Combining these results, we obtain the blue line in Figure 5.3. The blue line or *yes curve* is an estimate of the demand curve based solely on data derived from companies that paid the ransom. This curve is potentially biased by only including companies who paid and by potentially under-estimating the willingness to pay of companies that paid.

Applying Assumption 1(b) we can derive an estimated demand curve for those companies that did no pay a ransom. Now, the argument proceeds in the opposite direction. Here, if we observe, say, 65% of companies refusing to pay a ransom request of $100,000$ euros or below, then we infer that 35% of companies would have be willing to pay a ransom of $100,000$. This aggregation yields the red line in Figure 5.3. This estimated demand curve is potentially biased by only including those who did not pay and by potentially over-estimating the willingness to pay of those who did not pay.

While the two estimated demand curves in Figure 5.3 are derived using stylized assumptions they both give a similar picture of the underlying demand

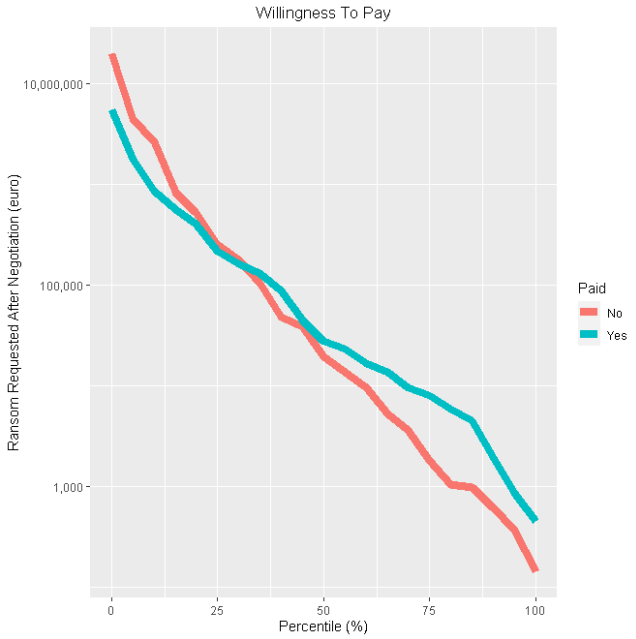


Figure 5.3: Ransom amount in euros paid by victims (blue line, *yes curve*) and those who did not pay (red line, *no curve*). Ordering the data according to ransom amounts under the assumption that victims who paid would pay less and those who did not pay would not pay if the ransom is larger, results in a demand-like curve.

curve. The *yes curve* gives a higher estimate of demand at low ransoms because it is based on those who paid. The *no curve* gives a higher estimate of demand at very high ransoms because it includes observations with very high ransoms that were not paid. In both cases, though, we see an approximate log-linear relationship between willingness to pay and demand with around 35% willing to pay a ransom of $100,000$ euros.

5.4.3 Hurdle Model of Ransom Paid

The results of the hurdle model are described in Table 5.5. The dispersion parameter is significant, which implies the negative binomial distribution is the appropriate fit.

Table 5.5: Hurdle model. The Zero Hurdle Model at the bottom models the first step whether victims decide to pay or not. The Count Model models the second step how much ransom a victim pays if the victim decides to pay in the first step. Estimate, std. error and z-value are rounded to two decimals, p-value to three decimals. N = 382.

Second Step: Count Model	**Estimate**	**Std. Error**	**z-value**	**p-value**	**Sign.**
Intercept	6.88	0.84	8.17	0.000	***
Year = 2020	1.02	0.43	2.37	0.018	*
Year = 2021	1.21	0.53	2.28	0.023	*
Year = 2022	-0.05	0.59	-0.09	0.931	
Insurance = Yes	1.03	0.29	3.60	0.000	***
Log Yearly Revenue Victim	0.39	0.09	4.49	0.000	***
Backups = Yes, not recoverable	0.12	0.36	0.33	0.741	
Backups = Yes, partially recoverable	0.32	0.46	0.70	0.485	
Backups = Yes, fully recoverable	0.51	0.52	1.00	0.320	
Data exfiltration = Yes	1.49	0.37	4.02	0.000	***
Data set = IR company + Police	0.71	0.54	1.30	0.193	
Data set = IR company	0.36	0.52	0.69	0.491	
Log(r_i)	-0.28	0.12	-2.31	0.021	*
First Step: Zero Hurdle Model	**Estimate**	**Std. Error**	**z-value**	**p-value**	**Sign.**
Intercept	-1.99	0.74	-2.67	0.008	**
Year = 2020	0.82	0.40	2.05	0.040	*
Year = 2021	-0.10	0.44	-0.23	0.819	
Year = 2022	0.09	0.52	0.18	0.859	
Insurance = Yes	-0.23	0.45	-0.52	0.606	
Log Yearly Revenue Victim	0.12	0.11	1.06	0.290	
Backups = Yes, not recoverable	0.62	0.36	1.70	0.090	
Backups = Yes, partially recoverable	-0.87	0.42	-2.06	0.039	*
Backups = Yes, fully recoverable	-3.31	0.61	-5.44	0.000	***
Data exfiltration = Yes	0.26	0.42	0.61	0.544	
Data set = IR company + Police	2.32	0.62	3.72	0.000	***
Data set = IR company	3.19	0.67	4.75	0.000	***

[1] Where Sign. is $^{*}p \leq .05$, $^{**}p \leq .01$, $^{***}p \leq .001$.

The findings of this study provide empirical support for the confirmation of H1, indicating that distinct factors influence the decision-making process con-

Table 5.6: Summary of Results for Different Hypotheses. The sign denotes the type of relationship, positive +, neutral = and negative -. Confirmed hypothesis have a X, whereas rejected hypotheses have a -.

Variable	Sign	Hypothesis	Confirmed
H1: Two-step approach	=	Different factors influence ransom paid and payment decision	X
H2: IR company and police	+	H2.1: Pay or not	X
	+	H2.2: Ransom amount	-
H3: Insurance	+	H3.1: Pay or not	-
	+	H3.2: Ransom amount	X
H4: Recoverable Backups	-	H4.1: Pay or not	X
	=	H4.2: Ransom amount	-
H5: Data Exfiltration	+	H5.1: Pay or not	-
	+	H5.2: Ransom amount	X
H6: Yearly Revenue	+	H6.1: Pay or not	-
	+	H6.2: Ransom amount	X
H7: Year encryption	=	H7.1: Pay or not	-
	+	H7.2: Ransom amount	X

cerning ransom payments compared to the actual ransom amount paid. Specifically, instances involving victims who hire an Incident Response (IR) company (β=3.19, p<0.001) or both the IR company and the police (β=2.32, p<0.001) demonstrate higher payment rates than those solely reporting to the police, which validates hypothesis H2.1. However, contrary to hypothesis H2.2, victims with assistance from IR companies (β=0.36, p=0.49) and the data set involving both the police and IR company (β=0.71, p=0.19) did not pay larger ransoms.

Regarding insurance coverage, victims with insurance do not appear to be more inclined to pay the ransom (H3.1) (β=-0.23, p=0.61). Nonetheless, they do pay larger ransom amounts, thus confirming hypothesis H3.2 (insurance coverage β=1.03, p<0.001). Taking the exponential of β leads to 2.7, which indicates that insurance leads to 2.7 times larger ransom paid.

The presence of recoverable backups significantly diminishes the likelihood of payment (H4.1), as follows from reduced probability of ransom payments when victims have partially (β=-0.87, p=0.04) or fully recoverable backups (β=-3.31, p<0.001). However, recoverable backups do not appear to influence the ransom amount paid (H4.2).

Data exfiltration does not lead to more frequent ransom payments (H5.1) (β=0.26, p=0.54); nonetheless, it does result in a larger ransom amount paid (H5.2), as supported by the positive relationship between data exfiltration and the ransom amount paid (β=1.49, p<0.001). Taking the exponential of β leads to 4.4, which indicates that data exfiltration leads to 4.4 times larger ransom paid.

The log yearly revenue of the victim does not appear to impact the decision to pay the ransom (H6.1) (β=0.12, p=0.29). However, it does lead to larger ransom payments, which confirms hypothesis H6.2 (log yearly revenue of the victim β=0.39, p<0.001). Since both the dependent and independent variable are logarithms, we could interpret the β as the elasticity: an 1% increase in a victim's yearly revenue causes a 0.12% rise in the ransom paid.

Lastly, the frequency of ransom payments did not change over the four years (H7.1). However, the ransom paid during 2020 and 2021 exceeded that of 2019 and 2022 (H7.2). Notably, ransom payments were higher in 2020 (β=1.02, p=0.02) and 2021 (β=1.21, p=0.02) compared to 2019. Table 5.6 provides an overview which hypotheses are confirmed or rejected.

5.5 Discussion and Conclusion

The present study set out to examine the ransom paid during ransomware attacks, analyzing 382 ransomware incidents reported to the Dutch Police and an IR company in the Netherlands. Drawing on economic literature, we proposed a two-step process that determines the ransom amount paid. Initially, victims decide whether to pay, followed by the decision of how much to pay, which we modeled using a hurdle model. Our model focused on personalized ransom pricing by ransomware criminals, since is most common for businesses and organizations, compared to individuals who more often encounter uniform pricing [26]. Our estimated hurdle model revealed distinct factors influencing each decision.

Our first research question focused on the first step in the two-step process: factors determining whether victims will pay the ransom. Our findings suggest that the decision to pay depends on backups measures and companies who hire an IR company. Furthermore, there was a difference in frequency of paying the ransom in 2020 compared to the other years examined in this chapter. Insurance,

yearly revenue and data exfiltration do not seem to influence the victims' decision to pay the ransom.

Our second research question focused on the factors determining how much ransom will pay, the second step of our two-step process. Our findings suggest that data exfiltration, insurance coverage and yearly revenue of the victim are important factors for determining how much ransom a victim will pay if they decide to pay. Furthermore, in 2020 and 2021 more ransom was paid than in 2019 and 2022. We did not find differences in ransom paid between victims with different backups measures and companies in the IR company data set.

Our third research question focused on whether the decision to pay and ransom amount paid depend on distinct factors. Based on the findings from the previous two research questions, we can conclude different factors influence the two steps in our model. Furthermore, the hurdle model supports the notion of a two-step choice process proposed by [10]. Companies first assess affordability and then evaluate the advantages and disadvantages of paying versus pursuing alternative data retrieval methods to minimize disruption and financial losses.

The significance of backups for the decision to pay ransom aligns with the rationale that having an alternative recovery procedure is crucial in avoiding costly downtime, in line with [16]. Additionally, our analysis showed that companies consulting the IR company were more willing to pay, as they sought guidance expert assistance in recovering from the ransomware attack. In case the victim considered payment, the IR company helps navigating the payment process, understanding associated risks, and potentially negotiating a discount on the ransom, as outlined by [38].

Previous research already addressed that insurance does not necessarily increase the probability of ransom payments [5] as was confirmed by our results. Nevertheless, having insurance does lead to larger ransom paid. Perhaps this is due to exposed moral hazard: since someone else is paying for the victim, the victim is willing to pay a larger amount. However, exposed moral hazard would also imply a larger proportion of victims be willing to pay the ransom. Perhaps ethical considerations or partial coverage by insurance might explain this difference in our results.

Likewise, the yearly revenue of a company did not influence the payment decision, but did influence the ransom paid. This result is in line with [27, 18]. This might due to victims being more financially capable to pay larger ransom and price discrimination strategies from the criminals [7, 18].

Contrary to prior claims [24, 21], data exfiltration did not directly lead to increased probability of ransom payments. However, our study found that victims tend to pay more when data exfiltration occurred, potentially to avoid reputation

costs linked to data publication. The difference in findings may arise from using a hurdle model: although the payment rates are significantly larger with data exfiltration than without, controlling for all other variables this difference seems to be insignificant. This finding illustrates the power a hurdle model: simultaneously estimating proportion paying and ransom amount paid.

Our results show that ransom payments in 2020 and 2021 are different from other years, which is congruent with previous findings [11, 9]. Perhaps, ransom payments in 2021 are influenced by major global events. Economically, the year was marked by the COVID-19 pandemic and the initial stages of the Ukraine conflict [32]. These events, coupled with evolving cyber insurance policies, such as Lloyd's exclusion clauses, may have impacted ransomware payment strategies. While our analysis suggests these factors as possible influences, it is important to note that it is impossible to be certain, given the complex interaction between global economic, political dynamics, and cybercrime."

It is often assumed that the willingness to report ransomware attacks to the police is typically low [27]. However, our investigation, which involved comparing data from the police and an IR company, revealed a notably high reporting rate of 62% among Dutch companies. This proportion exceeds the rates of 8-10% reported in studies focusing on the willingness to report online fraud cases to the Dutch Police [35]. This difference in reporting behavior could be attributed to victims being more inclined to report severe crimes to the police [33]. Even though our data set is limited to one IR company, the high reporting rate among Dutch companies is unlikely to be affected by lower willingness to report from victims managed by other IR companies, considering that the IR company featured in the present study accounted for half of the victims managed by any IR company in the Netherlands [24].

Nevertheless, it is good to mention that differences between the police and IR company data sets might be the result of internal processing of information and data. Typically, data from the police was unstructured and incomplete, whereas data from the IR company was typically more structured and complete. The differences found between the two data sets might be the result of this difference in data collection, processing and storage.

In conclusion, this study provides valuable insights into victim's decision-making process of paying the ransom during ransomware attacks. We analyzed 382 ransomware attacks reported to the Dutch Police and to an IR company, controlling for reporting bias. Our findings reveal a two-step process in ransom payments, with distinct factors influencing the decision to pay and the amount paid. These contributions aid in developing effective strategies to combat and mitigate the impact of ransomware attacks.

5.6 Limitations and Further work

Limitations of this study include:

1. Our study focused mostly on companies in the Netherlands. It might be difficult to generalize our results to other countries. In other countries ethical considerations of paying the ransom might be different than the Netherlands, possibly changing the significance and/or effect size of different factors influencing the two different steps in our study [13]. However, due to the sensitivity of the data, it might be hard to get data from other countries.

2. In our models we did not account for the perceived reputation of the attacker, which could significantly impact victim decisions on payment and ransom amount. Here reputation is the perceived probability of getting a key to regain access to files after payment [4].

3. Due to the sensitive nature of the data, only one person could code the data, see also [27]. This may introduce different types of biases, despite efforts to mitigate these biases through anonymous group discussions.

4. The potential endogeneity of ransom price on the decision to pay was outside the scope of the present paper. As the ransom requested may influence victim willingness to pay, criminals could adjust the ransom amount to maximize their profits. The potential supply-side of the demand curve could be modeled with endogeneity models [22].

Future research can explore several interesting avenues. Firstly, a focus on studying the endogeneity of price and willingness to pay could enhance our understanding of the dynamic between ransom requested and ransom paid. Secondly, accounting for the perceived trustworthiness of the attacker may influence the decision-making process, probably affecting both the likelihood of payment and the ransom amount. Lastly, generalizing the study's results to more countries could offer insights into potential variations in factors influencing the two steps, leading to more effective policy-making and a broader understanding of ransomware attack profitability.

This study provides valuable insights for policy makers and law enforcement in devising interventions to combat ransomware profitability. Two approaches can be considered:

1. Focus on the first step of the hurdle model: Encourage fewer victims to pay by emphasizing the importance of having recoverable backups. Promoting offline backups and conducting ransomware attack simulations can help prevent hasty decisions to pay.

2. Address the second step of the hurdle model: If victims decide to pay, they should pay less. Measures could include encouraging companies to take preventive measures against data exfiltration and engaging with cyber insurance companies to strategize on handling ransomware payments. Targeting large companies first in awareness campaigns may prove effective, as their refusal to pay can undermine ransomware profitability compared to smaller businesses.

In assessing the role of insurance providers in the ransomware economy, it's crucial to recognize that the current financial incentives might not encourage these companies to minimize ransom payments. In many instances, paying the ransom is the least costly option from a short-term perspective. This raises important questions about the need for regulatory intervention to correct what could be considered a market failure.

From a policy standpoint, several options are possible. One could consider a ban on insurance companies covering ransom payments. However, this may have no effect if an insurance payout still gives the company sufficient financial leverage to cover to the ransom themselves. Another possibility would be to restrict insurance payouts if a victim makes a ransom payment. However, this might lead to more companies going bankrupt, since they might not afford the ransom requested and also could not recover without the decryption key. Therefore it is important to consider the social welfare consequences of such a policy intervention.

Moreover, insurance companies could contribute to the fight against ransomware by increasing transparency and sharing valuable data with law enforcement. By doing so, we can collectively develop a richer understanding of the ransomware ecosystem, leading to better-informed strategies for combating these threats.

In conclusion, our recommendation is to consider more nuanced changes to insurance policies. These could offer a more effective approach for reducing the societal cost of ransomware attacks than more heavy-handed interventions like outright bans or additional taxes on ransom payments.

5.7 Ethics

We follow the principles from Menlo Report [2] to justify the ethical considerations made in this chapter:

Respect for Persons: Privacy and confidentiality of participants were prioritized. Individual cases were not considered, and data were aggregated at the sector levels to ensure the privacy of victims.

Beneficence: To maximize benefits and minimize harm, access to police investigation information was restricted to one member of the project with security clearance, while other team members received aggregated results. This approach, despite challenges to transparency, was deemed necessary for the large-scale empirical ransomware study. Additionally, understanding victims' decision-making about ransom payments may inform future criminals. Our research adheres to the principle of full-disclosure. Considering the entire study, we estimate that our model better informs victims and policy makers how to take preventive measures to prevent further harm than it educates criminals.

Justice: Equal opportunity was ensured for all ransomware attacks in the study, as selection was based solely on the keyword "ransomware" in police systems and attacks disclosed by the IR company. No additional emphasis was given to attacks with media attention or those involving the IR company.

Respect for Law and Public Interest: Specific information about certain groups, strains, or Dutch Police operations was excluded from the paper. Additionally, the IR company was involved in reviewing the paper to exclude potentially malicious information. The goal of the study is to inform potential victims and policy makers to take effective preventive measures.

This page is intentionally left blank.

Bibliography

[1] I. 471. *Conti leaks: Cybercrime fire team*. Intel 471. Retrieved July 24, 2023, from https://intel471.com/blog/conti-leaks-cybercrime-fire-team. 2022.

[2] M. Bailey, D. Dittrich, E. Kenneally and D. Maughan. 'The Menlo Report'. *IEEE Security & Privacy* 10.2, 2012, pp. 71–75.

[3] P. Bischoff. *Ransomware attacks cost the US $159.4bn in downtime alone in 2021*. Comparitech. https://www.comparitech.com/blog/information-security/us-ransomware-attacks-cost/. 2022.

[4] A. Cartwright and E. Cartwright. 'Ransomware and reputation'. *Games* 10.2, 2019, p. 26.

[5] A. Cartwright, E. Cartwright, J. MacColl, G. Mott, S. Turner, J. Sullivan and J. R. Nurse. 'How cyber insurance influences the ransomware payment decision: theory and evidence'. *The Geneva Papers on Risk and Insurance-Issues and Practice* 48.2, 2023, pp. 300–331.

[6] A. Cartwright, E. Cartwright, L. Xue and J. Hernandez-Castro. 'An investigation of individual willingness to pay ransomware'. *Journal of Financial Crime* 30.3, 2023, pp. 728–741.

[7] E. Cartwright, J. H. Castro and A. Cartwright. 'To pay or not: game theoretic models of ransomware'. *Journal of Cybersecurity* 5.1, 2019, tyz009.

[8] Centraal Bureau voor de Statistiek. *Online veiligheid en criminaliteit 2022*. `https://www.cbs.nl/nl-nl/publicatie/2023/19/online-veiligheid-en-criminaliteit-2022`. Accessed: 2024-06-19. 2023.

[9] Chainalysis. *Crypto Ransomware Revenue Down as Victims Refuse to Pay*. Retrieved July 23, 2023, from https://blog.chainalysis.com/reports/crypto-ransomware-revenue-down-as-victims-refuse-to-pay/. 2023.

[10] A. Y. Connolly and H. Borrion. 'Reducing ransomware crime: analysis of victims' payment decisions'. *Computers & Security* 119, 2022, p. 102760.

[11] Coveware. *Ransom Monetization Rates Fall to Record Low Despite Jump in Average Ransom Payments*. Coveware Blog. Retrieved July 25, 2023, from https://www.coveware.com/blog/2023/7/21/ransom-monetization-rates-fall-to-record-low-despite-jump-in-average-ransom-payments. 2023.

[12] J. G. Cragg. 'Some statistical models for limited dependent variables with application to the demand for durable goods'. *Econometrica: Journal of the Econometric Society*, 1971, pp. 829–844.

[13] A. Culafi. *Coveware: Rate of victims paying ransom continues to plummet*. TechTarget. Retrieved August 22, 2023, from https://www.techtarget.com/searchsecurity/news/366545539/Coveware-Rate-of-victims-paying-ransom-continues-to-plummet. 2023.

[14] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2021*. Tech. rep. Retrieved August 31, 2022. Luxembourg, 2021. URL: https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021.

[15] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2023*. Tech. rep. Retrieved August 31, 2023. Luxembourg, 2023. URL: https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-assessment-iocta-2023.

[16] E. Galinkin. 'Winning the Ransomware Lottery: A Game-Theoretic Approach to Preventing Ransomware Attacks'. *International Conference on Decision and Game Theory for Security*. Cham: Springer International Publishing, 2021, pp. 195–207.

[17] W. H. Greene. *Econometric Analysis*. Pearson Education India, 2003.

[18] P. Hack and Z. Y. Wu. *We wait, because we know you. Inside the ransomware negotiation economics*. 2021.

[19] N. Hassan. *Ransomware revealed*. Berkeley: Apress, 2019.

[20] J. Hernandez-Castro, A. Cartwright and E. Cartwright. 'An economic analysis of ransomware and its welfare consequences'. *Royal Society Open Science* 7.3, 2020, p. 190023.

[21] Z. Li and Q. Liao. 'Ransomware 2.0: to sell, or not to sell a game-theoretical model of data-selling ransomware'. *Proceedings of the 15th International Conference on Availability, Reliability and Security*. 2020, pp. 1–9.

[22] A. MacKay and N. Miller. *Estimating models of supply and demand: Instruments and covariance restrictions*. Working Paper 19-051. Harvard Business School Strategy Unit, 2023.

[23] A. McDowell. 'From the help desk: hurdle models'. *The Stata Journal* 3.2, 2003, pp. 178–184.

[24] T. Meurs and L. Holterman. *Whitepaper data-exfiltratie bij een ransomware-aanval*. Retrieved from https://executivefinance.nl/wp-content/uploads/2023/01/VCNL-Whitepaper-Exfiltratie.pdf. 2022.

[25] T. Meurs, M. Junger, A. Abhishta, E. Tews and E. Ratia. 'COORDINATE: A model to analyse the benefits and costs of coordinating cybercrime'. *JISIS* 12.4, 2022.

[26] T. Meurs, M. Junger, E. Tews and A. Abhishta. 'NAS-ransomware: hoe ransomware-aanvallen tegen NAS-apparaten verschillen van reguliere ransomware-aanvallen'. *Tijdschrift voor Veiligheid* 21.3-4, 2022, pp. 69–88.

[27] T. Meurs, M. Junger, E. Tews and A. Abhishta. 'Ransomware: How Attacker's Effort, Victim Characteristics and Context Influence Ransom Requested, Payment and Financial Loss'. *2022 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2022, pp. 1–13.

[28] G. Mott, S. Turner, J. R. Nurse, J. MacColl, J. Sullivan, A. Cartwright and E. Cartwright. 'Between a rock and a hard (ening) place: Cyber insurance in the ransomware era'. *Computers & Security* 128, 2023, p. 103162.

[29] K. Oosthoek, J. Cable and G. Smaragdakis. 'A Tale of Two Markets: Investigating the Ransomware Payments Economy'. *arXiv preprint* arXiv:2205.05028, 2022.

[30] N. Pattnaik, J. R. Nurse, S. Turner, G. Mott, J. MacColl, P. Huesch and J. Sullivan. 'It's more than just money: The real-world harms from ransomware attacks'. *International Symposium on Human Aspects of Information Security and Assurance*. Cham: Springer Nature Switzerland, 2023, pp. 261–274.

[31] T. B. Pepinsky. 'A note on listwise deletion versus multiple imputation'. *Political Analysis* 26.4, 2018, pp. 480–488.

[32] F. Teichmann, S. R. Boticiu and B. S. Sergi. 'The evolution of ransomware attacks in light of recent cyber threats. How can geopolitical conflicts influence the cyber climate?' *International Cybersecurity Law Review*, 2023, pp. 1–22.

[33] J. Tolsma, J. Blaauw and M. T. Grotenhuis. 'When do people report crime to the police? Results from a factorial survey design in the Netherlands, 2010'. *Journal of Experimental Criminology* 8, 2012, pp. 117–134.

[34] K. Wang, J. Pang, D. Chen, Y. Zhao, D. Huang, C. Chen and W. Han. 'A large-scale empirical analysis of ransomware activities in bitcoin'. *ACM Transactions on the Web (TWEB)* 16.2, 2021, pp. 1–29.

[35] S. G. V. de Weijer, R. Leukfeldt and S. van der Zee. 'Cybercrime reporting behaviors among small-and medium-sized enterprises in the Netherlands'. *Cybercrime in Context: The human factor in victimization, offending, and policing*. Cham: Springer International Publishing, 2021, pp. 303–325.

[36] J. Wolff. *Cyberinsurance Policy: Rethinking Risk in an Age of Ransomware, Computer Fraud, Data Breaches, and Cyberattacks*. MIT Press, 2022.

[37] T. Wood, E. Cecchet, K. K. Ramakrishnan, P. Shenoy, J. V. der Merwe and A. Venkataramani. 'Disaster recovery as a cloud service: Economic benefits & deployment challenges'. *2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 10)*. 2010.

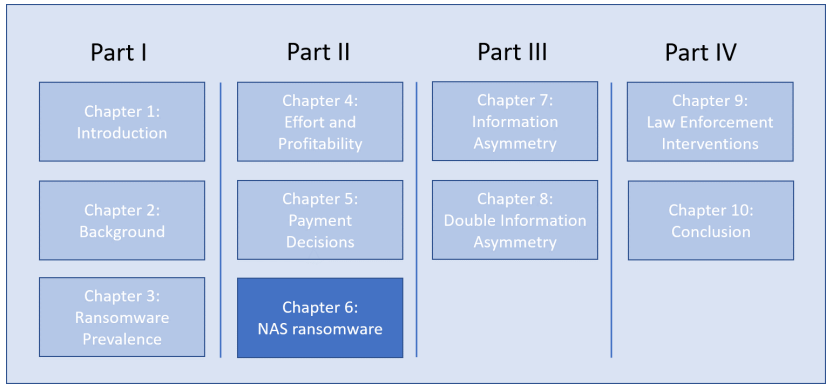
[38] D. W. Woods, R. Böhme, J. Wolff and D. Schwarcz. 'Lessons Lost: Incident Response in the Age of Cyber Insurance and Breach Attorneys'. *32nd USENIX Security Symposium (USENIX Security 23)*. 2023, pp. 2259–2273.

Pound wise, penny foolish

~ English proverb

Chapter 6

Ransomware on NAS Devices



The present study examines the impact of ransomware against Network Attached Storage (NAS) devices. NAS devices are external hard drives which are usually easily accessible through the internet. We analyse 502 ransomware attacks reported to the Dutch Police between 2019 and 2022, of which 104 (20.7%) targeted NAS devices. These attacks targeted both companies as individuals. Furthermore, we examine the police intervention against a NAS ransomware strain in October 2022. One limitation of this sample is possible low willingness to report to the Police. The aim of the present study is to compare ransomware attacks targeting NAS devices versus ransomware targeting other type of IT infrastructure.

6.1 Introduction

In recent years, the number of ransomware attacks has increased [35]. Many companies report significant damage [14, 40]. Besides companies, there have also been reports of ransomware attacks causing considerable damage to individuals. Often, these attacks are targeted at a specific device: the Network Attached Storage (NAS). The NAS is a device where multiple external hard drives can be placed, which are then accessible via the open internet. This allows you to easily access your files at home or on the go. NAS devices are an interesting alternative to cloud providers. In addition to storage and easy access, NAS is also used for making backups. The most well-known brands for NAS devices are QNAP, Netgear, Western Digital, Synology, and Asustor (see Figure 6.1).

Since NAS devices are accessible via the internet, already they are also vulnerable to various types of malware, such as ransomware. In previous years, the number of reports to the police of NAS related ransomware has increased from 6 in 2019 to 33 in the first seven months of 2022 [42]. Although these figures may present a skewed view due to potentially low reporting rates, other ransomware studies also detect an increase in the number of attacks with ransomware targeting NAS devices, or NAS ransomware for short [22, 19, 54, 58]. NAS ransomware attackers seem to request a relatively small amount of ransom around €500 [21, 11]. Since victims of NAS ransomware generally do not suffer material damage, such attacks might have not been prioritized for investigation by police agencies.

NAS ransomware represents a specific type of Internet of Things (IoT) attack [51]. The term IoT broadly refers to relatively small devices connected to the internet. Several studies have investigated IoT attacks, primarily focus-

Figure 6.1: Different brands of NAS devices: Netgear, Synology, Asustor, and Western Digital [27]

ing on persistent malware on IoT devices, such as botnets and cryptojackers [3, 50]. Persistent malware is designed to remain undetected on a computer or IoT device for extended periods of time, operating in the background without the user's knowledge. Such malware often establishes a botnet, a network of infected devices that can be controlled by the attacker. A cryptojacker, for instance, is a type of malware that covertly uses the IoT device's resources to mine cryptocurrency [48].

[3] investigated the IoT malware lifecycle, as it can be installed on NAS devices. The authors found that many IoT devices do not have a graphical user interface, making it difficult for users to set a password. Moreover, many network devices were found to be vulnerable to command injection, where arbitrary codes can be executed on the device. In addition, many devices had default passwords, which attackers could exploit using a brute-force attack. A brute-force attack involves using computing power to try as many passwords as possible to gain access to an account. Since NAS is connected to the open internet, this type of attack has a relatively large likelihood of success.

Previous insights into NAS ransomware attacks are based on research from cybersecurity companies [11, 10, 1]. To our knowledge, there has been no systematic scientific empirical research on NAS ransomware. Therefore, this study will focus on the following main research question:

Main Research Question: How do NAS ransomware attacks differ from regular ransomware attacks?

To answer the main research question, we consider Routine Activity Theory (RAT) [18]. RAT suggests that crime occurs when three elements converge in time and place: a motivated attacker, a suitable target, and the absence of guardianship. Although RAT was originally developed for offline crime, its applicability to cybercrime in a virtual environment has been debated [34, 60]. Nevertheless, we justify using RAT in the present study by using it as a framework to compare regular and NAS ransomware:

1. **Motivated Attacker.** For the element of *motivated attacker*, we pose sub-question 1: What is the difference in the modus operandi of the attacker between NAS ransomware and regular ransomware?

2. **Suitable Target.** For the element of *suitable target*, we pose sub-question 2: What is the difference in victim characteristics between NAS ransomware and regular ransomware?

3. Place and Time. The element of *place* in the NAS ransomware context is represented by the NAS device, as this is where the ransomware attack occurs. Based on the element of *time*, we pose the following sub-question 3: How does the development over time differ between NAS ransomware and regular ransomware?

4. Guardianship. For the element of *guardianship*, we examine a Dutch police intervention against a NAS ransomware variant named DeadBolt. Comparing results with police interventions against regular ransomware as studied in [38] will give us insights into potential differences of police interventions between NAS and regular ransomware. Sub-question 4: What is the effect of the DeadBolt operation on the number of NAS ransomware attacks?

We analyse data from two sources: the Dutch police and the Shodan internet scanner platform. The police data covers the period from January 1, 2019, to January 1, 2023, including 104 ransomware incidents (20.7%) targeted NAS devices, while 398 (79.3%) targeted other IT infrastructure, categorized as 'regular ransomware' in the present study. Additionally, we use a dataset from the internet scanner platform Shodan covering the period from December 2021 to January 2024, containing monthly data on the number of DeadBolt ransom notes [10]. The Shodan dataset allows us to analyse the effect of the Dutch police intervention against Deadbolt ransomware [11].

With the present study we hope to highlight the growing issue of ransomware targeting NAS devices, primary for two reasons: First, from a scientific perspective, comparing different types of ransomware attacks provides valuable insights into the operational methods of cyber attackers [30]. Second, understanding these attack methods is essential for developing targeted prevention strategies [38].

The outline of this paper is as follows: Section 2 reviews previous research. Section 3 details the data and methods used. Section 4 presents the results. Section 5 discusses the strengths and weaknesses of the research and offers suggestions for future studies. Section 6 draws conclusions regarding the extent to which the results support our hypotheses. Finally, Section 7 reflects on the insights gained from this study, as encouraged by RAT, to provide recommendations for various stakeholders involved in NAS ransomware: users, NAS device vendors, and local government agencies.

6.2 Previous Research

6.2.1 Modus operandi of regular and NAS ransomware

To investigate how regular and NAS ransomware differ, we first examine the modus operandi of the attacks using a crime script analysis [15, 25]. A crime script breaks down an attack into consecutive stages, with each stage representing a necessary step for executing the attack [15]. For ransomware attacks, several crime scripts have been created [31, 39, 36]. Keshavarzi and Ghaffary (2020) divide the ransomware attack into six steps:

1. **Infection**: This is the initial step where the victim's system is infected using methods such as spam emails, exploit kits, removable media, malvertising, and others.

2. **Installation**: The ransomware is installed by evading processes, modifying the registry, manipulating memory, or using a malicious installer.

3. **Communication**: The ransomware communicates with the attacker via hardcoded IP addresses or Domain Generation Algorithms (DGA) to exchange keys or send exfiltrated data.

4. **Execution**: The ransomware lists directories, encrypts or locks files, steals credentials, and modifies the Master Boot Record (MBR), rendering files inaccessible.

5. **Extortion**: The attacker demands a ransom, typically in exchange for decrypting the files, often utilizing anonymous networks to ensure untraceable payments.

6. **Emancipation**: Depending on the attacker's decision, they either decrypt the files after payment, unlock the system, or leave the files encrypted with no further action.

A disadvantage of this type of crime script is that it primarily focuses on encrypting a single computer, rather than considering the entire ICT network, whether segmented or not. Two steps can be added to this script. First, when encrypting an ICT network, lateral movement becomes important, as the malware must now gain access to the entire network of computers [1]. Lateral movement refers to the step-by-step process of gaining access to multiple computers across the network in order to achieve full control.

A second extension to this model is data exfiltration. Attackers sometimes download sensitive data from victims to their own servers [39]. This data often includes sensitive documents, such as passports, pay slips, and other business-critical information. If the victim refuses to pay, attackers may threaten to release the data publicly on a so-called leak site. Since 2019, this tactic has frequently been used in ransomware attacks [35, 37]. This extortion scheme is commonly referred to as double-extortion ransomware [43].

Compared to regular ransomware, a NAS ransomware attack typically consists of the following steps [5, 19, 10]:

1. **The reconnaissance phase**: The aim here is to identify potential victims. A common method involves using internet scanners like Shodan or Censys, which display NAS devices connected to the internet (see Figure 6.2). Shodan allows searches by IP address or for specific devices, such as NAS systems. This is often the first step in executing a NAS ransomware attack.

2. **Gaining access to the device and uploading the malware**: Once access is obtained, the malware can automatically encrypt the NAS drive.

3. **Waiting for the victim to pay**: After the encryption, the attacker demands a ransom from the victim.

4. **Money laundering**: The final step involves laundering the ransom money to make it untraceable.

Comparing the crime scripts, we could infer that the primary distinction between NAS ransomware and regular ransomware might be the degree of automation. NAS ransomware largely operates automatically, whereas a regular ransomware attack requires a more hands-on approach. In traditional attacks, the attacker invests time in lateral movement—gaining access to the entire network, including all connected computers and, if possible, any backups. Additionally, it appears that data exfiltration does not typically occur in NAS ransomware attacks [19]. Lastly, NAS ransomware generally does not involve negotiation processes.

Three well-known NAS ransomware variants are Ech0raix, Qlocker, and DeadBolt. In the present, we refer to a ransomware strain by using the file extension after encryption with ransomware.

Ech0raix has been active since June 2019 and is a form of ransomware-as-a-service (RAAS). This means that the ransomware can be purchased for a percentage of the ransom. Old Ech0raix attacks consisted mainly of brute-force attacks

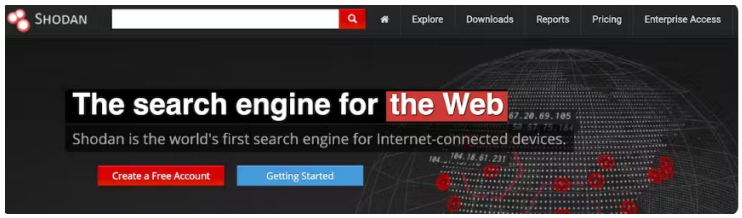


Figure 6.2: Screenshot of the search engine Shodan [53]

[19], where passwords are discovered through trial-and-error. The first versions of Ech0raix ransomware checked if the NAS is in a CIS country, a former Soviet country [54]. In these versions, the program does not encrypt data [4].

There appear to be different campaigns of Ech0raix ransomware attacks, where a campaign is defined as a sudden increase in attacks over a short period of time, often exploiting a specific vulnerability. One such alleged campaign occurred in February 2021, targeting vulnerabilities in NAS devices from the brand QNAP (CVE-2020-2501) [19]. This vulnerability involves a buffer overflow that does not require login credentials to access the device. A buffer overflow is a flaw that can be exploited to gain unauthorized access to a computer's memory. Another potential campaign in May 2021 exploited a three vulnerabilities to gain access to NAS devices [58].

Qlocker has been active since April 2021 and targets only QNAP NAS systems. According to Schouw [52], Qlocker has earned approximately €350,000. The attackers asked their victims for about 0.02-0.03 Bitcoin, which is approximately €600 to €1000 [52]. In December 2021, a series of attacks with Qlocker ransomware occurred, where the attackers accessed the NAS device using CVE-2021-28799, enabling them to access the credentials of the NAS device.

DeadBolt, active since December 2021, primarily targets QNAP NAS systems by exploiting a zero-day vulnerability [33]. Unlike other ransomware strains, DeadBolt employs a "spray and pray" approach, targeting small businesses and individuals rather than large organizations [33, 11]. Initially, the group demanded 0.03 Bitcoin (approximately €1000), later increasing the demand to 0.05 Bitcoin due to fluctuations in the BTC-USD exchange rate [5]. Despite warnings from QNAP, the Dutch police estimated that by October 2022, there were around 15,000 DeadBolt victims globally, including 1,100 in the Netherlands [45]. In the first half of 2022, DeadBolt earned approximately $187,000 [10], and by October 2022, Gomez et al. [21] identified 2,503 payments to DeadBolt,

amounting to an estimated $2.47 million in profits. The Dutch police intervened in October 2022, the impact of which will be further evaluated in the present study.

6.2.2 Theoretical Framework

To explain the behavior of attackers and victims, criminological theories are important, also for online crime [23, 24, 59]. The main theories for this research are Routine Activity Theory (RAT) and the Rational Choice Theory (RCT). Both theories help us form expectations about how attackers behave and how their targets differ. These expectations are directly linked to our hypotheses in the next section.

Routine Activity Theory (RAT), developed by Cohen and Felson [12, 18], provides insight into the causes of crime: to commit a crime, a motivated attacker must come together in time and space with an attractive target. In the absence of effective controls, the motivated attacker will target a suitable target [12]. Potential targets can be persons or objects that attackers perceive as vulnerable or particularly attractive. RAT emphasizes the importance of exposure to and vulnerability of targets and victims to potential attackers and reduced supervision to explain crime [2, 24, 44, 49]. From this perspective, it makes sense that potential victims who are highly visible online, such as financial institutions, have substantial financial resources, and may lack robust security measures, like hospitals, are particularly attractive to motivated attackers. As a result, they have a higher likelihood of becoming victims of attacks, such as ransomware [28, 34, 49].

As described above, we must be cautious with applying RAT to cybercrime [34, 60]. According to Yar [60], the basic elements of RAT remain the same in cybercrime, but these elements may be interrelated differently because the crime environment is virtual, and especially the elements 'absence of supervision,' 'place,' and 'space' might be fundamentally different than in the offline world. However, since this is an exploratory study and the elements of RAT are well defined in the context of NAS ransomware, we use these elements of RAT in the present study.

In the context of NAS ransomware, RAT suggests that NAS devices are "suitable targets" because they often have numerous vulnerabilities, making them easier to exploit. Additionally, many NAS devices are exposed to the internet without sufficient security measures, representing the "absence of capable guardianship" [34, 12]. Vendors may prioritize selling new products over securing existing ones, leaving these devices more vulnerable. This leads us to expect that NAS ransomware primarily exploits targets that are less protected and easier to

access compared to the larger, better-secured organizations often targeted by regular ransomware.

Rational Choice Theory (RCT) focuses on the decision-making process of attackers. According to Cornish and Clarke [16, 38], attackers weigh the costs and benefits of committing a crime and make rational choices to optimize their gain. Attackers adjust their strategies based on the potential rewards, the risks involved, and their ability to execute the crime. RCT allows us to explore the reasoning behind why different types of ransomware attackers target different victims.

In the context of NAS ransomware, RCT indicates that attackers target less protected and easier-to exploit devices, such as those owned by individuals or small businesses with weak security. While these targets may pay less [37], the lower effort required and potential for automation make high-volume attacks profitable[16]. In contrast, regular ransomware typically targets businesses with stronger security measures and demands higher ransoms. This aligns with the high-risk, high-reward model predicted by RCT, where attackers are willing to accept greater risks for the chance of significant financial gain.

6.2.3 Hypotheses

As outlined in the previous section, RAT and RCT suggest that the attack strategies for regular and NAS ransomware may differ, potentially leading to the use of different types of ransomware variants in NAS ransomware attacks compared to regular ransomware attacks. Accordingly, in the previous section we highlighted ransomware variants such as Ech0raix, QLocker, and DeadBolt, which are generally associated with NAS ransomware [5, 10, 19, 58]. In contrast, variants like Conti, Lockbit, and Revil are more commonly linked to regular ransomware [39, 37]. These findings suggest different ransomware variants associated with NAS ransomware versus regular ransomware. This leads us to the following hypothesis:

Hypothesis 1: Different ransomware variants are likely to be involved in NAS ransomware attacks than those that are involved in regular ransomware attacks.

RCT suggests that NAS ransomware attackers may adopt a volume-based strategy, targeting a large number of less-protected devices with relatively low ransom demands. This strategy maximizes overall profit through high attack volume rather than relying on large payouts from individual victims. Since individuals and small businesses typically have weaker security measures and more limited financial resources, NAS ransomware attackers view them as easy, low-risk targets. There is empirical evidence supporting the hypothesis that NAS

ransomware attacks demand smaller ransoms compared to regular ransomware attacks. Sowells [54] reports that Ech0raix demands between 0.05 and 0.06 bitcoins (approximately €500 to €600) for decryption. In contrast, Meurs [39] found that (regular) ransomware attacks on businesses in the Netherlands demanded an average ransom of €720,256 (sd = €2,632,673). There are several possible explanations for this. First, individuals generally have less money than businesses, so attackers may adjust their ransom demands based on the financial capacity of their victims. Second, individuals may be less willing to pay ransoms than businesses [37]. This leads us to the following hypothesis:

Hypothesis 2: The ransom requested in a NAS ransomware attack is lower than the ransom requested in a regular ransomware attack.

From a RAT perspective, NAS ransomware victims are more likely to be individuals or small businesses, as they generally have weaker security systems compared to larger organizations. Consequently, we hypothesize that NAS ransomware disproportionately affects individuals, whereas regular ransomware is more likely to target businesses that offer more substantial financial returns but are more challenging to breach. Accordingly, Rodríguez et al. [51] studied users of IoT devices. The authors used a random sample based on 128 users of IoT devices known to an internet provider. 91.2% of the NAS users were individuals and did not use the NAS device for business. This leads us to the following hypothesis:

Hypothesis 3: Individuals are more frequently victims of NAS ransomware than regular ransomware attacks.

NAS ransomware attacks are often associated to the disclosure of new vulnerabilities. Publicly revealed vulnerabilities provide attackers with opportunities to exploit devices that have not been updated or patched. This expectation aligns with RAT's focus on the convergence of a suitable target and a motivated offender in the absence of effective protection.

The previous section noted that the three largest NAS ransomware variants—Ech0raix, Qlocker, and DeadBolt—became active in June 2019, April 2021, and December 2021, respectively, often coinciding with the disclosure of vulnerabilities [5, 10, 22, 58]. This leads us to the following hypothesis:

Hypothesis 4: There is a relationship between the disclosure of new vulnerabilities in NAS devices and the timing and number of NAS ransomware attacks.

RCT further suggests that the vulnerabilities exploited by NAS ransomware attackers may be more predictable and standardized, as attackers tend to rely on automated methods to exploit numerous devices simultaneously. This approach contrasts with regular ransomware, which often involves more tailored,

targeted approaches to infiltrating businesses with stronger security measures and custom IT infrastructures. Accordingly, empirical studies demonstrate that regular ransomware attacks do not always start by exploiting vulnerabilities [13, 39]. In contrast, NAS ransomware appears to be closely tied to newly discovered and publicly disclosed vulnerabilities. This suggests that both types of ransomware have different entry points. Consequently, we expect that the trends in NAS ransomware over time will not correlate with those of regular ransomware.

Hypothesis 5: There is no temporal relationship between the number of reports of NAS ransomware and reports of regular ransomware.

RAT also suggests that when effective guardianship is in place, the frequency of attacks declines. For example, when organizations implement robust security measures, they reduce the opportunities for successful ransomware attacks. In the present study, we examine whether police interventions can act as capable guardians by reducing the number of NAS ransomware incidents. Limited research on police interventions in the online environment supports the potential effectiveness of such actions with minimal crime displacement [38, 55]. For instance, [55] examined the takedown of the Hansa market and its impact on vendor migration to other darknet markets. Their study of 220 vendors using PGP (Pretty Good Privacy) keys revealed that only a small number migrated with the same keys, indicating limited displacement. This leads us to the following hypothesis:

Hypothesis 6: A police intervention leads to a decrease in ransomware operations, even if there are no arrests.

In summary, these theories suggest key differences in how RAT elements apply to NAS and regular ransomware, which we explore through the six hypotheses. NAS ransomware attackers appear to focus on volume-based attacks with lower ransoms, while regular ransomware attackers target larger, more secure organizations with higher financial stakes. Furthermore, the timing of NAS ransomware attacks is likely tied to the disclosure of vulnerabilities, whereas regular ransomware may not be as strongly influenced by such disclosures and could be driven by other factors.

6.3 Data and Methods

6.3.1 Sample

We use two datasets: one from the Dutch police and another from the Shodan internet scanning platform. The police dataset was created by querying police systems with the keyword 'ransomware' and manually filtering the results to focus only on crypto ransomware incidents, excluding other types such as locker

ransomware, which are beyond the scope of this study. This dataset includes information on ransomware attacks targeting both organizations (80.3%) and individuals (19.7%), providing a unique opportunity to compare these two types of attacks.

Two aspects must be considered when using police data: the willingness to report a crime to the police and the use of police data for scientific research. Reporting cybercrime is relatively uncommon, with studies showing that victims of online fraud report only 8 to 10% of cases [56]. For regular ransomware, reporting rates are around 40% for medium-sized companies and 60% for large-sized companies in the Netherlands between 2019 and 2022 [41]. [41] also conclude that individuals and small companies may be less likely to report ransomware than medium and large companies. This should be considered when interpreting the results.

Furthermore, this study uses police data for scientific purposes. The legal basis for the use of police data for scientific research is Article 22 of the Police Data Act (WPG). It states that the Board of Procurators General grants permission for the use of police data for scientific research provided that the data is published in a completely anonymized form. This requirement is met in this chapter.

Between January 1, 2019, and January 1, 2023, 525 reports and/or complaints of ransomware attacks were made to the Dutch police. We excluded 23 reports of unsuccessful ransomware attacks. There remained 502 complaints of successful ransomware attacks. Of these, 104 (20.7%) were targeted at NAS devices and 398 (79.3%) were targeted at other IT infrastructure, defined in the present study as 'regular ransomware'.

To estimate the impact of the Dutch police intervention against DeadBolt [11, 45], a second dataset from the Shodan internet scanner platform was analysed [9]. This dataset was enabled by a specific technical characteristic of DeadBolt ransomware, which was previously analysed by [21].

DeadBolt exploits network-facing vulnerabilities in NAS devices, hijacking their login pages to display a ransom note titled "WARNING: Your files have been locked by DeadBolt." Infection required the NAS to be connected to the internet, making ransom notes potentially visible to the internet scanners. This allowed a relative objective measurement of the amount of unique Deadbolt victims over time. However, not all victims may have been observed, as some infected NAS devices might have been disconnected or cleaned before scanning. Data on DeadBolt infections was collected from the Shodan internet scanner by querying for "* DeadBolt", where '*' represents a wildcard [9].

Historical Shodan data from December 1, 2021, to January 31, 2024, was collected, showing the monthly number of DeadBolt victims per country. The peak was observed in August 2022, with 11,965 ransom notes identified on unique IP addresses.

6.3.2 Variables

In the police dataset, each complaint involved a number of variables that were manually coded. These variables are grouped according to the sub-questions into victim characteristics, modus operandi of the attack, and information about the complaint, such as the time of encryption and reporting.

1. **Victim Characteristics**

    a. **Sector**: For victims the sector of the company or, for individuals, their classification as 'Individual'. The sectors are defined by the Chamber of Commerce and include: Construction, Healthcare, Trade, ICT, MAS (Environmental and Agricultural Sector), Media, Recreation, and Transport.

    b. **Type of Enterprise**: Categories include Multinational (operating in multiple countries), BV (private limited companies), SMB (small and medium-sized enterprises with 1-5 employees), Foundation, Self-employed, or Individual. No Public Limited Companies were involved in this dataset. Missing enterprise types were looked up on Zoominfo.

    c. **Backup**: Victim's backup status, categorized as: no backup, backup but unable to restore files, partial restoration, or full restoration.

    d. **Paid**: Did the victim pay the ransom, and if so, how much?

    e. **Financial Damage**: Financial damage suffered by the victim. Companies often reported an amount, while individuals mentioned immaterial damage. This is coded as either immaterial damage or a specific monetary amount, if provided.

2. **Modus Operandi**

    f. **Ransomware Variant**: Is there a name on the ransom note or an extension of the files?

    g. **Ransom**: How much ransom did the attacker ask the victim to decrypt the files?

h. **Personal Ransom Note**: How often is the bitcoin address on the ransom note? If there is no bitcoin address, the victim must contact the attacker via email or TOR chat. With regular ransomware, a ransom amount is requested depending on victim characteristics [39]. Sometimes victims are willing to pay if the ransom amount is reduced. Since the amounts for victims of NAS ransomware are likely relatively low, this will probably reduce profits for the perpetrators behind NAS ransomware than for regular ransomware.

i. **Date of Attack**: When was the device (or devices) encrypted? Recorded in date and time.

3. **Metadata**

j. **Complaint**: This refers to the status of the incident report made by the victim to the police. It can either be a simple report, where the victim informs the authorities about the ransomware attack, or a formal complaint, which may lead to further legal action. Additionally, this includes whether the complaint has initiated an official investigation by law enforcement.

k. **Time Complaint**: When did the report/complaint take place? Recorded in date and time.

For the Shodan dataset, only the monthly count of ransom notes associated with unique IP addresses per country was recorded.

6.3.3 Analysis

We use various statistical tests to compare the characteristics of regular and NAS ransomware. For categorical variables, the chi-squared test (χ^2) is used, a standard method for evaluating the independence of categorical variables, as discussed by Plackett [46]. For numerical variables, comparisons are made using the t-test, which assesses whether the means of two samples differ. A key assumption of the t-test is that the samples are approximately normally distributed. Given the skewed distribution of variables such as the amount paid and the damage in euros, a log10 transformation is applied [42, 37]. The Shapiro-Wilk test for normality is then used to determine whether this transformation achieves an approximate normal distribution. If normality is not met, the non-parametric Mann-Whitney U-test is applied as an alternative to the t-test, following Zimmerman [61]. Additionally, temporal correlations between events, including NAS ransomware attacks, are examined. The relationship between

the timing of NAS ransomware encryption and traditional ransomware is tested using the Pearson product-moment correlation, an extension of Pearson correlation for interval data, as described by Puth, Neuhäuser, and Ruxton [47]. Lastly, to evaluate whether there was a decline in the number of DeadBolt ransom notes before and after the police intervention, a two-sample t-test is conducted [17].

6.4 Results

Table 6.1 presents the various descriptive results of this study. Sector, type of enterprise, backup, paid, and financial damage differ between NAS ransomware and regular ransomware. Additionally, the ransomware variant, ransom, and personal ransom note also differ from each other. Figure 6.3 shows the frequencies of ransomware attacks for regular and NAS ransomware. At first, there seems to be no correlation between the time of encryption in regular ransomware and NAS ransomware. NAS ransomware does appear to be associated with the four events from section 2, combined with the police intervention, which we refer to here as L1 through L5:

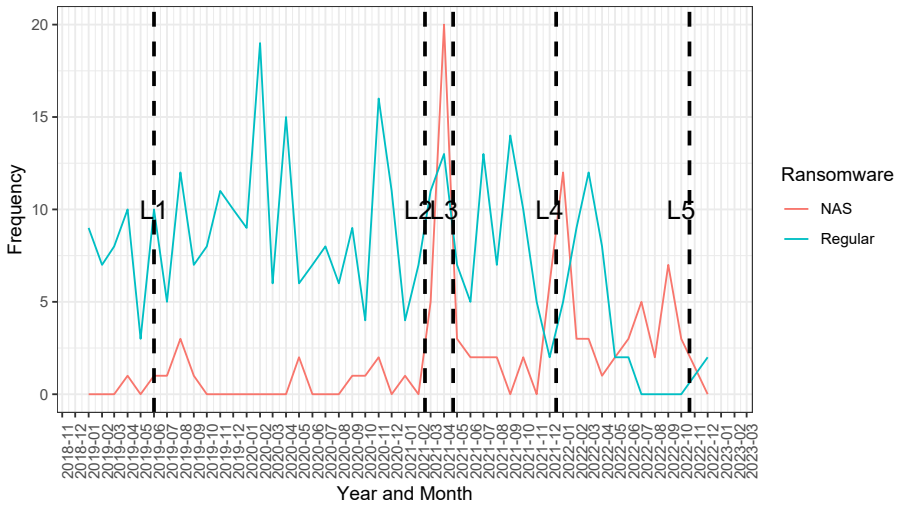


Figure 6.3: Frequency of ransomware attacks reported to the Dutch police between January 1, 2019, and January 1, 2023

Table 6.1: Comparison of Variables in NAS Ransomware and Regular Ransomware

Variables	Test, p-value	Category/Unit	NAS Ransomware	Regular Ransomware
1a. Sector	$\chi^2(10) = 157.97^{***}$	Construction	8 (8%)	48 (13%)
		Healthcare	2 (2%)	20 (6%)
		Trade	11 (11%)	105 (29%)
		ICT	4 (4%)	57 (16%)
		Agriculture	1 (1%)	11 (3%)
		Media	7 (7%)	13 (4%)
		Individuals	66 (66%)	33 (9%)
		Education	0 (0%)	14 (4%)
		Government	0 (0%)	10 (3%)
		Leisure	1 (1%)	20 (6%)
		Transport	1 (1%)	28 (8%)
1b. Business Type	$\chi^2(6) = 132.03^{***}$	BV (private limited companies)	12 (13%)	158 (42%)
		SMB	8 (8%)	75 (20%)
		Individuals	66 (69%)	33 (9%)
		Foundation	2 (2%)	8 (2%)
		Self-employed	8 (8%)	22 (6%)
		Multinational	0 (0%)	55 (15%)
		Public organization	0 (0%)	23 (6%)
1c. Backups	$\chi^2(3) = 32.71^{***}$	No backups	67 (70%)	127 (38%)
		Backups + no recovery	10 (10%)	65 (19%)
		Backups + partial recovery	9 (9%)	62 (18%)
		Backups + full recovery	10 (10%)	83 (25%)
1d. Paid	$\chi^2(1) = 14.17^{***}$	Paid	6 (6%)	80 (24%)
		Average paid euro (mean, sd)	€865 (sd=€420)	€207,709 (sd=€599,941)
		Average paid euro log10 (mean, sd)	2.89 (sd=0.22)	4.17 (sd=1.35)
1e. Financial loss	$t(392) = -7.53^{***}$	Loss not quantified	95 (91%)	233 (58%)
		Financial loss euro (mean, sd)	€4,257 (sd=€5,875)	€601,637 (sd=€3,181,015)
		Financial loss euro log10 (mean, sd)	3.27 (sd=0.61)	4.59 (sd=1.06)
2f. Ransomware strain	$\chi^2(8) = 190.28^{***}$	Top 5	Ech0raix (37%), DeadBolt (30%), Unknown (20%), Qlocker (5%), 0XXX (2%)	Unknown (32%), Phobos (9%), Revil (9%), Conti (5%), Lockbit (5%)
2g. Ransom requested	$W = 1326.5^{***}$	Ransom requested euro (mean, sd)	€1,404 (sd=€4,460)	€727,544 (sd=€2,623,434)
		Ransom requested euro log10 (mean, sd)	2.92 (sd=0.27)	4.41 (sd=1.20)
2h. Personal ransom note	$\chi^2(1) = 7.81^{***}$	No payment details on ransom note	0 (0%)	161 (45%)
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$				

- L1 (June 2019): The first wave, marked by the activation of the ransomware variant Ech0raix. Ech0raix exploited brute force methods to gain access to NAS devices by cracking login credentials.
- L2 (February 2021): The second wave began with the publication of the QNAP vulnerability CVE-2020-2501. This vulnerability allowed attackers to access devices without requiring login credentials.
- L3 (April 2021): The third wave saw the emergence of the ransomware variant QLocker, which exploited CVE-2021-28799 to access NAS device credentials.
- L4 (December 2021): The fourth wave was marked by the activation of the ransomware variant DeadBolt, which claimed to have discovered a zero-day vulnerability in QNAP NAS devices.
- L5 (October 2022): The Dutch police intervened against the ransomware variant DeadBolt.

During coding, it was noted that many victims of NAS ransomware stated that after the attack, they had searched the internet and found that a vulnerability had just been published for their specific type of NAS. Finally, the number of reports and complaints does not differ between NAS ransomware and regular ransomware ($\chi^2(3) = 5.05$, $p = 0.17$). Based on these results, we can make a statement about the outcome of the hypotheses. See Table 6.1 for an overview of the analysis.

Hypothesis 1: Different ransomware variants are involved in NAS than in regular ransomware. Based on '2f. Ransomware strain' in Table 1, we can say that the ransomware variants between regular and NAS ransomware differ. A significant portion of NAS attacks are carried out with Ech0raix, at 37%, while in regular ransomware there is more distribution across variants, with Unknown contributing the most at 32%. The findings support Hypothesis 1.

Hypothesis 2: The ransom requested in NAS ransomware attacks is lower than in regular ransomware attacks. As shown in Table 1, '2g. Ransom requested', the average ransom for regular ransomware attacks is €727,544, compared to just €1,404 for NAS ransomware. This means the ransom in regular attacks is, on average, 518 times higher. The findings support Hypothesis 2.

Hypothesis 3: More individuals are victims of NAS ransomware than of regular ransomware. As expected, '1a. Sector' (see Table 1) shows that 66% of NAS attacks target individuals, while only 9% of regular attacks affect individuals. The findings support Hypothesis 3.

Hypothesis 4: There is a correlation between the number of NAS ransomware attacks and the publication of vulnerabilities related to NAS devices.

Hypothesis 5: There is no temporal correlation between the number of NAS ransomware complaints and regular ransomware complaints.

Since both hypotheses rely on the same data, they are discussed together. To examine the temporal developments, Pearson product-moment correlations were calculated. Before applying the Pearson correlation, the time series were tested for stationarity, as only stationary data can be compared directly [57]. If the series were not stationary, differencing would be required. Both the Augmented Dickey-Fuller (ADF) and KPSS tests were used to assess stationarity. The KPSS test results indicated that both time series are stationary (NAS ransomware KPSS = 0.54, regular ransomware KPSS = 0.39), as the test statistics were below the critical value for stationarity. Similarly, the ADF test results supported stationarity for the NAS ransomware series with a Dickey-Fuller Statistic of -2.95 and a p-value of 0.06, while the regular ransomware series had a Dickey-Fuller Statistic of -2.55 with a p-value of 0.36.

- A correlation of $\rho = 0.42$ ($t(26) = 2.34$, $p = 0.03$) was found between events L1-L5 and NAS ransomware. The 95% confidence interval for this correlation is $\rho = 0.05 - 0.68$.
- No correlation was found between events L1-L5 and regular ransomware: $t(43) = 0.03$, $p = 0.97$.
- Additionally, there was no temporal correlation between regular and NAS ransomware: $t(47) = -0.17$, $p = 0.87$.

In summary, the data support Hypothesis 4, indicating a temporal correlation between events L1-L5 and NAS ransomware. However, consistent with Hypothesis 5, no correlation was found between events L1-L5 and regular ransomware, nor between regular and NAS ransomware. In Figure 6.3, the frequency of ransomware attacks reported to the Dutch police from January 1, 2019, to January 1, 2023, is presented. This concerns the date of file encryption, not the moment when the victim files a report. The lines L1-L5 represent specific events that seem to influence the number of encrypted NAS devices (see section 2).

DeadBolt Police Intervention

As previously described, DeadBolt is a ransomware variant commonly associated with NAS ransomware attacks. One of its key characteristics is the lack

of communication or negotiation with victims. Rather than setting up negotiation websites, DeadBolt directs victims to pay a specified Bitcoin amount to a provided address. Upon payment, the decryption key is automatically sent via the blockchain, embedded in a low-value Bitcoin transaction in the OP_RETURN field of the ransom address [11]. However, this unique method became a vulnerability that the Dutch National Police exploited to recover data for hundreds of victims without payment [45, 21, 11]. Cyber investigators from the Dutch police noticed that DeadBolt provided decryption keys before the victims' payments were confirmed on the blockchain, during the time when transactions are visible in Bitcoin's mempool but not yet finalized [11].

The police devised a plan using the replace-by-fee (RBF) technique to exploit this mechanism, sending and retracting payments to recover decryption keys without completing the ransom transactions. In collaboration with Europol, the operation involved identifying as many DeadBolt victims as possible and using a script to automate the send-and-retract process. This intervention, carried out in October 2022, successfully retrieved decryption keys for nearly 90% of the reported victims to Europol between January 2022 and October 2022 — 155 keys in total — disrupting DeadBolt's operations and forcing them to adopt a more manual, and therefore costlier, process for key distribution.

Figure 6.4 illustrates the number of DeadBolt ransomware incidents found by Shodan in various countries from December 2021 to January 2024. It shows a rise in incidents starting from early 2022, peaking around July 2022, followed by

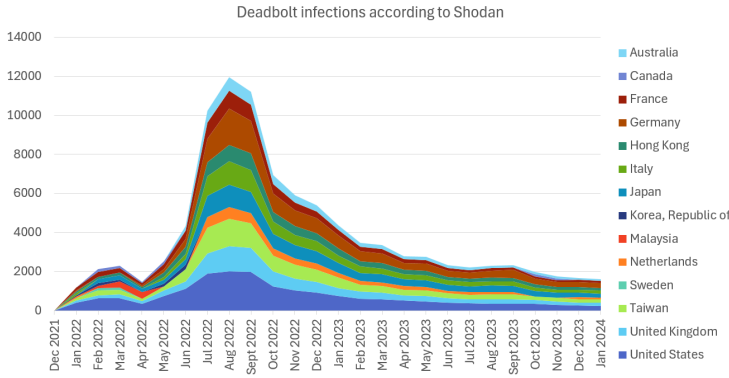


Figure 6.4: Frequency of ransomware attacks based on ransom notes of DeadBolt on Shodan between December 2021 and January 2024

a gradual decline. The countries represented include Australia, Canada, France, Germany, Hong Kong, Italy, Japan, The Republic of Korea, Malysia, the Netherlands, Sweden, Taiwan, the United Kingdom and the United States. The police intervention against DeadBolt ransomware occurred in October 2022, leading to a decline in incidents post-intervention. This is statistically supported by a two-sample t-test, which revealed a difference in the number of incidents before and after the intervention ($t(23) = 2.31$, $p = 0.0288$). Based on these findings, we can assess the outcome of Hypothesis 6:

Hypothesis 6: A police intervention results in a reduction of ransomware operations, even in the absence of arrests.

As expected, a decrease in DeadBolt ransom notes was observed following the police intervention. However, the interpretation of these results is constrained by limitations in both datasets. The police data is limited by its time period, preventing a comprehensive long-term analysis. Additionally, the Shodan data does not capture other (NAS) ransomware variants, making it difficult to definitively rule out the possibility of crime displacement. In other words, the attacker(s) behind DeadBolt may have shifted to another ransomware variant or a different type of malicious activity.

6.5 Discussion

The present study explores the differences between regular and NAS ransomware using victim reports filed with the police and data from the Shodan internet scanner. Police data provide a unique opportunity to compare these types of ransomware, as both individuals and businesses report incidents to law enforcement. In contrast, data from cybersecurity companies may be less representative, as such reports tend to focus on businesses, since individuals often cannot afford these services.

However, police reports have limitations, particularly selection bias [41, 32, 56]. By comparing the police dataset with that from the internet scanner Shodan, we can infer the willingness to report NAS ransomware to the Dutch police. The Shodan dataset shows a peak of 604 DeadBolt infections in the Netherlands in August 2022 (see Figure 6.4), while the police estimated a total of 1,100 infected devices by October 2022, using data from various internet scanners [45, 11]. However, the police dataset includes only 31 DeadBolt reports between January 2020 and January 2023, suggesting a reporting rate of 5.1% based on 604 infections and 2.8% based on the police's estimate of 1,100 devices. These rates align with previous research showing similar reporting levels for cybercrime, ranging from 5 to 10% [56].

One possible explanation for the low reporting rates among NAS ransomware victims is the perception that the police cannot help NAS ransomware victims. This hypothesis is supported by media coverage of the DeadBolt police intervention [45]. Our police dataset revealed that after the news covered the Deadbolt police intervention, 18 additional reports were filed within one month, compared to only 12 reports in the previous 10 months. Although 18 reports represent just 2% of the total Dutch DeadBolt victims in our dataset, the increase suggests that media coverage encouraged more victims to report. Police officers disclosed to the authors of the present study that many victims had initially refrained from reporting, believing law enforcement would be unable to assist. The successful intervention helped change this perception, but the increase in reports remained relatively small (2.0%). This belief in police ineffectiveness likely contributes to the overall low reporting rate for NAS ransomware incidents.

In contrast, reporting rates for regular ransomware, especially among medium and large businesses, are significantly higher, ranging from 40% to 60% [37, 41]. Prior research also shows that more severe cybercrime incidents are reported more frequently than less severe cybercrime [29, 32]. This discrepancy in reporting rates between NAS and regular ransomware could affect the accuracy of our findings when comparing these two types of ransomware.

For example, victims who pay the ransom may be even less likely to report the crime to the police than those who do not, introducing a bias in comparing NAS and regular ransomware. This hypothesis is supported by the discrepancy between the $2,47 Million profits observed on the blockchain from Bitcoin payments to DeadBolt [21] and the low number of reported payments in our study, in total around $2,500. Future research should explore alternative datasets, such as direct surveys of NAS device owners, to mitigate the bias present in police reports.

Another limitation of the study is that the police dataset was coded by a single individual, primarily due to the sensitive nature of the data. This approach can introduce several potential biases [6]. For instance, confirmation bias might occur if the coder unintentionally favors interpretations that align with their own preexisting beliefs. Additionally, personal bias can influence how data is categorized or interpreted based on the coder's own cultural and personal background. By discussing doubtful cases anonymously with the research group, we hoped to minimize these biases.

The observed differences between NAS and regular ransomware align with both Rational Choice Theory (RCT) and Routine Activity Theory (RAT). Our findings suggest that NAS ransomware attackers employ a volume-based strategy, characterized by lower ransom demands and more automated, scalable attacks.

This corresponds with RCT, as NAS ransomware attackers tend to target individuals with poorly secured NAS devices, requiring less effort but allowing for a larger number of attacks, thereby maximizing overall profit through volume rather than high individual payouts [16]. In contrast, regular ransomware, which often targets businesses and demands significantly higher ransoms, fits the traditional high-risk, high-reward model predicted by RCT.

RAT further helps explain differences between regular and NAS ransomware [12, 34, 60]. NAS ransomware victims, typically individuals or small businesses with insufficient security measures, present more frequent and accessible "suitable targets," while NAS devices are often exposed on the internet, increasing their vulnerability ("lack of capable guardianship"). In contrast, regular ransomware attacks are more likely to involve larger organizations with more robust security infrastructures, which are harder to exploit but offer larger rewards. The differences in victim characteristics and attack strategies between NAS and regular ransomware reflect these theories, with NAS ransomware following a model where ease of access and frequency of vulnerable targets drive the attacker's decision-making.

Finally, the decline in DeadBolt ransomware victims following the police intervention aligns with previous research on regular ransomware [38]. Analyzing a dataset of ransomware victims, [38] assessed five types of interventions: arrests, server takedowns, asset freezes, decryptor releases, and sanctions. Ransomware groups typically responded by ceasing operations, continuing, or rebranding. Nearly half of the interventions resulted in groups ceasing their activities. Furthermore, the authors observed that few ransomware groups rebranded, suggesting minimal crime displacement. The findings from this study and [38] should be replicated in future research to test generalizability and validity, but they offer a promising perspective for law enforcement and policymakers to combat both regular and NAS ransomware.

6.6 Conclusion

This study aimed to answer the **main research question**: what are the key differences between regular ransomware and NAS ransomware in terms of modus operandi, victim characteristics, and trends? Additionally, we explored the impact of the DeadBolt police intervention on NAS ransomware activity.

The **first sub-question** was: What are the differences in modus operandi between regular and NAS ransomware? The findings support Hypothesis 1, showing that different ransomware variants are involved in NAS ransomware attacks compared to regular ransomware. Hypothesis 2 was also confirmed,

revealing that NAS ransomware operates with a more automated and scalable revenue model. NAS ransomware demands lower ransom amounts and uses standardized ransom notes, reflecting a volume-based strategy.

The **second sub-question** was: How do victim characteristics differ between regular and NAS ransomware? The results support Hypothesis 3, indicating that NAS ransomware disproportionately targets individuals, while regular ransomware is more often aimed at businesses, particularly medium and large companies.

The **third sub-question** was: What are the trends between regular and NAS ransomware? The analysis confirms Hypothesis 4, showing a strong correlation between NAS ransomware attacks and the publication of vulnerabilities in NAS devices. This trend was not observed with regular ransomware, leading to the rejection of Hypothesis 5, which proposed that both ransomware types would follow similar patterns over time.

The **fourth sub-question** was: What was the impact of the DeadBolt police operation on NAS ransomware activity? Hypothesis 6 was confirmed, as the police intervention, which involved retrieving decryption keys without payment, led to a reduction in NAS ransomware activity. This demonstrated that law enforcement can disrupt ransomware operations, even without arrests.

One implication of this study is that scientific research focusing on ransomware should attempt to distinguish between regular and NAS ransomware. The different modus operandi, victim characteristics, and trends are sufficiently distinct between regular and NAS ransomware that empirical research likely needs to correct for these differences.

Furthermore, our study we used the Shodan internet scanner data to complement the police dataset. Future research could leverage internet scanners to investigate other emerging types of ransomware, such as NAS ransomware. While conventional ransomware typically does not make ransom notes accessible via the HTTP protocol, preliminary searches suggest that Shodan and Censys can reveal more than just ransom notes for NAS ransomware. By using these platforms to directly identify encrypted systems, researchers could uncover various new types of ransomware and differentiate them from conventional types. This approach may lead to a new taxonomy of ransomware, providing a broader understanding of the ransomware landscape and aiding in the development of targeted prevention strategies.

Additionally, research focusing on IoT malware can look more specifically at the differences between NAS ransomware and other types of IoT malware. In the present study, we looked at the differences between regular and NAS ransomware, but there may also be differences in modus operandi, victim characteristics,

and trends between NAS ransomware and other types of IoT malware. Investigating these differences can help in more effectively combating the various types of malware.

In conclusion, NAS ransomware differs from regular ransomware in several key areas: it involves distinct variants, follows a more automated revenue model, targets individuals more frequently, and is closely linked to the disclosure of vulnerabilities. Additionally, police interventions, such as the DeadBolt operation, can significantly reduce NAS ransomware activity (see Table 6.2).

Table 6.2: Summary of findings on the differences between NAS and regular ransomware

NAS Ransomware	**Regular Ransomware**
Most victims are individual citizens (66%).	Most victims are businesses (91%).
Average ransom: €1,404 (sd = €4,460).	Average ransom: €727,544 (sd = €2,623,434).
Ransom paid in only 6% of cases.	Ransom paid in 24% of cases.
No backups in 70% of cases.	No backups in 38% of cases.
Victims often suffer no direct financial damage beyond the ransom, but significant immaterial damage, such as loss of personal videos and photos (70%).	Average financial damage: €601,637 (sd = €3,181,015).
Ransom and bitcoin address provided directly on the ransom note (100%).	Ransom and bitcoin address only revealed after contact with attackers in 45% of cases.
Fewer steps involved in the attack: reconnaissance and encryption.	More steps involved in the attack: reconnaissance, persistence, lateral movement, data exfiltration, encryption, and negotiations.
4 attack campaigns over the last three years, associated with newly discovered vulnerabilities in NAS devices.	A small upward trend over the years, with no observed correlation to vulnerabilities of NAS devices.

6.7 Recommendations

In this section, we provide recommendations for three groups of stakeholders involved with NAS ransomware: users, vendors of NAS devices, and local government authorities.

6.7.1 Users of NAS Devices

Common cybersecurity advice includes using strong passwords, antivirus software, and keeping your software up-to-date [20]. Based on the present research, we can draw the following conclusions:

- A strong password is often not enough to prevent NAS ransomware. While it is effective against brute-force attacks, we observe a trend of attackers focusing more on vulnerabilities that bypass the password. Thus, while still advisable, a strong password is insufficient on its own.
- Updating NAS device software is crucial to prevent victimization by ransomware. Users often know this but still fail to update their software sufficiently [51]. Since NAS devices often have a functionality to update automatically, it is wise for users to enable this feature.
- It would be beneficial to shield the NAS device from the open internet, making it unsearchable via engines like Shodan and Censys. One way to create a barrier between the open internet and the NAS device is to set up the NAS device as a VPN server or by setting up a personal reverse proxy server.

6.7.2 Vendors of NAS Devices

Vendors selling NAS devices should inform customers about the risks of malware associated with the use of NAS. It is important to mention the types of malware that can affect a NAS and how to minimize the risk of malware and ransomware. As of writing this article, major vendors still do not provide information about potential risks and vulnerabilities of NAS equipment on their websites or in their stores. The national government could discuss with major vendors to agree on informing (potential) buyers of NAS devices. Manufacturers can also assist users by setting up automatic updates, alerts, and VPN solutions.

6.7.3 Local Government Authorities

Local government authorities can initiate prevention campaigns specifically aimed at preventing cybercrime among individuals. Awareness of NAS ransomware could lead to fewer victims. It is important for prevention campaigns to

address the risk perception of the target group, as this seems to correlate with the effectiveness of the campaigns [26]. Moreover, merely notifying NAS users is not enough. Rodríguez et al. [51] found that only 24% of NAS users took all self-protective measures after multiple notifications. This indicates that awareness alone is insufficient for self-protective measures against NAS ransomware. Additionally, only a relatively small group uses NAS devices. Local government authorities could also engage with residents in their area about how to safely store data and information in general. NAS ransomware, as described in the present study, can serve as a case study on how attackers operate and what you can do to avoid becoming a victim. This alternative approach not only combats NAS ransomware but also addresses other cybercrimes related to data storage, such as data theft [8].

Regarding local government authorities, the local police could play an important role in responding to ransomware incidents. Given the emotional impact of such attacks, especially when personal data is involved, it would be beneficial for the police to offer victims emotional support [7]. Additionally, the police can use the information from reports to monitor for patterns of attacks linked to the disclosure of specific vulnerabilities and subsequently inform manufacturers, users, and other government agencies about emerging threats.

Our study also underscores the value of innovative police interventions, such as the operation against the DeadBolt strain. These actions not only have the potential to reduce the number of ransomware attacks but also enable the recovery of victims' files without the need for ransom payments. Such proactive strategies could be crucial in minimizing both the financial and emotional harm caused by ransomware.

This page is intentionally left blank.

Bibliography

[1] Akamai. *What is ransomware?* Retrieved November 22, 2022, from https://www.akamai.com/our-thinking/cybersecurity/what-is-ransomware. 2021.

[2] N. Akdemir and C. J. Lawless. 'Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: A lifestyle routine activities approach'. *Internet Research*, 2020.

[3] O. Alrawi, C. Lever, K. Valakuzhy, K. Snow, F. Monrose and M. Antonakakis. 'The Circle Of Life: A Study of The Malware Lifecycle'. *30th USENIX Security Symposium (USENIX Security 21)*. 2021, pp. 3505–3522.

[4] A. Andrioaie. *An Increased Wave of eCh0raix Ransomware Attacks Hits QNAP NAS Devices*. Retrieved August 26, 2022, from https://heimdalsecurity.com/blog/more-ech0raix-ransomware-attacks-hit-qnap-nas-devices. 2021.

[5] I. Arghire. *QNAP Warns NAS Users of DeadBolt Ransomware Attacks*. Retrieved August 26, 2022, from https://www.securityweek.com/qnap-warns-nas-users-deadbolt-ransomware-attacks. 2022.

[6] R. Artstein and M. Poesio. 'Bias decreases in proportion to the number of annotators'. *Proceedings of FG-MoL 2005: The 10th conference on Formal Grammar and The 9th Meeting on*. Vol. 139. 2009.

[7] J. Borwell, J. Jansen and W. Stol. 'Comparing the victimization impact of cybercrime and traditional crime: Literature review and future research directions'. *Journal of Digital Social Research* 3.3, 2021, pp. 85–110.

[8] J.-W. Bullée and M. Junger. 'How effective are social engineering interventions? A meta-analysis'. *Information & Computer Security* 28.5, 2020, pp. 801–830.

[9] Censys. *Censys*. Retrieved March 4, 2024. 2022. URL: https://censys.io/.

[10] Censys. *Tracking Deadbolt Ransomware Across the Globe*. Retrieved November 11, 2022, from https://censys.io/tracking-deadbolt-ransomware-across-the-globe. 2022.

[11] Chainalysis Team. *Deadbolt Ransomware Strain Tricked into Giving Up Decryption Keys*. Accessed: 2024-08-08. 2023. URL: https://www.chainalysis.com/blog/deadbolt-ransomware-strain-tricked-into-giving-up-decryption-keys/.

[12] L. E. Cohen and M. Felson. 'Social change and crime rate trends - routine activity approach'. *American Sociological Review* 44, 1979, pp. 588–608.

[13] L. Y. Connolly, D. S. Wall, M. Lang and B. Oddson. 'An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability'. *Journal of Cybersecurity* 6.1, 2020, tyaa023.

[14] L. Connolly, D. Wall, M. Lang and B. Oddson. 'An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability'. *Journal of Cybersecurity* 6.1, 2020, pp. 1–18.

[15] D. B. Cornish. 'The procedural analysis of offending and its relevance for situational prevention'. *Crime Prevention Studies* 3.1, 1994, pp. 151–196.

[16] D. B. Cornish and R. V. Clarke. 'Rational choice perspective'. *Environmental criminology and crime analysis*. Ed. by R. Wortley and L. Mazerolle. Abingdon, UK: Willan, 2008.

[17] N. Cressie and H. Whitford. 'How to use the two sample t-test'. *Biometrical Journal* 28.2, 1986, pp. 131–148.

[18] M. Felson and L. E. Cohen. 'Human ecology and crime: A routine activity approach'. *Human Ecology* 8.4, 1980, pp. 389–406.

[19] S. Gatlan. *New eCh0raix Ransomware Brute-Forces QNAP NAS Devices*. Retrieved August 26, 2022, from https://www.bleepingcomputer.com/news/security/new-ech0raix-ransomware-brute-forces-qnap-nas-devices. 2021.

[20] S. Gatlan. *QNAP warns of eCh0raix ransomware attacks, Roon Server zero-day*. Retrieved November 12, 2022, from https://www.bleepingcomputer.com/news/security/qnap-warns-of-ech0raix-ransomware-attacks-roon-server-zero-day. 2021.

[21] G. Gomez, K. van Liebergen and J. Caballero. 'Cybercrime bitcoin revenue estimations: Quantifying the impact of methodology and coverage'. *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*. 2023, pp. 3183–3197.

[22] S. Hilt and F. Merces. *Backing Your Backup. Defending NAS Devices Against Evolving*. Retrieved December 13, 2022, from https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/reinforcing-nas-security-against-pivoting-threats. 2021.

[23] T. J. Holt, A. M. Bossler and K. C. Seigfried-Spellar. *Cybercrime and digital forensics: An introduction*. Abingdon, Oxon: Routledge, 2015.

[24] T. J. Holt, J. van Wilsem and E. R. Leukfeldt. 'Testing an integrated self-control and routine activities framework to examine malware infection victimization'. *Social Science Computer Review*, 2018.

[25] A. Hutchings and T. J. Holt. 'A crime script analysis of the online stolen data market'. *British Journal of Criminology* 55.3, 2015, pp. 596–614.

[26] E. F. ter Huurne. *Information Seeking in a risky world: the theoretical and empirical development of FRIS: A framework of risk information seeking*. 2008.

[27] H. Info. *Verschillende NAS-apparaten*. Accessed: 2023-04-25. 2022. URL: `https://hardware.info/`.

[28] M. Junger, L. Montoya, P. Hartel and M. Heydari. 'Towards the normalization of cybercrime victimization: A routine activities analysis of cybercrime in Europe'. *Proc. of the IEEE International Conference On Cyber Situational Awareness, Data Analytics And Assessment (CyberSA), London, UK*. IEEE. 2017, pp. 1–8.

[29] M. Junger, B. Veldkamp and L. Koning. *Fraudevictimisatie in Nederland (Fraud Victimization in the Netherlands)*. Retrieved December 13, 2022, from https://www.utwente.nl/nl/bms/fraudvic. 2022.

[30] M. Junger, V. Wang and M. Schlömer. 'Fraud against businesses both online and offline: Crime scripts, business characteristics, efforts, and benefits'. *Crime Science* 9.1, 2020, pp. 1–15.

[31] M. Keshavarzi and H. R. Ghaffary. 'I2CE3: A dedicated and separated attack chain for ransomware offenses as the most infamous cyber extortion'. *Computer Science Review* 36, 2020, p. 100233.

[32] L. Koning, M. Junger and B. P. Veldkamp. 'Reporting fraud victimization to the police: factors that affect why victims do not report'. *Psychology, Crime and Law*, in press.

[33] K. Lab. *New ransomware trends in 2022*. Retrieved August 26, 2022, from https://securelist.com/new-ransomware-trends-in-2022/106457. 2022.

[34] E. R. Leukfeldt. *Cybercriminal networks: Origin, growth and criminal capabilities*. Den Haag: Eleven International Publishers, 2016.

[35] S. Mansfield-Devine. *Sophos: The State of Ransomware 2022*. 2022.

[36] S. R. Matthijsse, M. S. van 't Hoff-de Goede and E. R. Leukfeldt. 'Your files have been encrypted: A crime script analysis of ransomware attacks'. *Trends in Organized Crime*, 2023, pp. 1–27.

[37] T. Meurs, E. Cartwright, A. Cartwright, M. Junger, R. Hoheisel, E. Tews and A. Abhishta. 'Ransomware Economics: A Two-Step Approach to Model Ransom Paid'. *2023 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2023, pp. 1–13.

[38] T. Meurs, R. Hoheisel, M. Junger, A. Abhishta and D. McCoy. 'What To Do Against Ransomware? Evaluating Law Enforcement Interventions'. *2024 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2024, pp. 1–13.

[39] T. Meurs, M. Junger, A. Abhishta and E. Tews. 'How attacker's effort, victim characteristics and context influence ransom requested, payment and financial loss'. *2022 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2022.

[40] T. Meurs, M. Junger, A. Abhishta and E. Tews. 'POSTER: How Attackers Determine the Ransom in Ransomware Attacks'. *IEEE S&P Conference*. 2022.

[41] T. Meurs, M. Junger, M. Cruyff and P. G. M. van der Heijden. 'Estimating The Number Of Unobserved Ransomware Attacks'. *Available at SSRN 4942706*, 2024.

[42] T. Meurs, M. Junger, E. Tews and A. Abhishta. 'NAS-ransomware: hoe ransomware-aanvallen tegen NAS-apparaten verschillen van reguliere ransomware-aanvallen'. *Tijdschrift voor veiligheid* 21.3-4, 2022, pp. 69–88.

[43] T. Meurs, E. Cartwright, A. Cartwright, M. Junger and A. Abhishta. 'Deception in double extortion ransomware attacks: An analysis of profitability and credibility'. *Computers & Security* 138, 2024, p. 103670.

[44] F. Miró-Llinares, J. Drew and M. Townsley. 'Understanding target suitability in cyberspace: An international comparison of cyber victimization processes'. *International Journal of Cyber Criminology* 14.1, 2020, pp. 139–155.

[45] R. Nieuws. *Unieke actie: politie bevrijdt gegijzelde computers dankzij truc met bitcoin*. Retrieved November 11, 2022, from https://www.rtlnieuws.nl/nieuws/nederland/artikel/5337913/unieke-actie-politie-probeert-gegijzelde-computers-met-slimme-truc. 2022.

[46] R. L. Plackett. 'Karl Pearson and the chi-squared test'. *International statistical review/revue internationale de statistique*, 1983, pp. 59–72.

[47] M. T. Puth, M. Neuhäuser and G. D. Ruxton. 'Effective use of Pearson's product–moment correlation coefficient'. *Animal Behaviour* 93, 2014, pp. 183–189.

[48] C. Putman and L. Nieuwenhuis. 'Business model of a botnet'. *Proc. of the 26th IEEE Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Valladolid, Spain*. IEEE. 2018, pp. 441–445.

[49] B. W. Reyns. 'Routine activity theory and cybercrime: A theoretical appraisal and literature review'. *Technocrime and criminological theory*. Routledge, 2017, pp. 35–54.

[50] E. Rodríguez, M. Fukkink, S. Parkin, M. van Eeten and C. Gañán. *Difficult for Thee, But Not for Me: Measuring the Difficulty and User Experience of Remediating Persistent IoT Malware*. arXiv preprint arXiv:2203.01683. 2022.

[51] E. Rodríguez, S. Verstegen, A. Noroozian, D. Inoue, T. Kasama, M. van Eeten and C. H. Gañán. 'User compliance and remediation success after IoT malware notifications'. *Journal of Cybersecurity* 7.1, 2021, tyab015.

[52] R. Schouw. *Qlocker ransomware maakt wereldwijde comeback op QNAP NAS-apparaten*. Retrieved November 12, 2022, from https://nl.hardware.info/nieuws/80324/qlocker-ransomware-maakt-wereldwijde-comeback-op-qnap-nas-apparaten. 2022.

[53] Shodan. *Shodan*. Accessed: 2022-10-25. 2022. URL: https://shodan.io/.

[54] J. Sowells. *eCh0raix Ransomware Targeting QNAP Devices*. Retrieved August 26, 2022, from https://hackercombat.com/ech0raix-ransomware-targeting-qnap-devices. 2022.

[55] R. Van Wegberg and T. Verburgh. 'Lost in the Dream? Measuring the effects of Operation Bayonet on vendors migrating to Dream Market'. *WebSci'18: Evolution of the Darknet*. Association for Computing Machinery (ACM). 2018, pp. 1–5.

[56] S. van de Weijer, R. Leukfeldt and S. van der Zee. 'Reporting cybercrime victimization: determinants, motives, and previous experiences'. *Policing: An International Journal*, 2020.

[57] A. Witt, J. Kurths and A. Pikovsky. 'Testing stationarity in time series'. *physical Review E* 58.2, 1998, p. 1800.

[58] E. Witteman. *QNAP patcht kritieke lekken die ransomware faciliteerden*. Retrieved August 26, 2022, from https://www.techzine.nl/nieuws/privacy-compliance/457599/qnap-patcht-kritieke-lekken-die-ransomware-faciliteerden. 2021.

[59] R. Wortley and L. Mazerolle, eds. *Environmental criminology and crime analysis*. London, UK: Willan, 2008.

[60] M. Yar. 'The Novelty of 'Cybercrime'. An Assessment in Light of Routine Activity Theory'. *European Journal of Criminology* 2.4, 2005, pp. 407–427.

[61] D. W. Zimmerman. 'Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances'. *The Journal of Experimental Education* 55.3, 1987, pp. 171–174.

Part III

Information Asymmetry

This page is intentionally left blank.

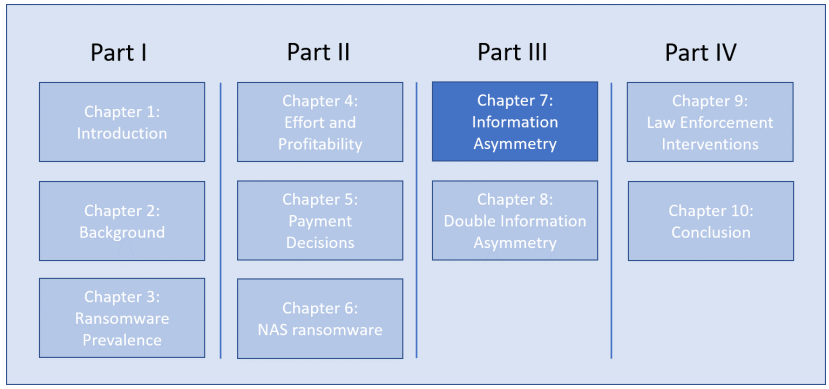
Доверяй, но проверяй

Trust, but verify

~ Russian proverb

Chapter 7

Deception in Double-Extortion Ransomware



Ransomware attacks have evolved with criminals using double-extortion schemes, where they signal data exfiltration to inflate ransom demands. This development is further complicated by information asymmetry, where victims are compelled to respond to ambiguous and often deceptive signals from attackers. This study explores the complex interactions between criminals and victims during ransomware attacks, especially focusing on how data exfiltration is communicated. We use a signaling game to understand the strategies both parties use when dealing with uncertain information. We identify five distinct equilibria, each characterized by the criminals' varied approaches to signaling data exfiltration, influenced by the strategic parameters inherent in each attack scenario.

7.1 Introduction

Crypto-ransomware attacks globally are a growing concern for our society. In the United States alone, an estimated 1,981 schools, 290 hospitals, 105 local governments and 44 universities and colleges were hit by crypto-ransomware attacks in 2022[44]. Crypto-ransomware (or ransomware) is a malicious software that aims to encrypt the files of victims [23]. Typically, if victims lack adequate backups, they can only regain access to those files after paying a ransom to the criminals [30]. Given that many victims are willing to pay a ransom, these malicious software have proved highly lucrative for criminals and provide a viable 'business model'.

In recent years, criminals have complemented file encryption with the stealing of sensitive data. This sensitive data might contain personally identifiable information of employees and/or customers, intellectual property or legal information. The criminals then threaten to publish the sensitive data or sell it to competitors if the victim does not pay the ransom. This is referred to as a double-extortion scheme [49, 46, 28]. Evidence suggests the double-extortion scheme leads to larger ransom requests and/or a higher willingness to pay, and therefore to more profits for the criminals compared to encryption-only-attacks [33, 39]. Hence, it is important to investigate *how double-extortion schemes may evolve, and how law enforcement can disrupt the business model of these criminals*.

During a ransomware attack, victims struggle with the critical issue of determining whether their data has been exfiltrated [38, 42]. While some may possess logs that facilitate the identification of accessed files, others are not as fortunate. Although irregularities in data flow on the affected network during an attack can be identified, it does not always confirm the theft of confidential information. This ambiguity creates a window of opportunity for criminals. Those engaged in encryption-only attacks can exploit this uncertainty, asserting that data exfiltration occurred to demand a higher ransom. Criminals may proactively offer, or victims may request, 'evidence' of data exfiltration. This exchange, termed as a 'signal' in our study, forms the core of the strategic dynamics between the attacker and the victim. It is a 'game', where the criminal's choice to provide evidence of exfiltration intersects with the victim's decision on whether or not to pay the ransom demand.

In 2020, Coveware reported that 70% of ransomware attacks were combined with data exfiltration [49]. [39] found that between 2019 and 2022, from 124 ransomware attacks, forensic analysis suggested there were traces of data exfiltration in 43% of the attacks. This resulted in a slightly larger portion of payments: 26% of the victims paid with likely data exfiltration, whereas 24% paid

without data exfiltration.The ambiguity around data exfiltration complicates the victims' response, with criminals sometimes falsely claiming data theft to inflate ransoms [39]. The Maze group initiated the double-extortion scheme in November 2019, exposing non-paying victims on a leak site and claiming data deletion for those who paid, though without conclusive proof [24, 28, 19, 18].

Criminals can use various strategies to signal that data was exfiltrated [38, 25]. One approach is to publish a small fraction of the exfiltrated data on a leak site. However, it is worth noting that this strategy carries a potential drawback, as the extent of the reputational harm incurred might be independent of the magnitude of the published data. Moreover, publishing some data still leaves open the question of how much additional data was exfiltrated. Another approach is to send a picture of the file tree to the victim. A potential drawback of this approach is that it gives the possibility to victims to determine the importance of the stolen files. It is also relatively easy to obtain without actually exfiltrating the files and so is not a particularly credible signal. Criminals may, therefore, decide not to signal even if data was exfiltrated. It could be they want to sell the data on darknet forums [33] or to conceal attacks where data exfiltration was unsuccessful. In such cases the past reputation of the ransomware group may inform on the likelihood of data exfiltration.

In this chapter, we employ a game-theoretic framework to evaluate the dynamics between criminals and victims in the context of ransomware attacks, focusing on the signaling of data exfiltration [29]. We are particularly interested in the criminals' decision to signal data theft and how victims respond to such signals. Signaling games, a well-explored concept in game theory, offer valuable insights into these complex interactions characterized by information asymmetry. Our analysis unveils five distinct equilibria, shaped by the criminals' varied approaches to signaling data exfiltration—from consistent signaling, no signaling at all, to conditional signaling based on actual data theft. These equilibria are not arbitrary but are influenced by the strategic parameters inherent in each ransomware attack scenario. We further calibrate the game with parameters that mirror real-world conditions. This calibration facilitates the identification of the most realistic equilibrium, enabling us to anticipate the likely ransom amounts and the corresponding payoffs for both parties involved. Based on our findings, we propose tangible strategies to dismantle the ransomware business model. These strategies are aimed at reducing the ransom amounts and undermining the criminals' payoffs, marking a significant step towards mitigating the impacts of these cyber-attacks.

Our article makes three contributions; First, it is, to the best of our knowledge, the first study to analyse the important signaling component of double-extortion

ransomware schemes. We draw on data from negotiations between criminals and victims to motivate this issue as being an important area of study. Second, we provide a theoretical analysis of the strategic consideration criminals face when signaling data exfiltration and the consequences for the payoffs of criminals and victims. Third, by understanding the incentives of criminals we can identify the optimal strategy of victims and examine defensive measures for victims and policy makers to decrease the negative welfare consequences of double-extortion ransomware.

The paper is organized as follows. In Section 2 we motivate our signaling game approach through empirical observations of double-extortion ransomware attacks and negotiations. In Section 3 we briefly overview the previous literature on the economic and game-theoretic modeling of ransomware. In Section 4 we introduce the signaling game. In Section 5 we state the main results. In Section 6 we conclude and provide policy recommendations. Proofs of propositions are provided in Section 7.

7.2 Motivation

The foundation of our game-theoretic model is based upon empirical observations that will be expanded upon in the following section, providing context and motivation for our theoretical model. In Section 5.2 we will calibrate the parameters in our model using this dataset. For a more detailed analysis we refer to our previous work [39].

We draw on empirical data from two datasets compiled by the lead author: 1) 525 ransomware attacks reported to the Dutch Police and 2) 117 ransomware attacks reported to an incident response company (IR company). Some general insights from the Dutch Police data have previously been reported in [39]. In that paper, the authors study how the criminal's effort, victim characteristics and context influence the ransom requested, payment and financial loss. A key finding of the study is that data exfiltration has a highly significant, positive impact on the ransom requested, proportion of victims who pay, and the victim's financial loss. This demonstrates the critical role that exfiltration plays in ransomware.

The foundation of our game-theoretic model is established upon empirical observations that will be expanded upon in the ensuing section, providing necessary context. For an in-depth analysis of the dataset, we point readers to our prior work [39].

In motivating the game theoretic approach used in this chapter we analysed the extended datasets introduced above. From the overall datasets we excluded attempted attacks, no encryption and attacks on individuals. This resulted in

1) 354 ransomware attacks reported to the Dutch Police and 2) 98 ransomware attacks reported to the IR company. In total we, therefore, analysed 452 ransomware attacks. For each attack a range of variables were coded based on the case logs provided by the Police and IR company. For this study we use the following variables: whether data is exfiltrated (yes/no/unknown), what the ransom requested was before and after negotiations (in euro) and whether the victim paid (yes/no/unknown). Furthermore, we looked at the negotiation text to understand the exchange of information of data exfiltration between the victim and criminals. We remark that the classification of data exfiltration (yes/no/unknown) is somewhat subjective for the very reasons that motivate this paper (namely, data exfiltration is hard to verify). Our classification benefits, however, from information that became available over time, and may not have been available at the time of the attack, e.g. whether data was subsequently published on a leak site.

7.2.1 Criminal Profits of Double-Extortion Ransomware

Here we state our main findings from the data analysis in terms of data exfiltration in relationship with payment, and ransom requested.

1. **Data exfiltration:** Overall, we find that in 50.4% of cases it was unknown whether data was exfiltrated, and in 49.6% of cases we believe that data was exfiltrated. Data exfiltration is assumed in 43% of cases in the Dutch Police, based on 134 cases, and in 53% of cases of the IR company, based on 98 cases. Commonly, the basis for assuming that data was exfiltrated is because: (1) log files show specifically that files have been exfiltrated, or (2) data was published on the leak site of the criminals. We see, therefore, that data exfiltration is common but not universal. We also see that whether data exfiltration took place remains unknown in many cases, even with the benefit of hindsight (e.g. data appearing on leak sites).

2. **Paid:** From the 452 ransomware attacks, 130 victims negotiated. In total, 119 victims paid the ransom (27.8%). Of these, 78.5% victims paid after negotiations and 21.5% paid without negotiations. If we focus on those subset of payments where we are relatively confident if data exfiltration took place, we find that data exfiltration leads more often to payment: 37.5% versus 28.9%. This difference is statistically significant, based on a chi-squared test ($\chi^2 = 5.42, df = 1, p = 0.02$). The reason why both payment percentages are above the total average of 27.8% is because, relatively, it is more often unknown whether data was exfiltrated for the vic-

tims who did not pay. So, these results show that data exfiltration leads to larger proportion of victims paying than no data exfiltration.

3. **Ransom requested:** The average ransom request before negotiation is 1,029,320 euro (sd=3.0 million euro). After negotiation the average ransom request is 578,956 euro (sd=1.9 million euro), a decrease of 44%. When data is exfiltrated, the ransom before negotiation is 2,960,281 euro (sd=4.7 million euro) and after negotiation 1,771,216 euro (sd=3.2 million euro), a decrease of 40%. Without data exfiltration the ransom before negotiation is 466,924 euro (sd=2.0 million euro) and after negotiation 135,346 euro (sd=0.2 million euro), a decrease of 70%. Tests that there is a difference in ransom requested with and without data exfiltration using a t-test is significant for both ransom requested before negotiations ($t = 63.17, df = 232, p < 0.001$) and after negotiations ($t = 66.05, df = 232, p < 0.001$)*. It appears that data exfiltration is highly profitable for the criminals. Furthermore, it seems that data exfiltration leads to less discount after negotiations than when data is not exfiltrated.

In conclusion, double-extortion ransomware seems to lead to a larger proportion of victims paying the ransom and a larger ransom requested, and, therefore, to more profits for criminals.

7.2.2 Exploration of Victim's Decision To Pay

If victims are more likely to pay a ransom, and pay a larger ransom, because of data exfiltration, it is naturally in their interests to ascertain whether data exfiltration has indeed taken place. As we discussed in the introduction this is difficult to do in the immediate aftermath of an attack. Hence criminals may want to signal data exfiltration, and victims may seek for information about data exfiltration. In Table 7.1 we provide six illustrative examples of criminals attempting to signal that data is exfiltrated. As you can see, data exfiltration was claimed in the ransom note. In two cases supplementary evidence was provided during negotiations. We also summarise the victims' decision-making process regarding ransom payment. Note that we display the anonymized text used by criminals, which includes grammar and style mistakes.

In the first four cases the victims were not convinced by the criminal's claim that data has been exfiltrated. The signal in this case was, therefore, seen as non-credible. Furthermore, in none of the four cases was data published on the

*In line with [39] we have taken the logarithm of the ransom to approximately normalize the data, which is required to validly perform a t-test. Not taking the logarithm also results in highly significant t statistics.

leak site after the victim did not pay. This may suggest the victims were probably correct to infer no data had been exfiltrated. The absence of data being published on a leak site does not, however, serve as conclusive evidence that no data has been compromised. In informal discussions, law enforcement officers have disclosed to the authors that criminals are occasionally selective in their choice of which victim's data they publish. By exclusively publishing data of large organizations, criminals can cultivate a reputation as a group focusing on prominent victims.

In the fifth and sixth cases the criminal showed a list of files which were exfiltrated. In the fifth case this led the victim to believe that data was exfiltrated and they made the decision to pay the ransom to prevent the publishing of the

Case	Criminal claim (anonymized raw text)	Additional signals	Victim decision-making
1.	"We gathered highly confidential/personal data. These data are currently stored on a private server. This server will be immediately destroyed after your payment. If you decide to not pay, we will release your data to public or re-seller. So you can expect your data to be publicly available in the near future. We only seek money and our goal is not to damage your reputation or prevent your business from running"	No	Although the victim did not believe data was exfiltrated, they did decide to pay because backups were inadequate to recover without payment.
2.	"For the ransom you get: Full decryption Fixing your network vulnerabilities and securing your network Removal of all your data from our servers."	No	Victim was not confident data was exfiltrated and did not pay.
3.	"If we don't hear back from you within 24 hours. I can sell them on the darknet and send the information to regulatory agencies, in your area I will send out offers to competitors to buy your data. In this case you will have the following problems : 1. Your customers will become victims of fraudsters (who will buy your data on the darknet). 2. Regulatory authorities (responsible for enforcing data protection laws) will start investigating your company for leaking your customers' personal data (leading to huge fines and loss of reputation). 3. Your competitors could easily get hold of your information."	No	Victim was not confident data was exfiltrated. However, due to the lack of adequate backups they did pay. During the negotiations no other claim of data exfiltration was made by the criminal
4.	"All your important files have been encrypted. Any attempts to restore your files with third-party software will be fatal for your files! Restore your data possible only buying private key from us. We have also downloaded a lot of private data from your network. If you do not contact us in a 5 days, we will post information about your breach on our public news webs."	No	Victim was not confident data was exfiltrated and did not pay.
5.	"Your data is stolen and encrypted. If you don't pay the ransom, the data will be published on our TOR darknet sites. Keep in mind that once your data appears on our leak site, it could be bought by your competitors at any second, so don't hesitate for a long time. The sooner you pay the ransom, the sooner your company will be safe."	List of .rar files of alleged exfiltrated data provided by criminal	Victim paid, because data of customers was stolen. They had backups, but decided to pay just for prevent the publication of exfiltrated data.
6.	"Price for you is X btc. You need to pay this amount and we will give you decrypt tool for all your machines, security report on how you were hacked, file tree on what we have downloaded a lot of data from your network that in case of not payment will be published on public news website and sold on the black-markets. We remove it after payment and wiping log is provided as well. To start a business we offer you to make payment in two stages. What amount you can pay today?"	A list of exfiltrated files was provided by criminals	Victim got the file tree of the exfiltrated data and decided that the data was not important. Furthermore, they did have backups. Therefore, they decided not to pay.

Table 7.1: Claims of the criminal of data exfiltration (raw text and anonymized), additional signals send by the criminal and the victim's decision-making whether to pay or not.

data on a leak site. The criminals, thus, benefited from sending a more credible signal. In the sixth case the victim decided, based on the list of exfiltrated files provided by the criminal, that data publication would be less costly than paying the ransom. As a consequence, the data of the victim was published on the leak page. In this example sending a signal appears to have backfired for the criminals, because it gave the victim the opportunity to estimate the reputational damage of data exfiltration.

Considering our dataset as a whole, we have examples of all possible combinations: (a) The criminals signaling data exfiltration when we believe there was data exfiltration, (b) not signaling data exfiltration when we believe there was data exfiltration, (c) signaling data exfiltration when we believe there was no data exfiltration, and (d) not signaling data exfiltration when we believe there was no data exfiltration. This makes it difficult for victims, law enforcement and policy makers to understand the optimal response when claims of data exfiltration are made. Given the large ransom amounts at stake it is of value to better understand the trade-offs that victims face.

It is reasonable to hypothesize that criminals engage in strategic considerations when deciding to signal data exfiltration or refrain from doing so, taking into account the potential impact on victims' willingness to pay. Indeed, the history of ransomware shows that criminals rapidly evolve their economic strategy to ones that make more money. Consequently, we develop a decision model to capture this behavior, using a game-theoretic framework of signaling. In the subsequent section, we provide a rationale for utilizing game-theoretic models in the context of ransomware and overview prior research on this subject.

7.3 Related Works

The goal of this section is to give a brief overview of past research on the economic and game-theoretic approach to ransomware attacks. Traditionally, ransomware research takes a more technical approach [9, 47]. However, recently the application of economic theory to analyse decision-making of criminals and victims of ransomware attacks have increased [11, 33, 30, 21]. This might be the result of ransomware criminals running there attacks as a business, where many decisions are made using economic reasoning [27].

Most ransomware criminals are financially motivated and conduct multiple attacks [39, 13]. Therefore it is important for them to optimize profits over multiple ransomware attacks. One important aspect is the use of different price discrimination strategies [25]. For example, the criminals change the ransom requested on victim characteristics, like yearly revenue [39]. Another aspect is

the use of data exfiltration: as concluded in Section 2, this increases the willingness to pay of victims, which leads to more profits for criminals. [13] identify four distinct fears of victims which might explain the increased willingness to pay: (1) incrimination (e.g. exposure to data protection authorities), (2) reputational damage/lost revenue (e.g. exposure of sensitive data which could cause loss of customers), (3) exposure of intellectual property, and (4) humiliation (e.g. exposing embarrassing information about customers or a particular employee in an executive role). These fears increase the willingness to pay and give an incentive for criminals to perform data exfiltration, or pretend that data is exfiltrated.

In addition to the previously mentioned empirical studies, game-theoretic models have been employed to explore the dynamics between criminals and victims within the context of ransomware attacks [11, 30, 21] and double-extortion ransomware schemes [34, 33, 35]. Game theory provides a valuable theoretical framework for examining the strategic decision-making process of different actors, making it highly applicable in the context of ransomware attacks [11]. This suitability arises from the well-defined roles of the actors involved, namely the criminal and victim, and clear decision options available to the victim, such as paying or not paying the ransom. Furthermore, the payoffs are mostly monetary and therefore easily quantified. From the game-theoretic framework we could infer whether there is a stable equilibrium and possible interventions to change that equilibrium to increase social welfare.

Several studies have applied a game-theoretic framework to double-extortion ransomware [33, 31, 35]. [33] demonstrate that when criminals employ a strategy involving both data encryption and data exfiltration, it consistently results in higher profits as opposed to solely relying on data encryption. Furthermore, the act of selling the exfiltrated data has been found to further increase the profitability for criminals, surpassing the potential reputation gains achieved by simply deleting the data upon receiving payment from victims.

One possible critique of using game-theoretic models in the ransomware context is the assumption of rational decision-making by both criminals and victims. Both criminals and victims may make impulsive, irrational decisions [11]. Rationality, in this context, however, does not imply a cold and unemotional decision-making process, but rather an understanding that criminal and victim need to take account of each other's strategic incentives, and have incentives to maximize their financial payoff. This aligns with the Rational Choice Model proposed by [15]. The Rational Choice Model (RCM) of crime states that criminals, or offenders, are rational decision-makers. Crime is purpose behaviour designed to meet the offender's commonplace needs for such things as money, status, sex

and excitement. Offenders are reasoning actors who weigh means and ends, costs and benefits, and make a rough rational choice for the course of action that seems to yield the most benefit [15, 16].

Research supports the Rational Choice Model of crime, for offline crime [51, 12] and online crime [3, 52]. Most relevant, experiments show that policy measures that influence the costs and benefits of crime, by increasing the effort and the risks, and decreasing the potential benefits, generally prevent crime offline [12] and online [7]. Taken together, we could conclude that the assumption of rationality, which is crucial for the application of a game-theoretic framework, can yield valuable insight in the context of double-extortion ransomware schemes.

So far, we considered studies which focus on the profitability of data exfiltration, applying game-theoretical models to ransomware and data exfiltration. These papers abstract away from a key aspect of the strategic environment: victims are often unsure whether data is exfiltrated. Information asymmetry between victim and criminal can be modeled with signaling games [43]. Signaling games are a widely used framework in economics and evolutionary biology to model a range of applied settings. They have been used, for instance, to model job seekers signaling their productivity to potential employers [48]. In this setting the signal could be years of schooling or high grades (even if that does not directly add to productivity). Signaling games have also been used to understand non-anonymous donations to charity [22]. In this case the donation can be a signal the donor is a pro-social, generous individual. In all these settings there is one party that has more information than the other (whether they are productive, pro-social etc.) and have incentives to signal a 'desirable' property. In our setting the criminal may want to signal data exfiltration.

The analysis of signaling games has produced some profound results. For instance, it has been shown that 'costly signaling' can result in which the actor with the most desirable attributes must incur large costs to signal their desirability. There is, for instance, evidence that university education is a costly signal of ability [6]. In our setting, this suggests that data exfiltration need not be unambiguously beneficial for the criminals. Another seminal finding is due to [2]. He set up a framework to analyse the information asymmetry between the buyer and seller of used cars. The seller knows the quality of the car, but the buyer does not. [2] shows that this information asymmetry can lead to a breakdown in trade. In short, sellers of good cars do not want to sell them cheap, but buyers are reluctant to pay a high price for a car that may be no good. We observe, therefore, an adverse selection market failure. The framework of [2] has also been used to study other use cases [31]. In our setting, it again suggests that data exfiltration need not be unambiguously beneficial for the criminals. In the

following section we will formally apply the theory of signaling games to the case of double-extortion ransomware attacks.

7.4 Model

7.4.1 Signaling Game

In this section we introduce a signaling game of double-extortion ransomware. The game involves two players, a criminal and a victim. It has three stages, which can be explained as follows. The variables are summarized in Table 7.2 and the signaling game is depicted in Figure 7.1.

Stage 1: We assume that the criminal attempts to exfiltrate data. However, this attempt may not succeed. We denote with α the probability the criminal exfiltrates data (DE) and denote with probability $1-\alpha$ the probability the criminal does not succeed in exfiltrating data (NDE). In game-theoretic terminology, this process is formulated as 'nature' determining with a probability α the state is DE and $1-\alpha$ the state is NDE [2, 43, 48]. The criminal learns in Stage 1 whether they are in state DE or NDE. The victim remains uninformed, although the probability of data exfiltration α is common knowledge.

Stage 2: The criminal chooses a ransom demand and whether to send a signal to the victim that data was exfiltrated (S) or not (NS). The signal could consists of sending a file tree or pictures of the file tree structure. Another possibility is sending a few exfiltrated files. Let R^S denote the ransom demand of the criminal if they send a signal and R^{NS} the demand if no signal is sent. Thus, the criminal either sends signal S and ransom demand R^S or chooses NS and ransom demand R^{NS}.[†]

The cost of sending a signal is k^D when data is exfiltrated and k^N when no data is exfiltrated. One interpretation of the cost of a signal is opportunity costs. For instance, the effort and time could have been used for another attack. Crucially, we assume that sending a signal when data is exfiltrated is less costly than when data is not exfiltrated, so $k^D < k^N$. This assumption arises from the notion that it might be harder to send a signal in state NDE than DE. Indeed, it

[†]The criminal could choose any ransom above 0 for any combination of both own type and signal. So, suppose, more generally, we denote by $R^S_{DE}, R^S_{NDE}, R^{NS}_{DE}$ and R^{NS}_{NDE} the ransom of a type DE or NDE if they signal or do not signal. There cannot be an equilibrium in which a criminal of type DE and NDE signal and $R^S_{NDE} \neq R^S_{DE}$; this would reveal the criminal if type NDE and, thus, make their signal ineffective. Similarly, there cannot be an equilibrium in which a criminal of type DE and NDE would not signal and $R^{NS}_{NDE} \neq R^{NS}_{DE}$; this would again reveal the criminal if type NDE and lower the ransom the victim would rationally pay.

could be that k^N is very large meaning that it is essentially impossible to send a signal if data is not exfiltrated.

Stage 3: Having seen whether the criminal sends a signal (S) or no signal (NS) and seen the ransom demand R^S or R^{NS}, the victim decides to pay or not. We assume that this is a binary yes/no decision with no possibility for negotiation.[‡] Final payoffs are now determined as depicted in Figure 7.1, which shows criminal payoff, victim payoff for each potential outcome.

In explaining the respective payoffs of criminal and victim we remark that game theoretic equilibria depend on the relative payoff differences across actions, rather than the absolute payoff. For the victim we are, thus, interested in the relative payoff difference from paying the ransom versus not paying the ransom. We assume that if the victim pays the ransom then they lose the ransom amount, R^S or R^{NS}, as well as 'legal fees', $L \geq 0$, which can include legal and other associated costs (including psychological and moral) of paying the ransom. We assume that if the ransom is paid the criminal returns access to, at least some, files and is reduces the amount of sensitive data published. If, therefore, the victim does not pay the ransom they lose $V \geq 0$ from higher recovery costs as well as $T \geq 0$ (if data was exfiltrated) from increased reputational costs resulting from publication of sensitive data. This motivates the payoff function in Figure 7.1.

In analysing the incentives of the criminal we need to consider the relative payoff differences from signalling or not not signalling. We assume that the attack costs the criminal c. If the victim pays then the criminal receives ransom R^S or R^{NS}. If the criminal signals then they pay the cost k^D or k^N, as stated previously.

The victim will have legal fees, recovery costs (like buying new hardware and software) and reputation costs under any scenario. We reiterate, however, that since we consider these costs to be constant across all outcomes, we do not include them in our analysis. Furthermore, we would like to stress that our model implicitly takes into account factors such as the importance of backups. For instance, if a victim has good backups then the recovery costs V would be low, and therefore the victim, as we will show, would not pay a large ransom. Similarly, if sensitive data is not exfiltrated then T would be low and the victim would again not pay a large ransom.

A (pure) strategy for the criminal involves the following conditional choices: (a) decide whether to signal or not if data is exfiltrated, (b) decide whether to

[‡]Alternatively, R^S or R^{NS} could be seen as the final ransom demands that will result from negotiation.

signal or not if data is not exfiltrated, and (c) determine ransom demands R^S or R^{NS} as appropriate. A (pure) strategy for the victim compromises conditional choices: (a) decide to pay or not to pay if the criminal signals, and (b) decide to pay or not to pay if the criminal does not signal.

	Variable	Description
Criminal	R^S	Ransom when signaling
	R^{NS}	Ransom when not signaling
	c	Cost of attack
	k^D	Cost of signal with data exfiltration
	k^N	Cost of signal no data exfiltration
	τ	The state or type of the criminal: data exfiltrated or not
Victim	T	Reputation cost
	V	Recovery cost without decryption key
	L	Legal fees of paying ransom
	α	Probability of data exfiltration
	μ	Probability the victim believes data is exfiltrated
	ϵ	Smallest gain that would induce victim to pay

Table 7.2: Variables used in the data exfiltration signaling game

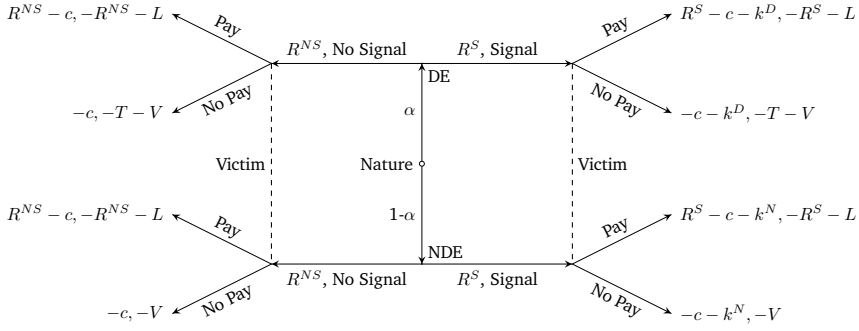


Figure 7.1: A schematic representation of set-up of the signaling game of data exfiltration

7.4.2 Bayesian equilibria of the signaling game

In the following we identify (pure strategy) Bayesian equilibria of the signalling game. A Bayesian equilibria takes into account that the victim starts with prior belief α that data was exfiltrated but can potentially update their beliefs once they observe the strategy of the criminal. We denote by μ the updated belief of the victim the criminal is type DE. A Bayesian equilibrium has the following basic properties: (a) The criminal maximizes their expected payoff given the strategy of the victim, (b) the victim updates their beliefs about state DE and NDE using Bayes rule, and (c) the victim maximizes their expected payoff given the strategy of the criminal and their own beliefs [20].

Where relevant it may be necessary to tie down beliefs for 'surprise events' or outcomes that are 'off the equilibrium path'. For instance, if the candidate equilibrium says that the criminal will signal, we need to specify the victim's beliefs if the criminal is observed to not signal. In this case we invoke the D1 Criterion which says that any deviation from the equilibrium path is assumed to be done by the type with the most incentive to deviate [5]. In this case (see the formal analysis for more details) it means the choice to not signal is seen as evidence that there was no data exfiltration.

As is standard in the analysis of signalling games, we distinguish between separating and pooling equilibria. A separating equilibrium has the property that the victim can distinguish the type of the criminal (DE or NDE) from their actions. A pooling equilibrium has the property that the criminal will act the same whether type DE or NDE and so the victim can not distinguish type. We identified two separating equilibria (we will call A1 and A2) and three pooled equilibria (we will call B1, C1 and C2). We first characterize the five equilibria. We then provide conditions on the parameters of the game under which the different equilibria exist. We would argue that all five types of equilibria can potentially be seen in the field.

A1. Separating equilibrium. Victim pays whether signal or not. The victim believes $\mu = 0$ when she receives no signal and $\mu = 1$ when they receive a signal. The victim is, thus, willing to pay the ransom when she gets the signal that data is exfiltrated if and only if $R^S < T + V - L$. Set $R^S = T + V - L - \epsilon$, where ϵ is arbitrarily close to 0. Likewise, the victim is willing to pay the ransom when there is no signal if and only if $R^{NS} < V - L$. Set $R^{NS} = V - L > 0$. Combined, the criminal has payoff $U = T + V - L - k_D - c - \epsilon$ in state DE and $U = V - L - c - \epsilon$ in state NDE. The criminal has no incentive to deviate from the equilibrium strategy in state DE if $k^D < T$. In interpretation, the extra revenue the

criminal can demand from signalling data is exfiltrated compensates for the cost of sending the signal. Similarly, the criminal has no incentive to deviate from the equilibrium strategy in state NDE when $k^N > T$. In interpretation, the cost of sending a signal (when data is not exfiltrated) is higher than the extra revenue from the ransom.

A2. Separating equilibrium. Victim only pays when receiving signal. As with equilibrium A1, the victim believes $\mu = 0$ when they receive no signal and $\mu = 1$ when she receives a signal. Following, the same logic as equilibrium A1 the victim is willing to pay ransom $R^S = T + V - L - \epsilon$ if their is a signal. The maximum ransom they are willing to pay if there is no signal is $R^{NS} < V - L$. If, therefore, $V < L$ the victim is not willing to pay a (positive) ransom. The payoff for the criminal in state DE is $U = T + V - L - c - k^N - \epsilon$ and their payoff in state NDE is $U = -c$. The criminal in state NDE has no incentive to signal if $T + V - L < K^N$. In interpretation the cost of signaling in state NDE is higher than the maximum ransom the victim is willing to pay.

B1. Pooled equilibrium: The criminal signal and the victim pays. The criminal sends a signal in both states DE and NDE. The victim should maintain the belief $\mu = \alpha$ when they receive a signal that data is exfiltrated. If they do not receive a signal than beliefs are set $\mu = 0$ (invoking the D1 Criterion). The maximum ransom a victim is willing to pay if a signal is sent is $R^S = V + \alpha T - L - \epsilon$. The maximum ransom they are willing to pay if no ransom is sent is $R^{NS} = V - L - \epsilon$. Thus, the criminal has no incentive to deviate from the equilibrium path in state NDE if $\alpha T > k^D$. In interpretation, the cost to the NDE type of signaling is sufficiently low that they signal even though no data was exfiltrated. This lowers the ransom a type DE can demand because their signal is less credible.

C1. Pooled equilibrium: The criminal does not send a signal and the victim pays. The criminal sends no signal in both state DE and NDE. The victim should maintain the belief $\mu = \alpha$ when they receive no signal that data is exfiltrated. If they do receive a signal than beliefs are set $\mu = 1$ (invoking the D1 Criterion). The victim is willing to pay ransom $R^{NS} = V + \alpha T - L - \epsilon$ when she does not receive a signal and $R^S = V + T - L - \epsilon$ when she does receive a signal. The criminal has no incentive to deviate when type DE if $(1 - \alpha)T < k^D$. In interpretation, the extra ransom is insufficient to cover the cost of sending a signal (even when data is exfiltrated). This

equilibrium also requires $V + \alpha T > L$ so that the victim is willing to pay a positive ransom.

C2. Pooled equilibrium: No signal and victim does not pay. We follow the same logic as equilibrium C1 but now consider the case where $V+\alpha T < L$. In this case the victim is not willing to pay a positive ransom if a signal is not sent. Moreover, the criminal has no incentive to deviate when type DE if $V + T - L < K^D$. The interpretation of this equilibrium is that it is too costly to pay for the victim and too costly for the criminal to send a credible signal. Clearly there would be no incentive for the criminal to attack in this scenario because they incur the cost c.

The Bayesian equilibria that exist in the game will depend on the specific parameters of the game, V, L, T, α, K^D and K^N. In the following three Propositions we characterise the set of conditions under which there exists separating equilibria A1 and A2 (**Proposition 1**), pooling equilibria B1 (**Proposition 2**), and pooling equilibria C1 and C2 (**Proposition 3**). Proof of propositions can be found in the Appendix.

Proposition 1 *If $V > L$ and $k^D < T < k^N$ there is a Bayesian equilibrium satisfying the D1 Criterion of the type A1. If $V < L$ and $k^D < T + V - L < k^N$ there is a Bayesian equilibrium satisfying the D1 Criterion of the type A2.*

Our first proposition shows that there exists a separating equilibrium if the cost of signalling is sufficiently low when the criminal is type DE and high when they are type NDE. Thus, the criminal only signals if data has been exfiltrated. The criteria for sufficiently low and high depends on the reputational costs T, recovery costs V and legal fees L. The victim pays if data is exfiltrated and pays if data is exfiltrated if and only if $V > L$.

Proposition 2 *If (a) $V > L$ and $\alpha T > k^N$, or (b) $V + \alpha T > L > V$ and $\alpha T + V - L > k^N$ there is a signaling equilibrium satisfying the D1 Criterion of the type B1.*

Our second proposition shows conditions under which there exists a pooling equilibrium where the criminal signals data is exfiltrated, irrespective of whether data is exfiltrated or not. This equilibrium exists if it is sufficiently low cost for the criminal to signal data exfiltration. The notion of sufficiently low depends on the ex-ante probability of data exfiltration α and the reputation cost T. The higher is αT then the more likely to obtain a pooling equilibrium with signalling.

Case	Type equilibrium	Condition
A1	Separating - victims pays	$L < V$ & $k^D < T < k^N$
A2	Separating - Only pay when signal	$V < L < V + T$ & $k^D < T + V – L < k^N$
B1	Pooling - Signal and pay	$V + \alpha T > L$ & $\alpha T > k^N$
C1	Pooling - No signal and pay	$V + \alpha T > L$ & $(1–\alpha)T < k^D$
C2	Pooling - No signal and no pay	$V + \alpha T > L$ & $V + T – L < k^D$

Table 7.3: Stable equilibria and conditions in signaling game

In interpretation, the victim is willing to pay a larger ransom if data exfiltration is signaled and so it is in the interests of the criminal to signal data exfiltration when type NDE (if k^D is sufficiently low).

Proposition 3 *If $V + \alpha T > L$ and $(1 - \alpha)T < k^D$ there exists a signaling equilibrium satisfying the D1 Criterion of the type C1. If $V + \alpha T > L > V$ and $T + V - L < k^D$ there exists a signaling equilibrium satisfying the D1 Criterion of the type C2*

Our final proposition shows conditions under which there exists a pooling equilibrium where the criminal does not signal data is exfiltrated, even if it is. This type of equilibrium exists if the cost of signaling is sufficiently high for type DE. Again, the reputation costs T are an important determinant of the meaning of sufficiently high. If the reputation costs are low then we are more likely to obtain a pooling equilibrium with no signalling. In interpretation, the victim is not willing to pay a larger ransom if data is exfiltrated and so there is less incentive for the criminal to signal exfiltration (if k^D is sufficiently high).

The five type of equilibria we have identified and conditions under which they exist are summarized in Table 7.3.

7.5 Theoretical Insights from the Game

7.5.1 Expected Payoffs

In the previous section we derived five types of Bayesian equilibria of the signaling game. In this section we perform simulations to better understand the interaction between different parameter values and the resultant payoffs of the criminal and victim. A summary of the equilibrium ransom amount and corresponding payoff of criminal and victim conditionally on the type of the criminal is depicted in Table 7.4.

Case	DE			NDE		
	R^S	Criminal	Victim	R^{NS}	Criminal	Victim
A1	$T+V-L-\epsilon$	R^S-c-k^D	$-R^S-L$	$V-L-\epsilon$	$R^{NS}-c$	$R^{NS}-L$
A2	$T+V-L-\epsilon$	R^S-c-k^D	$-R^S-L$	0	$-c$	$-V$
B1	$V+\alpha T-L-\epsilon$	R^S-c-k^D	$-R^S-L$	$V+\alpha T-L-\epsilon$	$R^{NS}-c-k^N$	$-R^{NS}-L$
C1	$V+\alpha T-L-\epsilon$	R^S-c	$-R^S-L$	$V+\alpha T-L-\epsilon$	$R^{NS}-c$	$-R^{NS}-L$
C2	0	$-c$	$-V-T$	0	$-c$	$-V$

Table 7.4: The ransom and payoffs of criminals and victims in the different equilibria depending on the type of the criminal.

If the criminal is type DE then they would prefer a separating equilibrium (A1 or A2) to a pooling equilibrium because they can charge a higher ransom and obtain a higher payoff. By contrast, if the criminal is type NDE they would prefer a pooling equilibrium (B1 or C1) because they can charge a higher ransom and obtain a higher payoff. As is standard in signalling games we, thus, obtain a complex interaction in which one type, DE in our game, has incentives to signal their type, while the other type, NDE, has an incentive to hide their type. The equilibrium outcome obtained will depend on the parameters of the game.

Having looked at expected payoffs for each type of criminal we can consider the ex-ante expected payoffs for both criminal and victim. The expected utility hypothesis of Von Neumann-Morgenstern states that the choice involving uncertainty of a decision-maker can be represented by the expected value of the cardinal utility functions [41, 50]. In other words, the total expected utility can be represented as the expected value of the separate utility functions multiplied by the probability of every state. In the current context this results in:

$$U(\Theta) = \alpha U_{DE}(\Theta) + (1-\alpha)U_{NDE}(\Theta) \tag{7.1}$$

Where U(·) represents the (cardinal) utility function or payoffs, α the probability of data exfiltration and Θ the parameters V, L, T, K^D, k^N.

Expected payoffs in the signaling game for each possible type of equilibrium are shown in Table 7.5. You can see that the victim has essentially the same payoff irrespective of the type of equilibrium. This is because the criminal is able to extract the maximum surplus from the victim. To explain, consider equilibrium C2 in which the victim does not pay. In this case they suffer the recovery loss V and, with probability α reputational damage T. This, ex-ante, is the most the victim can lose from the attack. In the other types of equilibria the victim pays the ransom (with positive probability) and gains up to ϵ from doing so. We, thus,

Case	Criminal	Victim
A1	$\alpha(T-k^D)+V-L-c-\epsilon$	$\alpha T-V+\epsilon$
A2	$\alpha(T+V-L-k^D-\epsilon)-c$	$\alpha T-V+\alpha\epsilon$
B1	$\alpha(T-k^D)-(1-\alpha)k^N+V-L-c-\epsilon$	$\alpha T-V+\epsilon$
C1	$\alpha T+V-L-c-\epsilon$	$\alpha T-V+\epsilon$
C2	$-c$	$\alpha T-V$

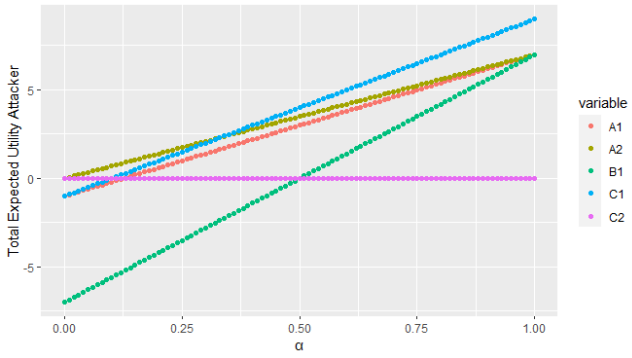
Table 7.5: Ex-ante expected payoff of criminal and victim before criminal type is determined.

see that ϵ can be interpreted as the smallest financial gain that would induce the victim to pay a ransom.

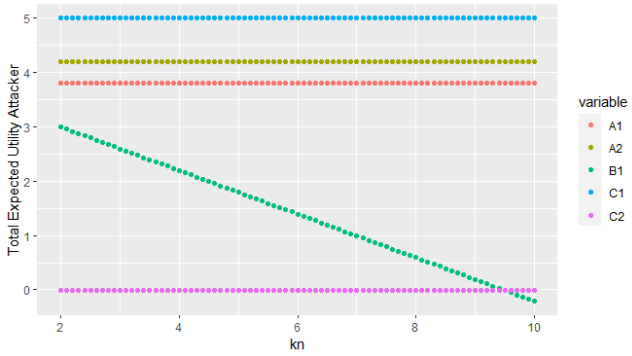
As we would intuitively expect, the victim's payoff loss from the attack is lower if the victim has active ready back-ups (which lowers V), less sensitive data (which lowers T), and measures in place to stop exfiltration (which lowers α). Crucially, these factors are beneficial to the victim irrespective of the type of equilibrium, and, thus, whether the victim pays the ransom or not, because they lower the ransom the criminal can demand. Preventive measures are, therefore, beneficial even if the victim pays the ransom.

While the victim's expected payoff does not depend on the type of equilibria, we can see in Table 7.5 that the criminal's payoff is highly dependent on the type of equilibria. To illustrate, in panels (a-c) of Figure 7.2 we plot the expected payoff of the criminal under each equilibrium type (assuming for now the equilibria exist) for fixed parameter values. We vary α, k^N and L in panels (a-c) respectively. We see that, for most parameter values, equilibria of type A2 or C1 maximize the criminal's payoff. By contrast, equilibria B1 never maximizes the criminal's payoff. This is noteworthy because equilibrium B1, in which the criminal signals data exfiltration, may appear a natural outcome. This type of equilibrium is not optimal for the criminal because they incur the costs of signaling exfiltration but cannot extract a higher ransom from signaling. Better for them to have equilibrium C1, in which they do not incur costs of signaling, or equilibrium A2, in which signaling enables a higher ransom.

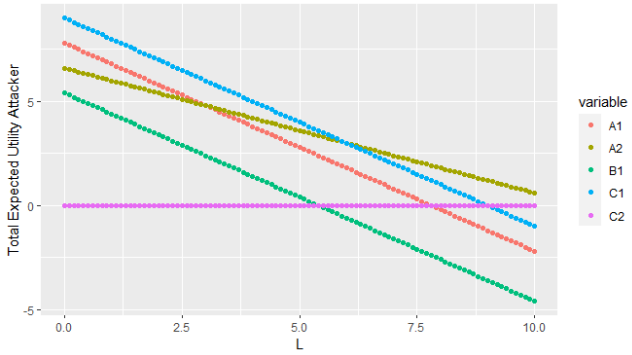
Figure 7.2(a) shows that increasing α leads to a larger expected payoff for the criminal (except for case C2). Thus, the criminal's payoff is higher if they have a higher ex-ante probability of data exfiltration. This suggests criminals have an incentive to improve their ability to exfiltrated data.



(a)



(b)



(c)

Figure 7.2: Total expected utility for the criminal when changing (a) α, (b) k^N, (c) L.

Figure 7.2(c) shows that a higher L will lead to a lower expected payoff for the criminal. This is because the higher L is reflected in a lower ransom paid. In interpretation, the legal fees are transferred from the victim to a third party (e.g. lawyers or insurers) rather than the criminals. This may be viewed as desirable from a societal perspective, although it does not materially benefit the victim.

Figure 7.2(b) shows that increasing k^N only impacts the criminal's profit in equilibrium B1. This is interesting, because increasing k^N, the cost of signaling data exfiltration when no data is exfiltrated, may seem a natural lever that victims could use to disrupt the criminal's business model. Our analysis suggests that increasing k^N may have limited impact. To explore this further we need to investigate which type of equilibria are most likely to exist in the field.

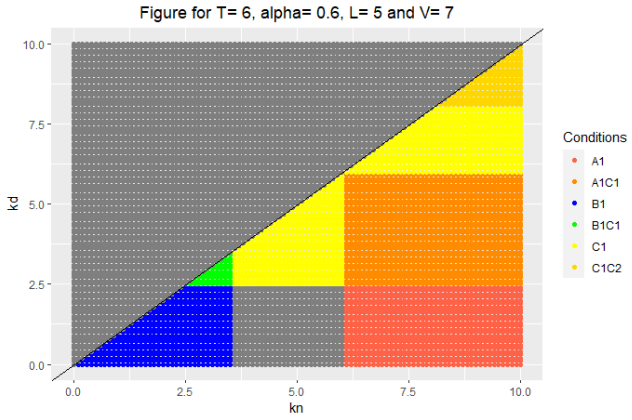
7.5.2 Overlapping Equilibria

As we have already demonstrated (see Propositions 1-3 and Table 7.3) each of the five equilibria we have identified will only exist under particular parameter values. Moreover, for a given set of parameters we may obtain multiple equilibria, a unique equilibria, or no equilibria. To illustrate, we provide three examples depicted in Figure 7.3. We indicate the existence of equilibria for combinations of k^N and k^D.

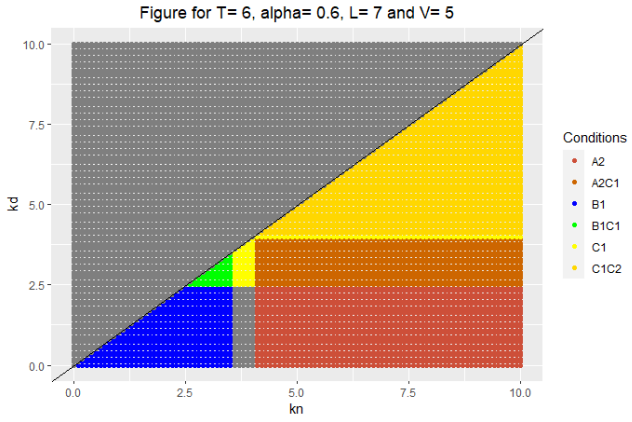
The first example, see Figure 7.2(a), has $T = 6$, $\alpha = 0.6$, $L = 5$ and $V = 7$. Here we see parameters for which the separating equilibrium A1 and the pooling equilibrium C1 exist. This occurs when k^N is large and k^D is 'intermediate'. In both equilibria the criminal does not signal if type NDE (because k^N is large). The equilibria differ in whether the criminal signals if type DE. Both equilibria are possible because k^D is an 'intermediate' range. If k^D is higher then only the pooling C1 equilibrium exists, while if it is lower only the separating A1 equilibrium exists.

We also see parameters for which both the pooling equilibrium B1, with signalling, and the pooling equilibrium C1, with no signalling, both exist. This happens for lower values of k^N and k^D. In interpretation, the criminal of type NDE will want to copy the equilibrium behavior of the type DE and it is too costly for the type DE to differentiate themselves. There are also parameter values for which there is no equilibrium. This happens for a small k^D and 'intermediate' k^N. In this case the type NDE wants to copy the type DE, but the type DE will want to differentiate themselves. There is, therefore, no stable (pure strategy) pooling equilibrium.

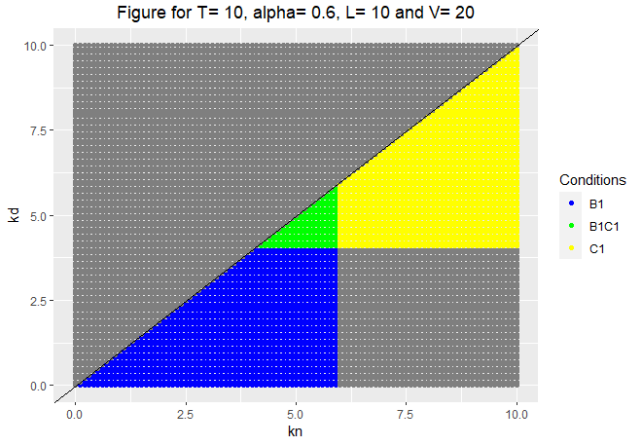
In our second example we set $T = 6$, $\alpha = 0.6$, $L = 7$ and $V = 5$. Thus, the legal fees are now larger than the recovery cost. See Figure 7.3(b). The high legal fees mean that the victim will not pay unless they have sufficiently high



(a)



(b)



(c)

Figure 7.3: Overlap of different equilibria for different parameters.

belief that data was exfiltrated. We, thus, see equilibrium A2. Otherwise, the types of equilibria we observe in our second example, as a function of k^N and k^D, are similar to those in our first example.

There are two interesting findings we will highlight from these examples. Consider our first example with $k^D = 2.4$ and $k^N = 2.5$. As we have discussed, there are two types of pooling equilibria for these parameters: B1 with signaling and C1 with no signaling. The criminal's expected payoff is 3.11 with equilibrium B1 and 5.55 with equilibrium C1. Clearly, therefore, the criminal would prefer equilibrium C1 over B1. This is because they avoid the cost of signaling data exfiltration. Criminals, however, have limited influence over which type of equilibria will emerge because it will depend on norms and historical precedence. It is possible, therefore, that a B1 equilibrium could emerge, in which criminals signal data exfiltration, even though this equilibrium is not the one they would prefer.

The second finding we would highlight is that an increase in k_N can, perhaps counter-intuitively, lead to an increased expected payoff for the criminal. To illustrate, consider the first example with $k^D = 1$ and $k^N = 2.5$. In our example this leads to equilibrium B1 with expected payoff for the criminal of 3.95. The low cost to signal data exfiltration (even if data is not exfiltrated) results in a pooling equilibrium where the criminal signals irrespective of type. Now suppose $k^N = 7.5$. In this case we obtain equilibrium A1 with an expected payoff for the criminal of 4.95. The increase in k^N increases the expected payoff of the criminal because it makes it easier for them to send a credible signal of data exfiltration. Hence, they are able to extract a higher ransom when of type DE. Also, the type NDE criminal no longer incurs the cost of signaling.

For different parameter ranges and increase in k^N can lower the expected payoff of the criminal. The general point, therefore, is that care is needed in evaluating interventions aimed at disrupting the criminal's business model. An increase in the costs of signaling data exfiltration can benefit the criminals by either making signals more credible and/or removing the incentives to send costly signals of data exfiltration.

7.5.3 Calibrating Parameter Values

For our third example we look to calibrate the parameters of the game, drawing on the data described in Section 2 with the objective to identify the most likely equilibria we would observe in the field.

1.Probability of data exfiltration α: Most criminals in our dataset try to exfiltrate data [39]. This seems in line with a Dutch whitepaper where 7

IR companies mentioned that in most ransomware attacks of there clients, data was exfiltrated [38]. This points to a high value of α. Data from Coveware in 2022 suggests that around 80% of ransomware attacks involve data exfiltration [17]. This, however, will include cases where the data exfiltrated could be deemed non-sensitive and of little value. The appropriate value of α in our model will, thus, be lower than such upper bounds. We suggest that setting $\alpha = 0.6$ strikes a reasonable balance.

2. Recovery cost V versus legal fees L: A key determinant of the type of equilibrium we obtain in our model is the relationship between V and L (see Table 7.3). There are various costs to paying a ransom that would contribute to L. These include prohibition and checks that payments are consistent with sanction legislation.[§] There are also costs to ransom negotiation and sourcing crypto-currency. Furthermore, there is evidence of negative psychological and moral consequences of paying [11, 14]. The simple reality, however, during the rapid rise of ransomware, is that a large proportion of victims pay the ransom. This trend predates the emergence of double-extortion and so is strong evidence that $V > L$ for most organisations. In other words, the financial gain from recovering access to encrypted files exceeds the costs of paying a ransom. This may be the case even if a business has back-ups, given that return of the files may allow a more rapid return to normal operations.

To give some perspective, The average financial loss reported by victims in our dataset is 555,820 euro (sd=3 million euro). The average loss when a ransom is paid is 399,098 euro (sd=0.8 million euro) while the average loss when a ransom is not paid is 674,672 euro (sd=3.9 million euro); a difference of around 275,000 euros. Furthermore, the average ransom paid is 330,326 euro (sd=0.8 million euro). Combining these two pieces of evidence, we might infer that $V - L$ is around 300,000 euro on average. In our calibration we, therefore, assume the recovery costs V are relatively large.

3. Reputation cost T: Another key determinant of the type of equilibrium in our model is the relationship between T and L and V. It is acknowledged that double-extortion has resulted in increased incentives to pay ransoms

[§]To the best of our knowledge only the United States of America state North-Carolina prohibited ransom payments by public entities. However, it is unclear what the penalty is and whether this also applies to double-extortion ransomware [32]. More generally, it is not clear that sanctions are a strong deterrent for payment [1].

[46, 40]. Indeed, analysis of our data revealed the ransom requested with data exfiltration is roughly 3 million euro and after negotiation roughly 1,7 million euro. Without data exfiltration the ransom was roughly 460,000 euro before negotiation and 135,000 euro without data exfiltration. This points to significant concerns about reputational costs [45]. Payment of a ransom does not, however, guarantee that data will not be leaked; nor does it protect the business against reputation damage or regulatory fines from the data breach [26]. We suggest, therefore, that the reputational 'savings' from paying a ransom are of secondary importance compared to recovery costs. Reputation costs are likely to be similar to legal fees in order of magnitude. Specifically, we set $V > T$ and $T = L$.

4. k^D **versus** k^N**:** The negotiations of the attacks analysed in Section 2 showed that some criminals did not send proof of data exfiltration even though analysis of logs established that data was exfiltrated. Likewise, in some cases where it was show that data was most likely not exfiltrated, the criminals said that data was exfiltrated. In most cases where is was considered likely data was exfiltrated the criminal sent proof by means of a file tree. Taken together, we will interpret evidence of signals being sent, as evidence that the costs k^D and k^N are relatively low compared to L, V and T. However, we will assume k^N is relatively large compared to k^D, because it is harder to, for example, make a file tree if no data is exfiltrated.

5. **Costs of attack** c**:** It is hard to quantify the costs criminals incur during an attack. [21] estimate the cost of a ransomware attack to be around 4,200 dollars. However, the cost of an attack seem to be related to so many variables that it is hard to give a complete estimate. For example, when the criminal is affiliated with a ransomware strain which is part of RaaS, then most probably they have to pay a part of the profits to the ransomware developers. On the other hand, the RaaS group helps with setting up the infrastructure and tooling for data exfiltration. In our sample, RaaS is more often associated with data exfiltration. This might indicate that the costs of setting up a leak site and performing data exfiltration is too much effort for an individual criminal. We elaborate on this case in the following subsection. Here we assume the costs of the attack being relatively low compared to V, T and L, based on [21].

Based on our calibration exercise we performed a simulation with $\alpha = 0.6$, $T = L = 10$ and $V = 20$. This takes into account that $V > T, L$. See Figure 7.3(c) for the set of equilibria. Since we expect signaling costs to be relatively low

compared with the other parameters, we would expect the lower-left quadrant of the graph to be most likely in real-life. This suggests equilibrium B1. Therefore, we would expect an equilibrium where (in the 'average' attack) the criminals signal that data is exfiltrated, whether data is exfiltrated or not, and the victim pays. This means criminals incur the costs of signaling. It also, as we now discuss, raises interesting questions about whether the criminals have an incentive to exfiltrate data.

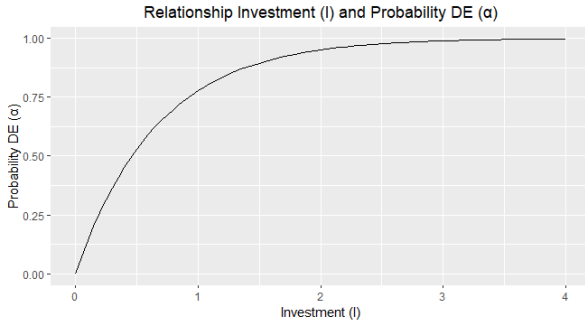
7.5.4 Increasing the Probability of Data Exfiltration

We consider B1 to be the most likely equilibrium in the field. In this case the criminal can obtain ransom $V + \alpha T - L - \epsilon$. And the expected payoff of the criminal is $\alpha(T - k^D) - (1 - \alpha)k^N + V - L - c + \epsilon$. Since (with equilibrium B1) $\alpha T > K^N$ and $k^D < k^N$ it holds that expected payoff is an increasing function of α. That is, increasing α will increase the total expected payoff of the criminal. We highlight, therefore, that while the criminals cannot extract a higher ransom from a particular attack if they exfiltrate data, they can gain across many attacks from a reputation for data exfiltration. We found cases where IR companies mentioned the reputation of the ransomware group as a possible indicator of data exfiltration: Although no evidence of data exfiltration is found during the forensic analysis, this group is well known for exfiltrating data. Reputation will be positively correlated with the value of α.[¶]

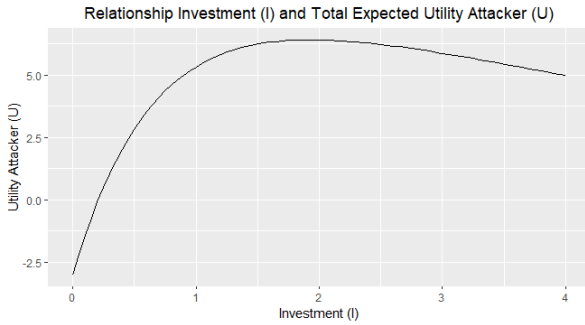
We extend our model to consider the case where the criminals can influence α by putting effort and/or investments into the attack, denoted as investment cost I. In this case the cost of an attack becomes a function of I: $c(I)$. The higher is I then the higher is α. Our signaling game is based on the assumption that investment cost I and α must be known, or common knowledge, before the game begins. The intuition is that the victim must have an idea how much the criminal has invested into data exfiltration and how that influences the probability of data exfiltration, α.

We assume that the relationship between investment I and α to be concave: initially the criminal has large investment cost. However, when the criminal decides to invest in the attack, we argue that once the right tooling and infrastructure is bought, additional investments would not increase the probability of data exfiltration substantially. Therefore the relationship between investment cost I and α is concave. While it is possible to utilize any concave function for our purposes, we have chosen to employ the Cumulative Distribution Function of the

[¶]For game theoretic analysis of ransomware and reputation we refer to [33, 10].



(a)



(b)

Figure 7.4: (a) Relationship between investment and α. (b) Relationship between investment and total expected utility criminal

exponential distribution as an illustrative example (see Figure 7.4(a), $\lambda = 1.5$).

$$\alpha(I) = 1 - e^{\lambda I} \tag{7.2}$$

Consequently, the expected payoff for the criminal in equilibrium B1 is:

$$U_{Total} = \alpha(I)(T - k^d) - (1 - \alpha(I))k^N + V - L - c - I + \epsilon \tag{7.3}$$

Plotting the total expected payoff of the criminal against investment, there seems an optimal investment which yield largest total expected payoff, see Figure 7.4(b). You can see that this results in an optimum value of α around 0.9. This

With parameters set to: T=10, k^D=1, k^N=3, L=0, V=0, c=0, ϵ=0.

is higher than we assumed in our calibration ($\alpha = 0.6$) but consistent with a relatively high α. If we increase the investment costs of data exfiltration then the optimal value of α is smaller. There are two other effects, however, we would highlight as likely to result in a smaller α in practice:

(a) Law Enforcement officers have informally disclosed that there are case studies where data exfiltration alerted the victim so that data encryption and exfiltration was not possible. In our model this suggests that increasing I could alert the victim so that they could quickly take action to stop the attack. Hence, increasing I does not necessarily increase α, and so the optimal value of α for the criminal may be correspondingly lower. In other words, the criminal may accept a lower probability of exfiltration α in order to increase the likelihood of the attack succeeding.

(b) If data exfiltration is costly the criminals might decide beforehand that they will not exfiltrate data. They might consider that the effort and costs of investments in infrastructure, and tooling might not be sufficient to cover the profits from data exfiltration. The criminal could then randomize between zero investment on some attacks and high investment on other attacks, to maximize their overall payoff. Given that a victim would not know if the criminals invested in data exfiltration in their particular attack, this again has the effect of lowering the value of α from that seen in Figure 7.4.

Endogenizing variables of our model, such as α might be an interesting way to generalize our model and could be the focus of further research.

7.6 Conclusion

7.6.1 Main Findings and Limitations

Our paper contributes to an understanding of the ransomware business model in a setting with data exfiltration and double-extortion. It is often the case (maybe typically the case) that the victim of a ransomware attack cannot know for sure whether data has been exfiltrated, particularly in the initial aftermath of the attack. If victims are willing to pay higher ransoms in the event of data exfiltration then it could be in the criminal's interests to signal data exfiltration. Drawing on a dataset provided by the Dutch Police and an incident response firm, we explored the issue of how credible victims find the claims of data exfiltration by attackers. Our findings indicate that victims display varying levels of confidence in whether data has actually been exfiltrated. Generally, the data suggests a greater willingness-to-pay on the part of victims when they believe their data has been compromised, incentivizing criminals to falsely claim that data has been exfiltrated.

We applied a signaling game model to analyse this information asymmetry, focusing on the interaction between a victim and criminal and Bayesian equilibrium strategies. Depending on various factors like the cost of sending signals and the reputation cost of actual data exfiltration, we identified five stable equilibria of the game. These range from scenarios where the attacker only signals when data is truly exfiltrated, to those where signals are sent or payments are made irrespective of the actual data exfiltration, to those where no signals are sent even if data is exfiltrated. Our analysis and calibration exercise suggests that the most likely real-world equilibrium outcome involves criminals signaling data exfiltration, whether or not exfiltration has actually occurred. Additionally, our study indicates that it could be strategically advantageous for criminals to invest more in attacks to enhance their chances of successful data exfiltration.

There are limitations in applying a game-theoretic framework to real-life situations. For instance, an assumption of common knowledge of game parameters is strong; given that ransomware remains fluid there is little opportunity for either victim or criminal to learn about each other through repeated interaction. Moreover, quantifying signaling costs and determining the extent of falsely generated credible signals pose challenges in real-life settings. Additionally, our model does not account for certain externalities, such as ethical considerations of the victim when deciding to pay the ransom, regardless of the costs or bankruptcy risks. Furthermore, the model does not account for multiple attacks by the criminal or the security behaviors of other potential victims, suggesting possible avenues for further research.

Another limitation of our analysis is that it does not explicitly differentiate between companies with and without recoverable data backups. Having recoverable backups does significantly influence the decision-making process of paying the ransom [37]. Therefore, our presented model misses an important factor influencing the decision to pay. Although the focus of the current study is only on the data-exfiltration during ransomware attacks, the proposed game theoretical models could be extended by differentiating situations with and without recoverable backups. A further way to extend our current game theoretic analysis is to consider the private information of the victim. Attackers may not know the value of the information in exfiltrated documents [38]. For instance, the documents could be in a foreign language (from the criminals perspective) or there are to many documents for the criminal to assess. In a companion paper we analyse the value of private information of the victim and concluded that private information decreases the payoff of the criminal and increases the payoff of the victim [36].

A final limitation to consider is the applicability of Nash equilibrium. While Bayesian Nash Equilibrium describe an outcome in which no one wants to change

their strategy, caution should be used in interpreting it as a prediction of behavior [8]. We may observe systematic deviations from Nash equilibrium or convergence on non-intuitive equilibria. Despite these limitations, we believe that a game theoretic analysis can still give useful insight on the incentives that ransomware criminals face. In our model we have seen the incentives for criminals to signal data exfiltration even if no exfiltration exists. This suggests that victims should be cautious of claims made by criminals, even if those claims seem credible. Importantly, there may be a 'ripple effect'; the more businesses believe data exfiltration has occurred when it did not, the more criminals have an incentive to falsely claim they have exfiltrated data. At face value, this would suggest it is in the victims interest to make it more costly for a ransomware criminal to falsely claim that data has been exfiltrated. We have seen, however, that this can have the perverse effect of benefiting the criminal (on average) because it can increase ransom demands if data is exfiltrated. It is important, therefore, to carefully consider the policy implications of our findings.

7.6.2 Recommendations for Policy Makers and Potential Victims

Our signaling game analysis yields several potential implications for policy makers and victims:

1. **Lowering the Probability of Data Exfiltration:** Victims should employ measures to decrease the likelihood of data exfiltration. In our model, this decreases α, which was one of the most important factors in determining the profitability of ransomware. The following strategies can be implemented to help achieve this:

    1. *Canary-files*: Victims can introduce "canary-files" throughout their network that generate alerts when copied or moved, thereby reducing the likelihood of successful data exfiltration [38].
    2. *Server take-down*: Engaging Law Enforcement to take down the server to which data is exfiltrated can disrupt the criminal's operations and possibly prevent successful data exfiltration. Although the criminal may have replicated the data on other servers, this action could leave a trace for investigation.
    3. *Spiking data*: Incorporating substantial amounts of fake data or deliberately contaminating the dataset can decrease the probability of valuable data being stolen [34].

2. Modifying Signal Credibility: Interventions aimed at altering the cost of signaling data exfiltration should be considered. In our models, these where variables k^D and k^N for when criminals did or did not exfiltrate data respectively.

1. *Increasing costs* k^N: Raising the costs associated with signaling data exfiltration can lead to a separating equilibrium, potentially resulting in higher payoffs for criminals. Paradoxically, investing in robust monitoring and logging systems may inadvertently increase the profitability of ransomware attacks. Criminals become more credible when they can provide a reliable signal, potentially justifying larger ransom demands. Therefore, increasing the costs of k^N does not seem an efficient defensive strategy on its own. To be effective it needs to be coupled with lowering the probability of data exfiltration α. If the probability of data exfiltration is low then an increase in signaling costs can disrupt the criminals profits because it becomes more readily apparent that data exfiltration has not taken place.
2. *Increasing costs* k^D: The criminals profit can be disrupted by efforts to increase the costs of signaling actual data exfiltration. However, implementing this approach may prove challenging. In our model, k^D represents the cost of analysing the data to provide a credible signal, and the opportunity costs of pursuing subsequent attacks. Extending the negotiation process or demanding extensive amount or time consuming evidence of data exfiltration might increase the costs of signaling. This strategy aligns with empirical evidence suggesting that prolonging negotiations can result in reduced ransom demands [39]. However, it remains unclear whether criminals will easily accept additional demands for evidence of data exfiltration.

3. Spillover Effects of Defensive Measures: It is crucial to recognize the externality effect resulting from victims defending their sensitive data. The results of this study indicate that increased data protection measures benefit not only individual businesses but also other organizations possessing vulnerable data. In particular, the more businesses invest in preventing data exfiltration the lower will be the population probability of data exfiltration α. As we have said, a decrease in α is an effective way to disrupt the criminal business model. Policy makers should acknowledge this positive externality effect and consider providing government support for cyber security investments. Neglecting this aspect may result in suboptimal

levels of cyber security prevention and recovery investments by businesses compared to the social optimum.

4. **Prohibition of Ransom Payments or tighter regulation:** Our framework does not directly capture the different pros and cons of banning ransom payments. It can, however, give insight if we think of the legal fees parameter, L, in our model. The higher is L then the lower the ransom the criminal can extract. Thus, banning ransom payments, to the extent it increases legal fees and fines, could be seen as beneficial, because it lowers the criminal's profit. However, a legal prohibition could drive ransom payments underground, and inadvertently lower the legal fees L because victims no longer seek the advice of lawyers and negotiators. Exploring the effect of secret ransom payment on social welfare could be a valuable direction for future research.

   As we have just argued, an increase in legal fees, which could include legal costs, but also negotiation costs and psychological costs, decreases the size of ransom. High legal fees, thus, disrupt the criminals profit. They do not, however, benefit the victim because they merely mean the victim is paying fees rather than ransom. One way to think of this from a societal perspective is in terms of regulation (short of banning payments). Higher levels of regulation (e.g. carefully enforced sanctions lists or requirements to alert law enforcement) would have the consequence of increasing L. The ideal would be to do this in a way that decreases ransom payments without driving ransom payments 'underground'.

In conclusion, our analysis provides valuable policy insights for addressing the challenges posed by double-extortion ransomware attacks. Implementing measures to lower the probability of data exfiltration and manipulating signal credibility can help mitigate the impact of such attacks. Additionally, policy-makers should consider the externality effect of increased data protection efforts and explore avenues for supporting cyber security investments to ensure social welfare is maximized.

7.6.3 Ethics

In conducting our research, we strictly adhere to the ethical principles outlined in the Menlo report [4], which includes the following criteria: respect for persons, beneficence, justice, and respect for law and public interest.

Respect for Persons: We prioritize the autonomy and agency of individuals involved in our research. Cases were anonymized and no personal identifiable

information is disclosed in this chapter. Our aim is to equip victims with effective methods to enhance their defensive mechanisms against cyber attacks.

Beneficence: Our research is guided by the principle of "do no harm" and aims to maximize probable benefits while minimizing potential harms. We conduct a thorough assessment of the risks and benefits associated with our research. Although the study of criminal decision-making risks of educating the criminal, we base our research on the principle of full-disclosure. Considering the entire study, we estimate that our model better informs victims and policy makers how to take preventive measures to prevent further harm than it educates criminals.

Justice: We uphold principles of fairness and equal consideration throughout our research. We strive to ensure that each individual is treated equitably and that the benefits resulting from our research are distributed to (potential) victims, companies and policy makers to prevent further harm of double-extortion ransomware.

Respect for Law and Public Interest: We conduct our research with a strong commitment to legal compliance and transparency. We engage in legal due diligence with the Dutch Police and IR company to ensure that our research adheres to applicable laws and regulations. Permission was granted to use their data after anonymizing victims and showing that no personal identifiable information is used in this chapter.

Acknowledgements

We would like to extend our sincere gratitude to the Dutch Police. In particular, we would like to thank Theo van der Plas and Cees van Tent for making the project possible. Furthermore, we thank the Cybercrime Unit East Netherlands and the Ransomware Taskforce for their expertise. Please note that the views expressed in this work are ours alone and do not necessarily reflect those of the Dutch Police. Furthermore, we would like to thank Northwave, in particular Pim Takkenberg, Erwin Maas, and Patrick van Looy.

Appendix: Proof of Propositions

Proposition 1 *If $V > L$ and $k^D < T < k^N$ there is a Bayesian equilibrium satisfying the D1 Criterion of the type A1. If $V < L$ and $k^D < T + V - L < k^N$ there is a Bayesian equilibrium satisfying the D1 Criterion of the type A2.*

Proof of Proposition 1. Equilibrium A1 can formally be written as follows: The criminal chooses to (i) Signal and set $R^S = T + V - L - \epsilon$ if type DE, and (ii) No Signal and set $R^{NS} = V - L - \epsilon$ if type NDE. The victim chooses (i) Pay if the criminal chooses Signal and asks ransom $R \leq R^S$, (ii) No Pay if Signal and $R > R^S$, (iii) Pay if No Signal and $R \leq R^{NS}$, and (iv) No Pay if No Signal and $R > R^{NS}$.

Consider the victim. Suppose the criminal has chosen signal and ransom R^S. The Bayesian updated belief of the victim should be $\mu(DE|S) = 1$. The expected payoff of the victim if they Pay is $U = -R^S - L = -T - V + \epsilon$. The expected payoff if they choose No Pay is $U = -T - V$. It is, thus, optimal to pay.

Suppose the Criminal has chosen No Signal. The Bayesian updated belief of the victim in this case is $\mu(NDE|NS) = 0$. The expected payoff of the victim if they Pay is $U = -R^{NS} - L = -V + \epsilon$. The expected payoff if they choose to No Pay is $U = -V + \epsilon$. This implies that it is optimal for the victim to pay if $L < V$ and optimal to Not Pay if $L \leq V$.

Consider now the incentive of the criminal. Suppose the criminal is type $\tau = DE$. On the equilibrium path they receive payoff $U = T + V - L - \epsilon - c - k^D$. We argue that, if the criminal chooses Signal, then they cannot gain from choosing $R \neq R^S$, provided $R^S > 0$: If they choose a ransom $R < R^S$ then the victim pays a smaller ransom, and if $R > R^S$ then the victim does not pay and the criminal has payoff $U = -c - k^D$. We have $R^S > 0$ if $T + V > L$. We also argue the criminal cannot gain from choosing No signal. In doing so, we distinguish two cases $L > V$ and $V > L$. If $L > V$ then the maximum ransom the victim is willing to pay is negative. Hence, the victim will choose No Pay and the criminal has payoff $U = -c$. It is, thus, optimal to Signal if $T + V - L > k^D$. If $L < V$ then the victim will pay a ransom up to $R = V - L - \epsilon$. The criminal can do no better than set ransom $R = V - L - \epsilon$. Thus, the criminals payoff is $U = V - L - \epsilon - c$. It is, thus, optimal to Signal if $T > k^D$.

Suppose the criminal is type $\tau = NDE$. The argument above naturally extends to this case, except we now derive, respective, conditions $T > k^N > k^D$ and $T + V - L > k^N > k^D$. Proving $\mu(NDE|NS) = 0$ is trivial: in the separating equilibrium the strategy of the Criminal is type $\tau = NDE$ is to not signal and

for type $\tau = DE$ to signal. Therefore $\mu(NDE|NS) = 0$ and $\mu(DE|S) = 1$ is consistent with the strategy of the criminal in the equilibrium.

Equilibrium A2 can formally be written as follows: The criminal chooses to (i) Signal and set $R^S = T + V - L - \epsilon$ if type DE, and (ii) No Signal and set $R^{NS} = 0$ if type NDE. The victim chooses (i) Pay if the criminal chooses Signal and asks ransom $R \leq R^S$, (ii) No Pay if Signal and $R > R^S$, (iii) No Pay if No Signal and $R \geq R^{NS}$. The arguments for the proof of existence of equilibrium A1 can naturally be applied to show proof of the existence of equilibrium A2.

■

Proposition 2 *If (a) $V > L$ and $\alpha T > k^N$, or (b) $V + \alpha T > L > V$ and $\alpha T + V - L > k^N$ there is a signaling equilibrium satisfying the D1 Criterion of the type B1.*

Proof of Proposition 2. The equilibrium has the following properties: The criminal chooses Signal and $R^S = \alpha T + V - L - \epsilon$ if $\tau = NDE$ and $\tau = DE$. The victim chooses (i) Pay if Signal and $R \leq R^S$, (ii) No Pay if Signal and $R > R^S$, (iii) Pay if No Signal and $R \leq V - L - \epsilon$, and (iv) No Pay if No Signal and $R \geq V - L$.

Consider the victim. Suppose the Criminal has chosen Signal and ransom R^S. The Bayesian updated belief of the victim should be $\mu(DE|S) = \alpha$. So, the expected payoff of the victim if they Pay is $U = -R^S - L = -\alpha T - V + \epsilon$. The expected payoff if they choose No Pay is $U = -\alpha T - V$. It is, thus, optimal to pay.

Suppose the Criminal has chosen No Signal. For now we assume the belief of the victim is $\mu(DE|NS) = 0$. Suppose the ransom is $R = V - L - \epsilon$. The expected payoff of the victim if they Pay the ransom is $U = -V + \epsilon$. The expected payoff if they choose No Pay is $U = -V$. It is, thus, optimal for the victim to Pay.

Consider now the incentive of the criminal. Suppose the criminal is type $\tau = DE$. On the equilibrium path they receive payoff $U = \alpha T + V - L - \epsilon - c - k^D$. We argue that, if the criminal chooses Signal, then they cannot gain from choosing $R \neq R^S$, provided $R^S > 0$: If they choose a ransom $R < R^S$ then the victim pays a smaller ransom, and if $R > R^S$ then the victim does not pay and the criminal has payoff $U = -c - k^D$. We have $R^S > 0$ if $\alpha T + V > L$. We also argue the criminal cannot gain from choosing No signal. In doing so, we distinguish two cases $L > V$ and $V > L$. If $L > V$ then the maximum ransom the victim is willing to pay is negative. Hence, the victim will choose No Pay and the criminal has payoff $U = -c$. It is, thus, optimal to Signal if $\alpha T + V - L > k^D$. If $L < V$ then the victim will pay a ransom up to $R = V - L - \epsilon$. The criminal can do no better

than set ransom $R = V - L - \epsilon$. Thus, the criminals payoff is $U = V - L - \epsilon - c$. It is, thus, optimal to Signal if $\alpha T > k^D$.

Suppose the criminal is type $\tau = NDE$. The argument above naturally extends to this case, except we now derive, respective, conditions $\alpha T > k^N > k^D$ and $\alpha T + V - L > k^N > k^D$.

It remains to show that $\mu(DE|NS) = 0$. Here, we invoke the D1 Criterion. Consider $V > L$. Suppose that the criminal chooses No Signal and ransom demand $R > 0$. Let p be the probability that the victim will pay. The type $\tau = DE$ will receive a weakly higher payoff than in equilibrium if

$$p^{DE} \geq \frac{\alpha T + V - L - k^D}{R}.$$

The type $\tau = NDE$ will receive a strictly higher payoff than in equilibrium if

$$p^{NDE} > \frac{\alpha T + V - L - k^N}{R}.$$

Given that $k^N > k^D$ we have $p^{NDE} < p^{DE}$. Thus, type $\tau = DE$ can be eliminated using the D1 Criterion. It follows that $\mu(DE|NS) = 0$ is consistent with the D1 Criterion. ■

Proposition 3 *If $V + \alpha T > L$ and $(1 - \alpha)T < k^D$ there exists a signaling equilibrium satisfying the D1 Criterion of the type C1. If $V + \alpha T > L > V$ and $T + V - L < k^D$ there exists a signaling equilibrium satisfying the D1 Criterion of the type C2*

Proof of Proposition 3.

Equilibrium C1 has the property: The criminal chooses Signal and $R^{NS} = \alpha T + V - L - \epsilon$ if $\tau = NDE$ and $\tau = DE$. The victim chooses (i) Pay if Signal and $R \leq R^S$, (ii) No Pay if Signal and $R > R^S$, (iii) Pay if No Signal and $R < R^{NS}$, and (iv) No Pay if No Signal and $R > R^{NS}$.

Consider the victim. Suppose the Criminal has chosen Signal and ransom R^S. The Bayesian updated belief of the victim should be $\mu(DE|S) = 1$. So, the expected payoff of the victim if they Pay is $U = -R^B - L = -T - V + \epsilon$. The expected payoff if they choose No Pay is $U = -T - V$. It is, thus, optimal to pay.

Suppose the Criminal has chosen No Signal. For now we assume the belief of the victim is $\mu(DE|NS) = \alpha$. Suppose the ransom is $R^{NS} = \alpha T + V - L - \epsilon$. The expected payoff of the victim if they Pay the ransom is $U = -\alpha T - V + \epsilon$. The expected payoff if they choose No Pay is $U = -V$. It is, thus, optimal for the victim to Pay if $(1 - \alpha)T < k^D$.

Consider now the incentive of the criminal. Suppose the criminal is type $\tau = DE$. On the equilibrium path they receive payoff $U = \alpha T + V - L - \epsilon - c$. We argue that, if the criminal chooses Signal, then they cannot gain from choosing $R \neq R^{NS}$, provided $R^{NS} > 0$: If they choose a ransom $R < R^{NS}$ then the victim pays a smaller ransom, and if $R > R^{NS}$ then the victim does not pay and the criminal has payoff $U = -c - k^D$. We have $R^{NS} > 0$ if $\alpha T + V > L$. We also argue the criminal cannot gain from choosing No signal. In doing so, we distinguish two cases $L > V$ and $V > L$. If $L > V$ then the maximum ransom the victim is willing to pay is negative. Hence, the victim will choose No Pay and the criminal has payoff $U = -c$. It is, thus, optimal to Signal if $\alpha T + V - L > k^D$. If $L < V$ then the victim will pay a ransom up to $R = \alpha T + V - L - \epsilon$. The criminal can do no better than set ransom $R = \alpha T + V - L - \epsilon$. Thus, the criminals payoff is $U = \alpha T + V - L - \epsilon - c$. It is, thus, optimal to Not Signal if $(1 - \alpha T) < k^D$ or $V + T - L < k^D$.

Suppose the criminal is type $\tau = NDE$. The argument above naturally extends to this case, except we now derive, respective, conditions $(1 - \alpha T) < k^D < k^N$ and $T + V - L < k^D < k^N$.

It remains to show that $\mu(DE|S) = 1$. Here, we invoke the D1 Criterion. Consider $V > L$. Suppose that the criminal chooses Signal and ransom demand $R > 0$. Let p be the probability that the victim will pay. The type $\tau = DE$ will receive a weakly higher payoff than in equilibrium if

$$p^{DE} \geq \frac{T + V - L - k^D}{R}.$$

The type $\tau = NDE$ will receive a strictly higher payoff than in equilibrium if

$$p^{NDE} > \frac{T + V - L - k^N}{R}.$$

Given that $k^N > k^D$ we have $p^{NDE} < p^{DE}$. Thus, type $\tau = NDE$ can be eliminated using the D1 Criterion. It follows that $\mu(NDE|S) = 0$ is consistent with the D1 Criterion. It follows that $\mu(DE|S) = 1$ is consistent with the D1 Criterion. ■

Bibliography

[1] C. Abely. 'Ransomware, Cyber Sanctions, and the Problem of Timing'. *BCL Rev. E. Supp. I-* 63, 2022, p. 47.

[2] G. A. Akerlof. 'The market for "lemons": Quality uncertainty and the market mechanism'. *The quarterly journal of economics* 84.3, 1970, pp. 488–500.

[3] L. Allodi, F. Massacci and J. M. Williams. 'The work-averse cyber attacker model: Theory and evidence from two million attack signatures'. en. *SSRN Electron. J.*, 2017.

[4] M. Bailey, D. Dittrich, E. Kenneally and D. Maughan. 'The menlo report'. *IEEE Security & Privacy* 10.2, 2012, pp. 71–75.

[5] J. S. Banks and J. Sobel. 'Equilibrium selection in signaling games'. *Econometrica: Journal of the Econometric Society*, 1987, pp. 647–661.

[6] K. Bedard. 'Human capital versus signaling models: university access and high school dropouts'. *Journal of political economy* 109.4, 2001, pp. 749–775.

[7] N. L. Beebe and V. S. Rao. *Using situational crime prevention theory to explain the effectiveness of information systems security*. Las Vegas, 2005.

[8] J. Brandts and C. A. Holt. 'An experimental test of equilibrium dominance in signaling games'. *The American Economic Review* 82.5, 1992, pp. 1350–1365.

[9] R. Brewer. 'Ransomware attacks: detection, prevention and cure'. *Network Security* 2016.9, 2016, pp. 5–9.

[10] A. Cartwright and E. Cartwright. 'Ransomware and reputation'. *Games* 10.2, 2019, p. 26.

[11] E. Cartwright, J. Hernandez Castro and A. Cartwright. 'To pay or not: game theoretic models of ransomware'. *Journal of Cybersecurity* 5.1, 2019, tyz009.

[12] R. V. Clarke. 'Situational crime prevention'. *Environmental criminology and crime analysis*. Routledge, 2016, pp. 305–322.

[13] L. Connolly, M. Lang, P. Taylor and P. Corner. 'The Evolving Threat of Ransomware: From Extortion to Blackmail', 2021.

[14] S. Corbet and J. W. Goodell. 'The reputational contagion effects of ransomware attacks'. *Finance Research Letters* 47, 2022, p. 102715.

[15] D. Cornish and R. Clarke. 'Understanding crime displacement: An application of rational choice theory'. *Criminology* 25.4, 1987, pp. 933–948.

[16] D. B. Cornish and R. V. Clarke. 'The reasoning criminal: Rational choice perspectives on offending', 2014.

[17] A. Culafi. *Coveware: Double-extortion ransomware attacks fell in Q1*. 2022. URL: https://www.techtarget.com/searchsecurity/news/252516732/Coveware-Double-extortion-ransomware-attacks-fell-in-Q1.

[18] T. Cymru. 'Analyzing ransomware negotiations with CONTI: An in-depth analysis', 2022.

[19] Ecrime. *Gallery of 97 ransomware and data leak sites*. 2023. URL: https://ecrime.ch/screenshots/.

[20] D. Fudenberg and J. Tirole. *Game theory*. MIT press, 1991.

[21] E. Galinkin. 'Winning the Ransomware Lottery: A Game-Theoretic Approach to Preventing Ransomware Attacks'. *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Virtual Event, October 25–27, 2021, Proceedings 12*. Springer. 2021, pp. 195–207.

[22] A. Glazer and K. A. Konrad. 'A signaling explanation for charity'. *The American Economic Review* 86.4, 1996, pp. 1019–1028.

[23] D. Gonzalez and T. Hayajneh. 'Detection and prevention of crypto-ransomware'. *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE. 2017, pp. 472–478.

[24] S. Greengard. 'The worsening state of ransomware'. *Communications of the ACM* 64.4, 2021, pp. 15–17.

[25] P. Hack and Z.-Y. Wu. 'we wait, because we know you." inside the ransomware negotiation economics.'" *NCC Group, Nov* 12, 2021.

[26] N. Hodge. *Paying ransom to avoid GDPR fine an unwise gambit*. 2023. URL: https://www.complianceweek.com/cybersecurity/paying-ransom-to-avoid-gdpr-fine-an-unwise-gambit/33507.article.

[27] K. Huang, M. Siegel and S. Madnick. 'Systematically understanding the cyber attack business: A survey'. *ACM Computing Surveys* 51.4, 2018, pp. 1–36.

[28] Q. Kerns, B. Payne and T. Abegaz. 'Double-Extortion Ransomware: A Technical Analysis of Maze Ransomware'. *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 3*. Springer. 2022, pp. 82–94.

[29] D. M. Kreps and J. Sobel. 'Signalling'. *Handbook of game theory with economic applications* 2, 1994, pp. 849–867.

[30] A. Laszka, S. Farhang and J. Grossklags. 'On the economics of ransomware'. *Decision and Game Theory for Security: 8th International Conference, GameSec 2017, Vienna, Austria, October 23-25, 2017, Proceedings*. Springer. 2017, pp. 397–417.

[31] A. Laszka, E. Panaousis and J. Grossklags. 'Cyber-insurance as a signaling game: Self-reporting and external security audits'. *Decision and Game Theory for Security: 9th International Conference, GameSec 2018, Seattle, WA, USA, October 29–31, 2018, Proceedings 9*. Springer. 2018, pp. 508–520.

[32] J. Lewis. 'North Carolina prohibits public sector entities paying ransom ransomware cyberattack'. *The National Law Review*, 2022. URL: https://www.natlawreview.com/article/north-carolina-prohibits-public-sector-entities-paying-ransom-ransomware-cyberattack.

[33] Z. Li and Q. Liao. 'Game Theory of Data-selling Ransomware'. *Journal of Cyber Security and Mobility*, 2021, pp. 65–96.

[34] Z. Li and Q. Liao. 'Preventive portfolio against data-selling ransomware—A game theory of encryption and deception'. *Computers & Security* 116, 2022, p. 102644.

[35] Z. Li and Q. Liao. 'Ransomware 2.0: to sell, or not to sell a game-theoretical model of data-selling ransomware'. *Proceedings of the 15th International Conference on Availability, Reliability and Security*. 2020, pp. 1–9.

[36] T. Meurs, E. Cartwright and A. Cartwright. 'Double-sided Information Asymmetry in Double Extortion Ransomware'. *14th International Conference on Decision and Game Theory for Security, GameSec 2023*. 2023.

[37] T. Meurs, E. Cartwright, A. Cartwright, M. Junger, R. Hoheisel, E. Tews and A. Abhishta. 'Ransomware Economics: A Two-Step Approach To Model Ransom Paid'. *18th Symposium on Electronic Crime Research, eCrime 2023*. 2023.

[38] T. Meurs and L. Holterman. *Whitepaper data-exfiltratie bij een ransomware-aanval*. 2022. URL: https://executivefinance.nl/wp-content/uploads/2023/01/VCNL-Whitepaper-Exfiltratie.pdf.

[39] T. Meurs, M. Junger, E. Tews and A. Abhishta. 'Ransomware: How attacker's effort, victim characteristics and context influence ransom requested, payment and financial loss'. *Symposium on Electronic Crime Research, eCrime 2022*. 2022.

[40] G. Mott, S. Turner, J. R. Nurse, J. MacColl, J. Sullivan, A. Cartwright and E. Cartwright. 'Between a rock and a hard (ening) place: Cyber insurance in the ransomware era'. *Computers & Security* 128, 2023, p. 103162.

[41] Y.-K. Ng. 'Expected subjective utility: Is the Neumann-Morgenstern utility the same as the neoclassical's?' *Social Choice and Welfare* 1.3, 1984, pp. 177–186.

[42] P. S. Nyakomitta and S. O. Abeka. 'A Survey of Data Exfiltration Prevention Techniques'. *International Journal of Advanced Networking and Applications* 12.3, 2020, pp. 4585–4591.

[43] M. J. Osborne et al. *An introduction to game theory*. Vol. 3. 3. Oxford university press New York, 2004.

[44] D. Palmer. *The ransomware problem isn't going away, and these grim figures prove it*. 2023. URL: https://www.zdnet.com/article/these-grim-figures-show-that-the-ransomware-problem-isnt-going-away/.

[45] N. Pattnaik, J. R. Nurse, S. Turner, G. Mott, J. MacColl, P. Huesch and J. Sullivan. 'It's more than just money: The real-world harms from ransomware attacks'. *International Symposium on Human Aspects of Information Security and Assurance*. Springer. 2023, pp. 261–274.

[46] B. Payne and E. Mienie. 'Multiple-extortion ransomware: The case for active cyber threat intelligence'. *ECCWS 2021 20th European Conference on Cyber Warfare and Security*. Academic Conferences Inter Ltd. 2021, p. 331.

[47] R. Richardson and M. North. 'Ransomware: Evolution, mitigation and prevention'. *International Management Review* 13.1, 2017, p. 10.

[48] M. Spence. 'Competitive and optimal responses to signals: An analysis of efficiency and distribution'. *Journal of Economic theory* 7.3, 1974, pp. 296–332.

[49] H. Tuttle. 'Ransomware attackers turn to double extortion'. *Risk Management* 68.2, 2021, pp. 8–9.

[50] J. Von Neumann and O. Morgenstern. 'Theory of games and economic behavior'. *Theory of games and economic behavior*. Princeton university press, 1944.

[51] R. Wortley and M. Townsley. 'Environmental criminology and crime analysis: Situating the theory, analytic approach and application'. *Environmental criminology and crime analysis*. Routledge, 2016, pp. 20–45.

[52] Z. Xu and Q. Hu. 'The role of rational calculus in controlling individual propensity toward information security policy non-compliance behavior'. *Proceedings of the 51st Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, 2018.

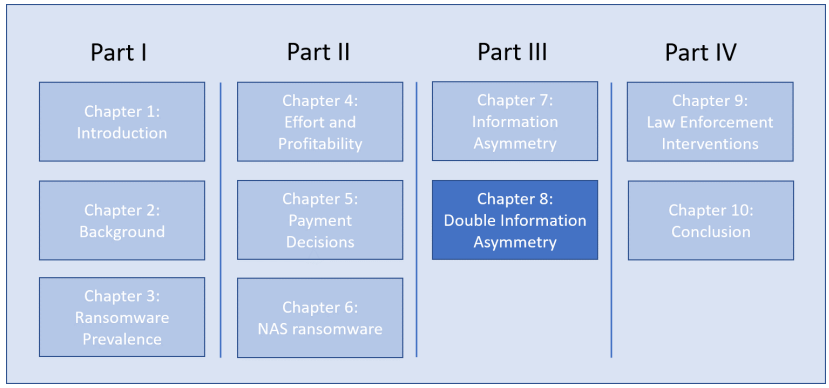
This page is intentionally left blank.

Into the unknown

~ Elsa, Frozen

Chapter 8

Double Deception in Double-Extortion Ransomware



In this chapter we model two important sources of asymmetric information between victim and attacker: (a) Victims are typically uncertain whether data is exfiltrated, due to for example misconfigured monitoring systems. (b) It is hard for attackers to estimate the value of compromised files. We use game theory to analyse the payoff consequences of such private information. Specifically, we analyse a signaling game with double-sided information asymmetry: (1) attackers know whether data is exfiltrated and victims do not, and (2) victims know the value of data if it is exfiltrated, but the attackers do not.

8.1 Introduction

The last decade has seen a rapid rise in crypto-ransomware attacks impacting individuals, businesses, charities and public organisations [8, 7, 26, 13, 27, 31]. Crypto-ransomware, or ransomware for short, is broadly defined as the use of crypto-techniques to encrypt the files of a victim, after which the attackers ask for a ransom to decrypt the files [38, 23]. Ransomware has proved highly profitable for criminal gangs, primarily because many victims pay the ransom in order to receive the decryption keys [29, 25]. It has also proved highly disruptive for business and society, leading to an increasing policy focus on how to disrupt the business model of criminals. The ransomware gangs, however, are continually evolving their strategy, not only in terms of technical sophistication but also economic sophistication, to maximize gains [16].

Since roughly 2019, ransomware groups have been experimenting with double-extortion [14, 6]. In this case the attackers not only encrypt files, but also exfiltrate data with the purpose to sell or publish the data if the victim does not pay [20, 19, 27, 23]. The criminals can, thus, extort the victim for access to the decryption key and to avoid data leak. Double-extortion has been shown to increase the ransom requested and ransom amount paid [24, 26]. Specifically, Meurs et al. [26] analysed 353 ransomware attacks reported to the Dutch Police and found a significant positive effect of data exfiltration on ransom requested. In a follow-up study, Meurs et al. [25] analysed 429 ransomware attacks reported to the Dutch Police and an Incident Response company. They applied a two-step statistical procedure to measure how data exfiltration influences both the frequency of ransom payments and the ransom amount paid. No significant effect of data exfiltration on frequency of ransom paid was identified. However, the ransom amount paid was 5.5 times larger with data exfiltration than without data exfiltration. This trend aligns with the observations of Matthijsse et al. [23], who reported that cyber security experts consider double-extortion tactics to have become a standard modus operandi among ransomware criminals.

Given the prevalence of double-extortion, one important issue for victims of a ransomware attack is determining whether data was exfiltrated [24]. Due to the deletion of log files by attackers, or misconfigured monitoring systems, victims often do not know whether data was exfiltrated [33, 34]. This means that an attacker who has not exfiltrated data can still threaten the publication of data, to get a larger ransom paid. On the flip side, the claims of an attacker that has exfiltrated data may be viewed as less-credible, empty threats, by the victim. Attackers are, thus, increasingly trying to send credible signals that data was

exfiltrated. For instance, to back up their claim, some attackers send evidence of exfiltration by means of a file tree of the exfiltrated data or a couple of files. Such signals could, however, still be sent, even if at a higher cost, by attackers who have not exfiltrated data.

Meurs et al. [24] explored this one-sided information asymmetry with a game theoretic signaling game. In the signaling game, the attacker learns whether data is exfiltrated or not and then decides whether to send a signal to the victim, or not. Calibrating their results with empirical data, the authors concluded that a pooling equilibrium is most likely in real-life, where attackers send a signal of data exfiltration, regardless of actual data exfiltration. The authors, thus, concluded that victims should be sceptical when attackers claim that data is exfiltrated. Moreover, criminals benefit from their private information. In this chapter we extend the game theoretic approach by taking into account a further important information asymmetry between victim and criminal.

In practice it is hard for attackers to determine the value of encrypted files for the victim. The filenames and files which contain text are often in a foreign or technical language, and the sensitivity of data is difficult to judge without insider understanding. Furthermore, it takes effort to estimate the importance of, potentially, millions of files. Attackers are, therefore, likely to be imperfectly informed of the value of files, even if data is exfiltrated. Combined, therefore, we have two information asymmetries in double-extortion ransomware attacks. First, the victim does not know whether data was exfiltrated or not, but the attacker does. Second, the victim knows whether potentially exfiltrated data is valuable or not, but the attacker does not. Here, we define valuable data for the victim, as data which would result in large reputation costs if made accessible for the general public, competitors or similar.

To our knowledge, no previous studies have modelled this two-sided information asymmetry of data exfiltration, and analysed how it impacts the profitability of attacks. Most empirical [26] and game-theoretical modeling [20, 19] of double-extortion ransomware has focused on the extra profits for attackers by conducting data exfiltration and encryption, compared to only data encryption. We address the relationship between the uncertainty of data exfiltration and profitability by analysing a signaling game. Signaling games provide a way to model a strategic game with incomplete information and sequential choice [10, 15, 1, 22, 30]. The basic premise is that a player holding extra information could try to influence the other players by sending a credible signal of their information. Signalling games provide a natural framework with which to explore double-extortion and the payoff consequences of asymmetric information. For a more detailed explanation of signaling games we refer to Osborne [30].

The value in studying double-extortion ransomware through the lens of game theory is to better understand how the business model of the criminals can be disrupted. In our setting, the victims will pay a ransom if it is financially beneficial to do so. Disruption of the business model, thus, comes from lowering the ransom that victims would be willing to pay. The implicit assumption is that lower ransoms would make ransomware less appealing for the criminals. A specific focus of the current paper is to analyse whether private information on the part of the victim, on the value of encrypted files, results in lower equilibrium ransoms. We find that it does. Private information can, thus, be one instrument to disrupt the ransomware business model and lower criminal profits. In practical terms this means victims should reveal as little as possible about the value of encrypted files, either through private negotiation or public communication.

Our work provides the following key contributions: First, we provide a game-theoretical framework to analyse the double-sided information asymmetry in double-extortion ransomware attacks. The framework consists of a signaling game, wherein the attacker can send a costly signal of data exfiltration that can inform the victim's beliefs and payment decision. Second, we identify four separating and eight pooling equilibria of the game and their underlying conditions. The type of equilibria that exists in the game will depend on the parameters of the game, particularly the cost of signaling data exfiltration, the cost to recover files without decryption, the reputation loss from data leakage, and the probability the victim's files contain valuable data. We identify the factors determining how much surplus the attacker can extract from the victim. Third, we analyse the impact that private information of the victim has on the profitability of the attack. Through examples, we show that the payoff loss to the criminal from now knowing the value of files can range from zero to over 20%.

We remark that our paper adds to a growing literature using game theory to analyse the ransomware decision process [5, 11, 4, 9]. Prior game-theoretical studies have focused on the interaction of ransomware and victim's decision to invest in security measures like backups or insurance [38, 2, 32, 36]. For instance, Laszka, Farhang and Grossklags [17] focused on modeling the ransomware ecosystem as a whole and how backup decisions affect the ransomware ecosystem. Vakilinia et al. [35] take a different approach in exploring how a double sided auction can facilitate the negotiation between attacker and victim to achieve a 'fair' ransom. Galinkin [11] analyses measures that an attacker can disrupt the business model of the attackers by lowering the profitability of ransomware attacks. The main intervention suggested is that of back-ups. We note, however, that in a setting with double-extortion, back-ups are not enough to combat the

ransomware threat. We must also consider the reputational costs from the publication of exfiltrated data.

We proceed as follows. In Section 2 we introduce the signalling game. In Section 3 we provide our main results. In Section 4 we conclude.

8.2 Signaling Game

We consider a two-player game between a criminal, henceforth called the attacker, and a victim. In application we will focus on the victim being an organisation but our analysis does not preclude the victim being an individual. We take as given that the victim has been subject to a ransomware attack and their data has been encrypted. The attacker is demanding a ransom for the decryption key.

If the victim does not pay the ransom then it will cost V_P to recover normal operations. The size of V_P will depend on a range of factors such as the availability of (functional) back-ups, the victim's reliance on the encrypted files for day-to-day operations, and the speed with which the organisation can return to normal operations. If the victim does pay the ransom then we assume the attackers will provide the decryption key and it will cost V_{NP} for the victim to restore normal operations. The size of V_{NP} may include factors such as the cost of decrypting files and the speed with which they can be decrypted. From a game theoretic point of view, the predictions of our model depend solely on the difference in recovery cost from paying versus not paying $V_P - V_{NP}$. Thus, to simplify the model, and without loss of generality, we set $V_{NP} = 0$ and $V_P = V$. We make the very mild assumption that $V \geq 0$ and so access to the decryption key cannot increase recovery costs. We will comment below on the case $V = 0$ where the decryption key is essentially 'worthless'.

We take it as given that, as well as encrypting files, the attacker attempted to exfiltrate data of the victim. We model two forms of asymmetric information or, equivalently, incomplete information between the victim and attacker:

- The attempt to exfiltrate data may or may not have been 'successful'. Let α denote the prior probability that data was exfiltrated. Crucially, we assume that the attacker knows if data is exfiltrated but the victim does not know. The incomplete information of the victim means the criminal can threaten to publish data even if no data was exfiltrated. In modelling games of incomplete information it is standard to distinguish (Harsanyi) types of a player [12, 10]. In this case the attacker can be of type 'data was successfully exfiltrated' or type 'data was not exfiltrated'. We use the terms DE and NDE, respectively, to distinguish the type of attacker.

- Exfiltration of data will cause reputational damage to the victim. Crucially, we assume that the victim knows the size of this damage but the criminal does not. For instance, the victim knows whether the data includes sensitive information about customers, employees etc. We assume that there are two types of victim: those with sensitive data, called high type, and those without, called low type. If exfiltrated data were to be leaked then the victim would incur reputation costs T_1 or $T_0 < T_1$ depending on whether they are high or low type, respectively. If the data is not leaked then we assume there is no reputation cost. The prior probability the victim is high type is β.

The game has three stages.

1. Following the approach of Harnsanyi [12], Nature determines the type of the victim (high or low type) and the type of the criminal (data exfiltrated or no data exfiltrated) in Stage 1 of the game. The victim learns their type (with probability β they are high type), and the attacker learns whether data was exfiltrated (with probability α it is exfiltrated).

2. In stage 2 the attacker chooses (a) whether or not to send a signal that data has been exfiltrated, and (b) the size of ransom demand. The signal can, for instance, consist of a picture of the file tree of the exfiltrated data, or a sample of exfiltrated data. The cost to the attacker of sending a signal when data is exfiltrated is k_D and the cost when data is not exfiltrated is k_N. We assume that it is more costly to send a signal if no data is exfiltrated, hence, $k_D < k_N$. The attacker can choose any ransom demand. To simplify notation we denote by R^S the ransom demand of the attacker if they send a signal and R^{NS} the demand if no signal is sent.*

3. In stage 3 the victim observes whether or not a signal was sent, and learns the ransom demand. The victim then chooses whether to pay the ransom or not. To simplify the analysis we assume an ultimatum bargaining game in which there is no opportunity for negotiation. Thus, the victim is given a take-it-or-leave it offer and the choice to pay or not ends the game.

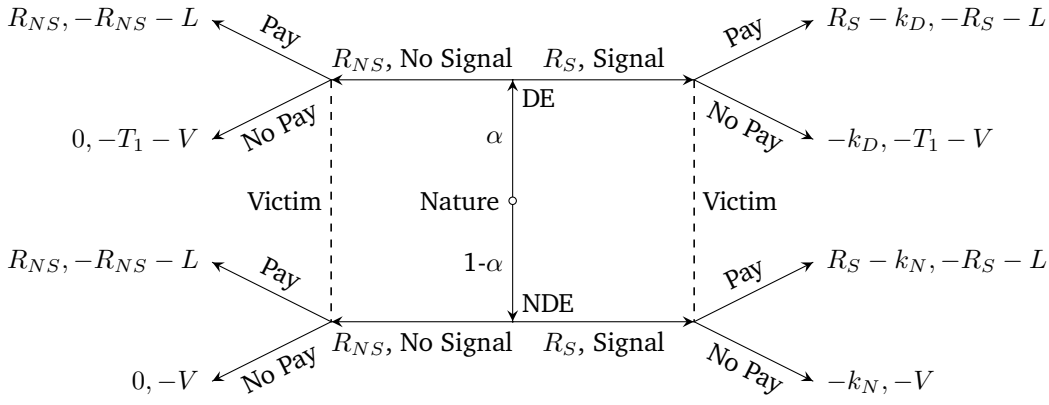
*The attacker could choose any ransom above 0 for any combination of both own type and signal. So, suppose, more generally, we denote by $R^S_{DE}, R^S_{NDE}, R^{NS}_{DE}$ and R^{NS}_{NDE} the ransom of a type DE or NDE if they signal or do not signal. There cannot be an equilibrium in which an attacker of type DE and NDE signal and $R^S_{NDE} \neq R^S_{DE}$; this would reveal the attacker if type NDE and, thus, make their signal ineffective. Similarly, there cannot be an equilibrium in which an attacker of type DE and NDE would not signal and $R^{NS}_{NDE} \neq R^{NS}_{DE}$; this would again reveal the attacker if type NDE and lower the ransom the victim would rationally pay.

Table 8.1: Variables used in the data exfiltration signaling game

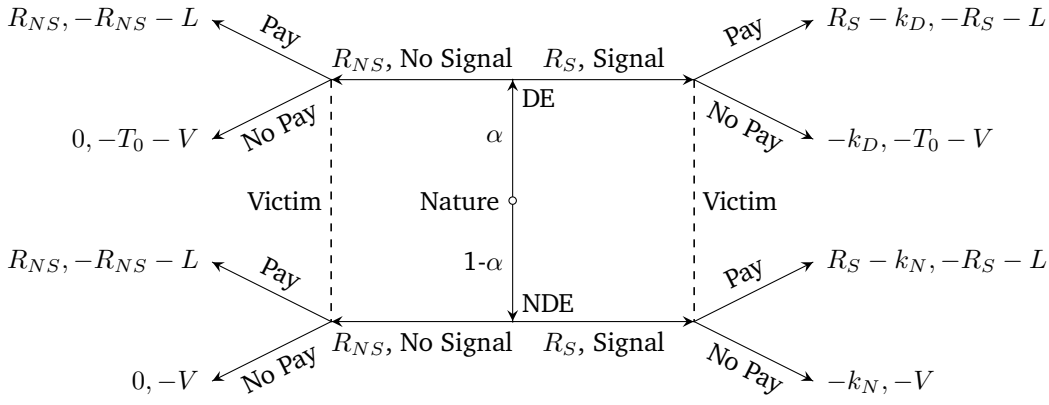
	Variable	Description
Attacker	R_S	Ransom when signaling
	R_{NS}	Ransom when not signaling
	k_D	Cost of signal with data exfiltration
	k_N	Cost of signal without data exfiltration
	β	Probability of data being valuable
Victim	T_1	Reputation cost for valuable data
	T_0	Reputation cost for non-valuable data
	V	Recovery cost without decryption key
	L	Legal fees of paying ransom
	α	Probability of data exfiltration
	μ	Belief on probability of data exfiltration
	ϵ	The smallest unit of currency

The prior probability of data exfiltration α is assumed to be common knowledge to attacker and victim. This means that in stage 3 of the game the victim can form a belief on the probability that data was exfiltrated. This belief will be based on prior belief α together with the observed action of the criminal in stage 2 to signal or not (along with the ransom demand). Let μ denote the updated belief of the victim. The value of μ will be determined. The prior probability the victim is high type β is also assumed to be common knowledge to attacker and victim.

The variables of the game are summarized in Table 8.1. One additional variable we introduce is $L \geq 0$ which captures the legal fees and costs (including psychological and moral) of paying a ransom. We also introduce variable ϵ to represent the smallest unit of currency. This will allow us to characterise the optimal ransom in a more succinct way. We exclude from the analysis any fixed costs incurred by the attacker and victim that are not dependent on the strategic elements of the game. For instance we do not include the cost to the attacker of implementing the attack. We can exclude such costs, without loss of generality, because they will not influence the equilibrium outcomes of the game. We depict the game in Figure 8.1.



Case T_1 (Prob. β): Important files exfiltrated.



Case T_0 (Prob. 1-β): No important files exfiltrated.

Figure 8.1: Description of the game.

8.3 Results

In the following we solve for Bayesian equilibria of the game [10]. Informally, a Bayesian equilibrium has the property that both attacker and victim: (1) maximise their expected payoffs given their beliefs and the strategy of the other, (2) update their beliefs using Bayes rule. Thus, in equilibrium, players appropriately interpret information, and have no incentive to change their actions given their beliefs and the actions of the other player. It is standard to consider Bayesian equilibria as a benchmark solution concept in signalling games to capture and analyse the incentives of players [15].

We focus on Bayesian equilibria that satisfy the, so called, D1 Criterion [10]. To briefly explain the motivation for this refinement, we remark that if players act consistent with a Bayesian equilibrium then there may be nodes with zero probability of being reached. A Bayesian equilibrium does not tie down beliefs at such nodes because Bayes rule is indeterminate. The D1 Criterion is used to place 'common sense' restrictions on beliefs. Specifically, The D1 Criterion imposes extra conditions on beliefs by saying that any deviation from the equilibrium path is assumed to be done by the type with the most incentive to deviate [3].

The D1 Criterion is useful to rule out equilibria sustained by 'non-intuitive beliefs' [15]. For instance, consider a candidate equilibria in which the attacker chooses to not signal if they are type DE or NDE. On the equilibrium path the attacker should not signal. Thus, Bayes rule does not impose any restrictions on beliefs if the attacker does signal. Yet, informally, as we shall below, a type DE has the most incentive to deviate and signal. The D1 Criterion would, thus, require the victim to believe the deviation was by a type DE. This rules out 'non-intuitive' equilibria that are only sustained by the victim believing a signal of data exfiltration must indicate that data was not exfiltrated.

To focus the analysis on what we believe are the most realistic cases, we distinguish and characterize three broad types of equilibrium: (a) separating equilibria in which the type DE signals data is exfiltrated and the type NDE does not, (b) a pooling equilibria in which both the type DE and NDE signal that data is exfiltrated, and (c) a pooling equilibria in which both the type DE and NDE do not signal that data is exfiltrated. We exclude from analysis hybrid equilibria in which the attacker randomises their actions. In the following we discuss separating and pooling equilibria in turn before analysing the impact of private information. Throughout, we assume that if the victim is indifferent between paying and not paying then they will not pay.

8.3.1 Separating Equilibrium

A separating equilibrium has the basic characteristic that the attacker signals data exfiltration if they are of type DE (i.e. data was exfiltrated) and does not signal if they are of type NDE (i.e. data was not exfiltrated). The existence of a separating equilibrium and the exact form of any equilibrium will depend on the parameters of the game. In total, we identified four types of separating equilibria that can exist, which we will label A1-A4. These are summarised in Table 8.2. As you can see the equilibria differ by whether or not the victim pays the ransom. In equilibrium A1 the victim pays irrespective of their type and whether the attacker signals. In equilibrium A2 the victim pays unless they are the low type and the attacker signals. In equilibrium A3 the victim pays the ransom if the attacker signals but does not pay the ransom if the attacker does not signal. In equilibrium A4 the victim only pays if they are a high type and the attacker signals.

Equilibrium	Attacker		Victim			
	DE	NDE	T_1		T_0	
			Signal	No signal	Signal	No signal
A1	Signal	No signal	Pay	Pay	Pay	Pay
A2	Signal	No signal	Pay	Pay	No pay	Pay
A3	Signal	No signal	Pay	No Pay	Pay	No Pay
A4	Signal	No Signal	Pay	No Pay	No pay	No pay
B1	Signal	Signal	Pay	Pay	Pay	Pay
B2	Signal	Signal	Pay	Pay	No pay	Pay
B3	Signal	Signal	Pay	No pay	Pay	No pay
B4	Signal	Signal	Pay	No pay	No pay	No pay
C1	No signal	No signal	Pay	Pay	Pay	Pay
C2	No signal	No signal	Pay	Pay	No Pay	Pay
C3	No signal	No signal	Pay	Pay	Pay	No Pay
C4	No signal	No signal	Pay	Pay	No Pay	No Pay

Table 8.2: Equilibria satisfying the D1 criterion in the signaling game.

In all four equilibria A1-A4 the high type victim pays if they receive a signal of data exfiltration. The equilibria differ in whether a low type victim pays if they receive a signal of data exfiltration and/or whether the victim (high or low type) pays if they receive no signal. To provide some intuition for the four equilibria we identify three ransom demands that prove particularly relevant:

$$R^*_{S0} = T_0 + V - L - \epsilon; R^*_{S1} = T_1 + V - L - \epsilon; R^*_{NS} = \max\{V - L - \epsilon, 0\}. \quad (8.1)$$

Informally, see the proof of Theorem 8.3.1 for the full details, R^*_{S0} and R^*_{S1} are the maximum ransom the low type and high type, respectively, are willing to pay if they believe data has been exfiltrated. While, R^*_{NS} is the maximum ransom the victim is willing to pay if they believe data has not been exfiltrated. We readily see that if $V \leq L$ the victim would not pay any positive ransom demand if they know data has not been exfiltrated.

If data exfiltration is believed to have taken place then the high type is willing to pay a larger ransom than the low type, $R^*_{S1} > R^*_{S0}$. This provides a strategic trade-off for the attacker: (a) if they ask for a high ransom, R^*_{S1}, then they extract maximum surplus from the high type victim, but the low type will not pay the ransom. (b) If they ask for a low ransom, R^*_{S0}, then both the low and high type victim will pay the ransom but they do not fully extract surplus from the high type. This trade-off between asking a high or low ransom can be captured by the following term:

$$\Phi_S = \beta(R^*_{S1} - R^*_{S0}) - (1-\beta)R^*_{S0} = \beta(T_1 - T_0) - (1-\beta)(T_0 + V - L - \epsilon). \quad (8.2)$$

The first term in Φ_S is the expected gain for the attacker from charging a high ransom and extracting maximum surplus from the high type, while the second term is the expected loss from charging a ransom the low type is not willing to pay.

We are now in a position to state our first main result. As the preceding discussion preempts we need to consider combinations of $V \gtrless L$ and $\Phi \gtrless 0$ giving rise to the four different cases and equilibria.

There exists a separating equilibrium satisfying the D1 criterion if and only if the following conditions hold:

(A1) If $L < V$ and $\Phi_S < 0$ then $k_D < T_0 < k_N$.

(A2) If $L < V$ and $\Phi_S > 0$ then $k_D < \beta T_1 - (1-\beta)(V - L) < k_N$.

(A3) If $L > V$ and $\Phi_S < 0$ then $k_D < T_0 + V - L < k_N$.

(A4) If $L > V$ and $\Phi_S > 0$ then $k_D < \beta(T_1 + V - L) < k_N$.

Proof. We first consider the strategy of the victim. Suppose the attacker sends a signal and ransom demand R_S. Suppose the victim infers the attacker is type DE. In other words, $\mu = 1$. If the victim is low type and pays the ransom their expected payoff is $-R_S - L$. Their expected payoff if they do not pay is $-T_0 - V$. It follows the low type victim will optimally pay the ransom if and only if $-R_S - L > -T_0 - V$ or equivalently $R_S < T_0 + V - L$. They would, therefore, pay ransom R^*_{S0}. If the victim is high type and pays the ransom their expected payoff is $-R_S - L$. Their expected payoff if they do not pay is $-T_1 - V$. It follows the high type victim will optimally pay the ransom if and only if $R_S < T_1 + V - L$. They would, therefore, pay ransom R^*_{S1}. Given that $T_1 > T_0$ we also have that the high type would pay ransom R^*_{S0}.

Now suppose the attacker does not send a signal and sets ransom demand R_{NS}. Suppose the victim infers the attacker is type NDE. In other words, $\mu = 0$. If the victim is low type and pays the ransom their expected payoff is $-R_{NS} - L$. Their expected payoff if they do not pay is $-V$. It follows the low type victim will optimally pay the ransom if and only if $-R_{NS} - L > -V$ or equivalently $R_{NS} < V - L$. They would, therefore, pay ransom R^*_{NS} if $V > L$ and not pay if $V < L$. The same logic holds if the victim is high type.

We now consider the incentives of the attacker. Suppose the attacker is type DE. Also suppose that on the equilibrium path they signal and set ransom R^*_{S0}. Their expected payoff in equilibrium is $\pi(S, R^*_{S0}) = T_0 + V - L - \epsilon - k_D$. In exploring incentives to deviate from the equilibrium path, we first consider the possibility the attacker signals but sets a different ransom demand $R_S \neq R^*_{S0}$. If $R_S < R^*_{S0}$ then the expected payoff of the attacker is $\pi(S, R_S) = R_S - k_D < \pi(S, R^*_{S0})$ and so the attacker receives a lower payoff than on the equilibrium path. If $R^*_{S1} > R_S > R^*_{S0}$ (and $\mu = 1$) then the high type victim would pay the ransom but the low type victim would not. The expected payoff of the attacker is, therefore, $\pi(S, R_S) = \beta R_S - k_D \leq \beta R^*_{S1} - k_D$. It follows the attacker prefers the equilibrium path if and only if $\pi(S, R^*_{S1}) \leq \pi(S, R^*_{S0})$ or, equivalently, $\beta(T_1 + V - L - \epsilon) \leq T_0 + V - L - \epsilon$. Rearranging gives the condition on $\Phi_S < 0$. Reversing this argument we can say it is on the equilibrium path for the attacker of type DE to signal and set ransom R^*_{S1} if and only if $\Phi_S > 0$.

We next consider the possibility that an attacker of type DE chooses to not signal. Suppose they set ransom demand R_{NS} (and are inferred to be type NDE). Their expected payoff is at most $\pi(NS, R_{NS}) = R^*_{NS}$. We then have four different cases to consider. (a) Suppose $V > L$ and $R^*_S = R^*_{S0}$. It follows the attacker prefers the equilibrium path if and only if $V - L - \epsilon < T_0 + V - L - \epsilon - k_D$ or,

equivalently, $k_D < T_0$. (b) Suppose $V > L$ and $R_S^* = R_{S1}^*$. It follows the attacker prefers the equilibrium path if and only if $V - L - \epsilon < \beta(T_1 + V - L - \epsilon) - k_D$ or, equivalently, $k_D + (1 - \beta)(V - L - \epsilon) < \beta T_1$. (c) Suppose $V < L$ and $R_S^* = R_{S0}^*$. It follows the attacker prefers the equilibrium path if and only if $0 < T_0 + V - L - \epsilon - k_D$ or, equivalently, $k_D < T_0 + V - L$. (d) Suppose $V < L$ and $R_S^* = R_{S1}^*$. It follows the attacker prefers the equilibrium path if and only if $0 < \beta(T_1 + V - L - \epsilon) - k_D$ or, equivalently, $k_D < \beta(T_1 + V - L - \epsilon)$.

Next suppose the attacker is type NDE. Extending the logic of the preceding discussion there is no incentive for the attacker to choose a ransom other than R_{NS}^*. We focus, therefore, on the incentive to signal and choose ransom demand R_S^*. We again have four different cases to consider. (a) Suppose $V > L$ and $R_S^* = R_{S0}^*$. On the equilibrium path the attacker has expected payoff $\pi(NS, R_{NS}^*) = V - L - \epsilon$. It follows the attacker prefers the equilibrium path if and only if $V - L - \epsilon > T_0 + V - L - \epsilon - k_N$ or, equivalently, $k_N > T_0$. (b) Suppose $V > L$ and $R_S^* = R_{S1}^*$. It follows the attacker prefers the equilibrium path if and only if $V - L - \epsilon > \beta(T_1 + V - L - \epsilon) - k_N$ or, equivalently, $k_N + (1 - \beta)(V - L - \epsilon) > \beta T_1$. (c) Suppose $V < L$ and $R_S^* = R_{S0}^*$. It follows the attacker prefers the equilibrium path if and only if $0 > T_0 + V - L - \epsilon - k_N$ or, equivalently, $k_N > T_0 + V - L$. (d) Suppose $V < L$ and $R_S^* = R_{S1}^*$. It follows the attacker prefers the equilibrium path if and only if $0 > \beta(T_1 + V - L - \epsilon) - k_N$ or, equivalently, $k_N > \beta(T_1 + V - L - \epsilon)$.

It remains to check the D1 criterion is satisfied. The only game path we need to consider in any detail is that where the attacker does not signal and sets ransom $R_{NS} \neq R_{NS}^*$. We have assumed the victim will infer the attacker is type NDE. Given that $K_N > k_D$, the attacker has most incentive to not signal when of type NDE. This assumption, therefore, naturally satisfies the D1 criterion.□ ■

In interpretation of Theorem 8.3.1 we can see that there exists a separating equilibrium if and only if k_D is sufficiently small and k_N is sufficiently large. In other words, a separating equilibrium exists if it is 'cheap' for the attacker to signal when they have exfiltrated data and 'expensive' for the attacker to signal if they have not exfiltrated data. This would imply, for instance, that if victims have invested in good monitoring systems to identify data exfiltration, they could make it harder for the attacker of type NDE to send a credible signal; then, k_N would increase and we would expect the improved monitoring to result in a separating equilibrium. We explore these issues in mode detail after analysing pooling equilibria.

8.3.2 Pooling Equilibrium with Signal

We turn our attention now to pooling equilibria. We focus first on pooling equilibrium in which the attacker signals. That is, the attacker signals that data is exfiltrated whether they are type NDE or DE. Given that the attacker will signal irrespective of type, a signal does not convey any useful information to the victim on whether or not data has been exfiltrated. We identify four types of such pooling equilibria, which we will label B1-B4. These are summarised in Table 8.2. Equilibria B1-B4 (like A1-A4) differ in terms of whether the victim will pay.

Two ransom demands that we identified as being particularly relevant in determining pooling equilibria are:

$$R^*_{P0} = \alpha T_0 + V - L - \epsilon; \tag{8.3}$$
$$R^*_{P1} = \alpha T_1 + V - L - \epsilon, \tag{8.4}$$

Informally, R^*_{P0} and R^*_{P1} are the maximum ransom the low and high type, respectively, are willing to pay if they believe the attacker has exfiltrated data with probability α.

As with the separating equilibrium, the optimal ransom demand of the attacker involves a trade-off between setting a high ransom R^*_{P1} that only the high type will pay and a low ransom R^*_{P0} that both the high and low type will pay. This trade-off is captured by the term:

$$\Phi_P = \beta\alpha(T_1 - T_0) - (1 - \beta)(\alpha T_0 + V - L - \epsilon). \tag{8.5}$$

We can now state our second result.

There exists a pooling equilibrium in which the attacker signals, satisfying the D1 criterion, if and only if the following conditions hold:

(B1) If $L < V$ and $\Phi_P < 0$ then $k_N < \alpha T_0$.

(B2) If $L < V$ and $\Phi_P > 0$ then $k_N < \beta\alpha T_1 - (1 - \beta)(V - L)$.

(B3) If $L > V$ and $\Phi_P < 0$ then $k_N < \alpha T_0 + V - L$.

(B4) If $L > V$ and $\Phi_P > 0$ then $k_N < \beta(\alpha T_1 + V - L)$.

Proof. Consider the strategy of the victim. Suppose the attacker sends a signal and ransom demand R_S. Suppose the victim infers the attacker is type DE with probability $\mu = \alpha$. If the victim is low type and pays the ransom their expected payoff is $-R_S - L$. Their expected payoff if they do not pay is $-\alpha T_0 - V$. It

follows the low type victim will optimally pay the ransom if and only if $-R_S - L > -\alpha T_0 - V$ or equivalently $R_S < \alpha T_0 + V - L$. They would, therefore, pay ransom R^*_{P0}. If the victim is high type and pays the ransom their expected payoff is $-R_S - L$. Their expected payoff if they do not pay is $-\alpha T_1 - V$. It follows the high type victim will optimally pay the ransom if and only if $R_S < \alpha T_1 + V - L$. They would, therefore, pay ransom R^*_{P1}. Given that $T_1 > T_0$ we also have that the high type would pay ransom R^*_{P0}.

Now suppose the attacker does not send a signal and sets ransom demand R_{NS}. Suppose the victim infers the attacker is type NDE. In other words, $\mu = 0$. If the victim is low type and pays the ransom their expected payoff is $-R_{NS} - L$. Their expected payoff if they do not pay is $-V$. It follows the low type victim will optimally pay the ransom if and only if $-R_{NS} - L > -V$ or equivalently $R_{NS} < V - L$. They would, therefore, pay ransom R^*_{NS} if $V > L$ and not pay if $V < L$. The same logic holds if the victim is high type.

Next consider the incentives of the attacker. Suppose the attacker is type DE. Also suppose that on the equilibrium path they signal and set ransom R^*_{P0}. Their expected payoff in equilibrium is $\pi(S, R^*_{P0}) = \alpha T_0 + V - L - \epsilon - k_D$. Suppose the attacker signals but sets a different ransom demand $R_S \neq R^*_{P0}$. If $R_S < R^*_{P0}$ then the expected payoff of the attacker is $\pi(S, R_S) = R_S - k_D < \pi(S, R^*_{P0})$ and so the attacker receives a lower payoff than on the equilibrium path. If $R^*_{P1} > R_S > R^*_{P0}$ (and $\mu = \alpha$) then the high type victim would pay the ransom but the low type victim would not. The expected payoff of the attacker is, therefore, $\pi(S, R_S) = \beta R_S - k_D \leq \beta R^*_{P1} - k_D$. It follows the attacker prefers the equilibrium path if and only if $\beta(\alpha T_1 + V - L - \epsilon) \leq \alpha T_0 + V - L - \epsilon$. Rearranging gives $\Phi_P < 0$. Reversing this argument we can say it is on the equilibrium path for the attacker of type DE to signal and set ransom R^*_{P1} if and only if $\Phi_P > 0$.

Now consider the possibility that an attacker of type NDE chooses to not signal. Suppose they set ransom demand R_{NS} (and are inferred to be type NDE). Their expected payoff is at most $\pi(NS, R_{NS}) = R^*_{NS}$. We then have four different cases to consider. (a) Suppose $V > L$ and $R^*_S = R^*_{P0}$. It follows the attacker prefers the equilibrium path if and only if $V - L - \epsilon < \alpha T_0 + V - L - \epsilon - k_N$ or, equivalently, $k_N < \alpha T_0$. (b) Suppose $V > L$ and $R^*_S = R^*_{P1}$. It follows the attacker prefers the equilibrium path if and only if $V - L - \epsilon < \beta(\alpha T_1 + V - L - \epsilon) - k_N$ or, equivalently, $k_N + (1 - \beta)(V - L - \epsilon) < \beta \alpha T_1$. (c) Suppose $V < L$ and $R^*_S = R^*_{P0}$. It follows the attacker prefers the equilibrium path if and only if $0 < \alpha T_0 + V - L - \epsilon - k_N$ or, equivalently, $k_N < \alpha T_0 + V - L$. (d) Suppose $V < L$ and $R^*_S = R^*_{P1}$. It follows the attacker prefers the equilibrium path if and only if $0 < \beta(\alpha T_1 + V - L - \epsilon) - k_N$ or, equivalently, $k_N < \beta(\alpha T_1 + V - L - \epsilon)$.

One can show, using $k_D < k_N$, that the analogous conditions for a type DE to prefer signalling to not signalling are less binding.

It remains to check the D1 criterion is satisfied. The only game path we need to consider in any detail is that where the attacker does not signal and sets ransom $R_{NS} \neq R^*_{NS}$. We have assumed the victim will infer the attacker is type NDE. Given that $K_N > k_D$, the attacker has most incentive to not signal when of type NDE. This assumption, therefore, naturally satisfies the D1 criterion. □ ■

In interpretation of Theorem 8.3.2 there exists a pooling equilibrium with signalling if and only if k_N is sufficiently small. In other words, there exists a pooling equilibrium with signalling if and only if it is cheap for the attacker to signal even if data has not been exfiltrated. In practical terms this would suggest, for instance, a pooling equilibrium will exist if the victim does not have any monitoring capabilities to identify or evaluate a data breach. It would also be the case if the criminals can easily extract some information, e.g. file tree or sample file, that would allow them to signal data exfiltration even though data was not exfiltrated.

8.3.3 Pooling Equilibrium with No Signal

We now focus on pooling equilibria in which the attacker does not signal. That is, the attacker chooses to not signal that data is exfiltrated whether they are type NDE or DE. Given that the attacker does not signal, irrespective of type, the lack of signal does not convey any useful information to the victim on whether or not data has been exfiltrated. We identify four types of such pooling equilibria, which we will label C1-C4. These are summarised in Table 8.2 and again differ in terms of whether the victim will pay. We see that in all of the equilibria C1-C4 the high type pays whether there is a signal or not. The equilibria differ in whether the low type will pay.

In stating our third result we remark that all of the ransom demands previously identified, $R^*_{S0}, R^*_{S1}, R^*_{P0}, R^*_{P1}$, and the values of Φ_S and Φ_P prove relevant. To help navigate the statement of the theorem we note that

$$\Phi_S - \Phi_P = (1-\alpha)(\beta T_1 - T_0). \tag{8.6}$$

Thus, it can be the case that $\Phi_S > \Phi_P$ or vice versa. We also remark that it is possible to simultaneously have $\Phi_P > 0$ and $\Phi_S < 0$ (when $V < L$) and $\Phi_P < 0$ and $\Phi_S > 0$ (when $V > L$) (see the proof for more details). Equilibria C1-C4 largely depend on different combinations of whether Φ_S and Φ_P are positive or negative. We can now state our third result.

There exists a pooling equilibrium in which the attacker does not signal, satisfying the D1 criterion, if and only if the following conditions hold:

(C1) If $\Phi_P < 0$ and $\Phi_S < 0$ then $(1-\alpha)T_0 < k_D$.

(C2) If $\Phi_P < 0$ and $\Phi_S > 0$ then $\beta T_1 - \alpha T_0 - (1-\beta)(V - L - \epsilon) < k_D$.

(C3) If $\Phi_P > 0$ and $\Phi_S < 0$ then $T_0 - \beta\alpha T_1 + (1-\beta)(V - L - \epsilon) < k_D$.

(C4) If $\Phi_P > 0$ and $\Phi_S > 0$ then $\beta(1-\alpha)T_1 < k_D$.

Proof. Consider the strategy of the victim. Suppose the attacker does not send a signal and sets ransom demand R_{NS}. Suppose the victim infers the attacker is type DE with probability $\mu = \alpha$. If the victim is low type and pays the ransom their expected payoff is $-R_{NS} - L$. Their expected payoff if they do not pay is $-\alpha T_0 - V$. It follows the low type victim will optimally pay the ransom if and only if $-R_{NS} - L > -\alpha T_0 - V$ or equivalently $R_{NS} < \alpha T_0 + V - L$. They would, therefore, pay ransom R^*_{P0}. If the victim is high type and pays the ransom their expected payoff is $-R_{NS} - L$. Their expected payoff if they do not pay is $-\alpha T_1 - V$. It follows the high type victim will optimally pay the ransom if and only if $R_{NS} < \alpha T_1 + V - L$. They would, therefore, pay ransom R^*_{P1}. Given that $T_1 > T_0$ we also have that the high type would pay ransom R^*_{P0}.

Now suppose the attacker signals and sets ransom demand R_S. Suppose the victim infers the attacker is type DE. In other words, $\mu = 1$. If the victim is low type and pays the ransom their expected payoff is $-R_S - L$. Their expected payoff if they do not pay is $-T_0 - V$. It follows the low type victim will optimally pay the ransom if and only if $-R_S - L > -T_0 - V$ or equivalently $R_S < T_0 + V - L$. They would, therefore, pay a positive ransom R_S if $T_0 + V > L$ and not pay if $T_0 + V < L$. Similarly, the high type would pay ransom R_S if $T_1 + V > L$. We recall that $T_1 + V > L$ by assumption.

Next consider the incentives of the attacker. Suppose the attacker is type DE. Also suppose that on the equilibrium path they do not signal and set ransom R^*_{P0}. Their expected payoff in equilibrium is $\pi(NS, R^*_{P0}) = \alpha T_0 + V - L - \epsilon$. Suppose the attacker does not signal but sets a different ransom demand $R_{NS} \neq R^*_{P0}$. If $R_{NS} < R^*_{P0}$ then the expected payoff of the attacker is $\pi(NS, R_{NS}) = R_{NS} < \pi(NS, R^*_{P0})$ and so the attacker receives a lower payoff than on the equilibrium path. If $R^*_{P1} > R_{NS} > R^*_{P0}$ (and $\mu = \alpha$) then the high type victim would pay the ransom but the low type victim would not. The expected payoff of the attacker is, therefore, $\pi(NS, R_{NS}) = \beta R_{NS} \leq \beta R^*_{P1}$. It follows the attacker prefers the equilibrium path if and only if $\beta(\alpha T_1 + V - L - \epsilon) \leq \alpha T_0 + V - L - \epsilon$. Rearranging gives $\Phi_P < 0$. Reversing this argument we can say it is on the equilibrium path for the attacker of type DE to not signal and set ransom R^*_{P1} if and only if $\Phi_P > 0$.

Now consider the possibility that an attacker of type DE chooses to signal. Suppose they set ransom demand R_S (and are inferred to be type DE). We have

several different cases to consider. Before doing so we consider the relationship between Φ_S and Φ_P. Rearranging equation 8.2 we see that $\Phi_S < 0$ if and only if

$$\beta < \frac{T_0 + V - L - \epsilon}{T_1 + V - L - \epsilon}. \tag{8.7}$$

Rearranging equation 8.5 we see that $\Phi_P > 0$ if and only if

$$\beta > \frac{\alpha T_0 + V - L - \epsilon}{\alpha T_1 + V - L - \epsilon}. \tag{8.8}$$

To obtain $\Phi_P > 0$ and $\Phi_S < 0$ we, therefore, would require

$$\frac{\alpha T_0 + K}{\alpha T_1 + K} < \frac{T_0 + K}{T_1 + K} \tag{8.9}$$

where $K = V - L - \epsilon$. This simplifies to $KT_1 < KT_0$ which is possible if $V < L$. Similarly, $\Phi_P < 0$ and $\Phi_S > 0$ is only possible if $V > L$.

(a) Suppose $T_0 + V > L$, $\Phi_P < 0$ and $\Phi_S < 0$. Given that $\Phi_P < 0$ we know $R^*_{NS} = R^*_{P0}$. Also, given that $\Phi_S < 0$ we know that, if the attacker signals, they would maximize their payoff by setting ransom R^*_{S0} (see the Proof of Theorem 1). It follows the attacker prefers the equilibrium path if and only if $T_0 + V - L - \epsilon - k_D < \alpha T_0 + V - L - \epsilon$ or, equivalently, $T_0(1 - \alpha) < k_D$.

(b) Suppose $T_0 + V > L$, $\Phi_P < 0$ and $\Phi_S > 0$. Given that $\Phi_S > 0$ we know that, if the attacker signals, they would maximize their payoff by setting ransom R^*_{S1}. It follows the attacker prefers the equilibrium path if and only if $\beta(T_1 + V - L - \epsilon) - k_D < \alpha T_0 + V - L - \epsilon$ or, equivalently, $\beta T_1 - \alpha T_0 - (1-\beta)(V - L - \epsilon) < k_D$.

(c) Suppose $T_0 + V > L$, $\Phi_P > 0$ and $\Phi_S < 0$. Given that $\Phi_P > 0$ we know $R^*_{NS} = R^*_{P1}$. Also, given that $\Phi_S < 0$ we know that, if the attacker signals, they would maximize their payoff by setting ransom R^*_{S0}. It follows the attacker prefers the equilibrium path if and only if $T_0 + V - L - \epsilon - k_D < \beta(\alpha T_1 + V - L - \epsilon)$ or, equivalently, $T_0 - \beta\alpha T_1 + (1 - \beta)(V - L - \epsilon) < k_D$.

(d) Suppose $T_0 + V > L$, $\Phi_P > 0$ and $\Phi_S > 0$. Given that $\Phi_P > 0$ we know $R^*_{NS} = R^*_{P1}$. Also, given that $\Phi_S > 0$ we know that, if the attacker signals, they would maximize their payoff by setting ransom R^*_{S1}. It follows the attacker prefers the equilibrium path if and only if $\beta(T_1 + V - L - \epsilon) - k_D < \beta(\alpha T_1 + V - L - \epsilon)$ or, equivalently, $\beta(1 - \alpha)T_1 < k_D$.

(e) If $L > T_0 + V$ then $\Phi_P > 0$ and $\Phi_S > 0$. Thus, $R^*_{NS} = R^*_{P1}$ and, if the attacker signals, they would maximize their payoff by setting ransom R^*_{S1}. It follows the attacker prefers the equilibrium path if and only if $\beta(T_1 + V - L - \epsilon) - k_D < \beta(\alpha T_1 + V - L - \epsilon)$ or, equivalently, $\beta(1 - \alpha)T_1 < k_D$.

To derive the conditions in C1-C4 stated in the Theorem we note that if $\Phi_S < 0$ then it must be the case that $L < T_0 + V$. Similarly, if $\Phi_S < 0$ then it must be the case that $L < \alpha T_0 + V < T_0 + V$.

It remains to check the D1 criterion is satisfied. The only game path we need to consider in any detail is that where the attacker signals. We have assumed the victim will infer the attacker is type DE. Given that $K_N > k_D$, the attacker has most incentive to signal when of type DE. This assumption, therefore, naturally satisfies the D1 criterion. □ ■

In interpretation of Theorem 8.3.2 there exists a pooling equilibrium with no signalling if and only if k_D is sufficiently large. In other words, there exists a pooling equilibrium with no signalling if and only if it is expensive for the attacker to signal even if data has been exfiltrated. In practical terms this would suggest, for instance, a pooling equilibrium will exist if the victim requires detailed evidence of data exfiltration that would require the criminal to analyse the data in more detail. Or it could be the case that the process of signalling exfiltration, for example communicating with the victim, is costly in terms of time and opportunity cost.

8.3.4 Equilibrium Existence

Depending on the parameters of the game there may exist a separating equilibrium, a pooling equilibrium, both, or neither. To illustrate, consider the parameters $L = 0, V = 5, \alpha = 0.9, \beta = 0.5, T_0 = 1$ and $T_1 = 5$. Then $\Phi_S < 0$ and so there exists a separating equilibrium if and only if $k_D < 1 < k_N$. Also $\Phi_P < 0$ and so there exists a pooling equilibrium with signalling if $k_N < 0.9$. Thus, for $k_N < 0.9$ there is a pooling equilibrium with signalling, for $0.9 < k_N < 1$ there is neither a separating nor pooling equilibrium with signalling, and for $1 < k_N$ there is a separating equilibrium. The relative size of the cost for the attacker to signal data exifltration when they have not exfiltrated data is, thus, crucial to determining the equilibrium outcome.

We remind that the existence of a pooling equilibrium with signalling relies on K_N being sufficiently small while the existence of a pooling equilibrium with no signalling relies on K_D being sufficiently large. Given that $K_D < K_N$ it is generally not the case that there can exist both a pooling equilibrium with signalling and one without. There are, however, parameter values where this is possible. For instance, with the parameters introduced above there is a pooling equilibrium with no signalling if $0.1 < k_D$. Thus, if $0.1 < k_D < k_N < 0.9$ there exists both a pooling equilibrium with signalling and a pooling equilibrium with no signalling.

The existence of multiple equilibrium can capture different norms or historical precedent of the ransomware environment. Consider, for instance, a setting in which ransomware criminals never signal data exfiltration. Does an attacker who has exfiltrated data have an incentive to deviate and signal exfiltration? If data exfiltration is suspected without a signal ($\alpha = 0.9$) then the attacker can ask a relatively high ransom without signalling. The extra ransom that can be asked if data exfiltration is signalled may not, therefore, be enough to cover the costs of data exfiltration (k_D). Thus, it is an equilibrium to not signal.

Now consider the same parameters but a setting in which all ransomware criminals signal data exfiltration. Does an attacker who has not exfiltrated data have an incentive to not signal and save on the cost of signalling? If data exfiltration is suspected with a signal ($\alpha = 0.9$) then the attacker can extract a relatively high ransom if they signal (even though data is not exfiltrated). The loss in revenue from not signalling may, therefore, be more than the saving in signaling cost (k_N). Thus, it is an equilibrium to signal. In a setting with multiple equilibria, historical precedent and learning dynamics may determine which equilibrium (signal or not) is prevalent at the time [37].

8.3.5 Expected Equilibrium Payoffs

A key objective of our work is to analyse the payoff consequences, for both victim and attacker, of private information on the side of the victim. In Table 8.3 we detail the expected payoff of the attacker and victim in equilibria A1-A4, B1-B4 and C1-C4. These are ex-ante expected payoffs before own type is known. For instance, in equilibrium A1 there is probability α the attacker is type DE and obtains payoff $R^*_{S0} - k_D$ and probability $1-\alpha$ the attacker is type NDE and obtains payoff R^*_{NS}. The expected payoff is, therefore, $\alpha(R^*_{S0} - k_D) + (1-\alpha)R^*_{NS}$ Given that ϵ can be arbitrarily small we have omitted it from calculations of expected payoff.

In interpreting the payoffs in Table 8.3 it is important to keep in mind equilibrium existence. For instance, care is needed in saying payoffs are, say, higher in equilibrium C1 than B1 or A1 because these respective equilibria may exist for different parameter values. Our analysis will take this into account. We can, however, say at a broader level that the attacker's payoff, everything else the same, is highest in the pooling equilibria with no signalling (C1-C4). The intuition being that the attacker does not incur any costs of signaling. From a policy perspective, to deter ransomware it would, therefore, be beneficial to move away from a pooling equilibria with no signalling (C1-C4) to either a separating equilibrium (A1-A4) or a pooling equilibrium with signaling. As discussed in the

Table 8.3: Expected payoff of attacker and victim in equilibrium.

Equilibrium	Attacker	Victim
A1	$\alpha T_0 + V - L - \alpha k_D$	$-\alpha T_0 - V$
A2	$\alpha(\beta(T_1 + V - L) - k_D) + (1-\alpha)(V - L)$	$-\alpha(\beta T_1 + (1-\beta)T_0) - V$
A3	$\alpha(T_0 + V - L - k_D)$	$-\alpha T_0 - V$
A4	$\alpha(\beta(T_1 + V - L) - k_D)$	$-\alpha(\beta T_1 + (1-\beta)T_0) - V$
B1 & B3	$\alpha T_0 + V - L - \alpha k_D - (1-\alpha)k_N$	$-\alpha T_0 - V$
B2 & B4	$\beta(\alpha T_1 + V - L) - \alpha k_D - (1-\alpha)k_N$	$-\alpha(\beta T_1 + (1-\beta)T_0) - V$
C1 & C2	$\alpha T_0 + V - L$	$-\alpha T_0 - V$
C3 & C4	$\beta(\alpha T_1 + V - L)$	$-\alpha(\beta T_1 + (1-\beta)T_0) - V$

previous sub-section this may involve changing the norms of the ransomware environment.

Another policy insight that we can take from Table 8.3 is the importance of pre-empting a ransomware attack. In particular, pre-emption and appropriation preparedness for an attack can lower the recovery costs of an attack V, the reputational damage T_1 and T_0, and potentially decrease the probability of being a high type β and reduce the probability of data exfiltration α. All of these would reduce the losses of the victim in the event of a breach. This shows up very clearly in our model because the attacker is able to extract maximum surplus from the victim.

To analyse the consequences of private information we need to consider an alternative game in which the attacker knows the type of the victim and so knows if the reputational damage that would result from data publication is T_0 or T_1. We can apply Theorems 8.3.1, 8.3.2 and 8.3.3 to distinguish the conditions under which there exist separating and pooling equilibirum in this revised game. Specifically, by setting $\beta = 0$ or 1 we derive the following corollaries. If the victim is known to be type $i = \{0, 1\}$ there exists a separating equilibrium satisfying the D1 criterion if and only if the following conditions hold:

A1$A2$. If $L < V$, then $k_D < T_i < k_N$.

A3A4. If $L > V$, then $k_D < T_i + V - L < k_N$.

Proof. Suppose $\beta = 0$. Then $\Phi_S < 0$. Applying Theorem 8.3.1 we obtain conditions: (A1) $L < V$ and $k_D < T_0 < k_N$, and (A3) $L > V$ and $k_D < T_0 + V - L < k_N$. Suppose $\beta = 1$. Then $\Phi_S > 0$. Applying Theorem 8.3.1 we obtain conditions: (A2) $L < V$ and $k_D < T_1 < k_N$, and (A4) $L > V$ and

$k_D < T_1 + V - L < k_N$. □ ■ If the victim is known to be type $i = \{0, 1\}$ there exists a pooling equilibrium with a signal satisfying the D1 criterion if and only if the following conditions hold:

B1$B2$. If $L < V$ then $k_N < \alpha T_i$.

B3B4. If $L > V$ then $k_N < \alpha T_i + V - L$.

Proof. Suppose $\beta = 0$. Then $\Phi_P < 0$. Applying Theorem 8.3.2 we obtain conditions: (B1) $L < V$ and $k_N < \alpha T_0$, and (B3) $L > V$ and $k_N < \alpha T_0 + V - L$. Suppose $\beta = 1$. Then $\Phi_P > 0$. Applying Theorem 8.3.2 we obtain conditions: (B2) $L < V$ and $k_N < \alpha T_1$, and (B4) $L > V$ and $k_N < \alpha T_1 + V - L$. □ ■ If the victim is known to be type $i = \{0, 1\}$ there exists a pooling equilibrium with no signal satisfying the D1 criterion if and only if the following conditions hold:

C1C4. If $(1 - \alpha)T_i < k_D$.

Proof. Suppose $\beta = 0$. Then $\Phi_S < 0$ and $\Phi_P < 0$. Applying Theorem 8.3.3 we obtain condition (C1) $(1 - \alpha)T_0 < k_D$. Suppose $\beta = 1$. Then $\Phi_S > 0$ and $\Phi_P > 0$. Applying Theorem 8.3.3 we obtain condition: (C4) $(1 - \alpha)T_1 < k_D$. □ ■

With these three corollaries we can derive the expected payoff of the attacker and victim in a game where the victim's type is known. Table 8.4 details the relevant payoffs. For instance, the expected payoff of the attacker under equilibrium A3A4 if the victim is type 0 is $\alpha(T_0 + V - L - k_D)$ and the expected payoff of the attacker under equilibrium A3A4 if the victim is type 1 is $\alpha(T_1 + V - L - k_D)$. Some care is needed in deriving ex-ante expected payoffs because the existence of equilibrium A3A4 for the low type does not guarantee existence of equilibrium A3A4 for the high type, and vice-versa. Even so, by calculating which equilibrium emerges for each type we can determine an ex-ante expected payoff. For instance, if equilibrium A3A4 does exist for both the low type and high type then the attackers ex-ante expected payoff (before victim type is known) is $\alpha(\beta T_1 + (1 - \beta)T_0 + V - L - k_D)$.

8.3.6 The Value of Private Information

We are now in a position to analyse and quantify the payoff consequences of private information for the victim. For any set of parameters $L, V, T_0, T_1, k_D, k_N, \alpha$ and β we can: (i) determine which, if any equilibrium will hold in a game with incomplete information on victim's type, (ii) determine which equilibium will hold in the games where victim's type is known to be high or low, (iii) calculate expected payoffs of the attacker and victim with and

Table 8.4: Expected payoff of attacker and victim in equilibrium when type is known.

Equilibrium	attacker	Victim
A1A2 ($i = \{0, 1\}$)	$\alpha T_i + V - L - \alpha k_D$	$-\alpha T_i - V$
A3A4 ($i = \{0, 1\}$)	$\alpha(T_i + V - L - k_D)$	$-\alpha T_i - V$
B1B4 ($i = \{0, 1\}$)	$\alpha T_i + V - L - \alpha k_D - (1 - \alpha)k_N$	$-\alpha T_i - V$
C1C4 ($i = \{0, 1\}$)	$\alpha T_i + V - L$	$-\alpha T_i - V$

without incomplete information on victim's type, and (iv) quantify the payoff impact of private information. We provide three examples.

For our first example we consider parameters $L = 1, V = 3, \alpha = 0.5, T_0 = 2, T_1 = 4, k_D = 0.1$ and $k_N = 6$. Imputing the parameter values into Theorems 8.3.1-8.3.3 it becomes apparent that there exists a separating equilibrium for any value of β and does not exist a pooling equilibrium (with or with no signal) for any value of β. This example, thus, focuses on the case of a separating equilibrium. In Figure 8.2 we plot expected payoffs (as given in Tables 8.3 and 8.4) as a function of β.

You can see in Figure 8.2 that the payoff of the attacker is substantially lower when the type of the victim is not known. The difference reaches a maximum at the point of transition between equilibria A1 and A2 given by $T_0 = \beta T_1 - (1 - \beta)(V - L)$ or equivalently

$$\beta = \frac{T_0 + V - L}{T_1 + V - L}. \tag{8.10}$$

For the parameters in our example this gives $\beta = 2/3$. If the type of the victim is unknown the expected payoff of the attacker is 2.95. If the type of the victim is known the ex-ante expected payoff of the attacker is 3.62. So, the attacker's payoff is 18.43% lower if it does not know the type of the victim.

You can see in Figure 8.2 that the victim's payoff is higher if the attacker does not know their type and $\beta < 2/3$. The intuition being that the attacker sets the ransom as if the victim is low type (equilibrium A1) and, thus, the high type is not exploited as much as they would have been if type was known. If $\beta > 2/3$ we see that the payoff of the victim is the same whether or not the attacker knows their type. In this case the attacker sets the ransom as if the victim is high type (equilibrium A2). This means the high type is maximally exploited by the attacker, while the low type does not pay the ransom and, therefore, suffers recovery and reputational losses. The net effect for the victim is the same as if the

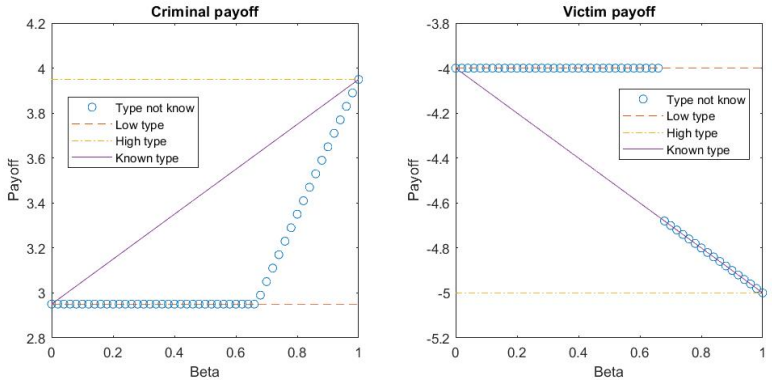


Figure 8.2: Expected payoff of the attacker and victim when $L = 1, V = 3, \alpha = 0.5, T_0 = 2, T_1 = 4, k_N = 6, k_D = 0.1$. An example of a separating equilibrium.

attacker knew their type and they were maximally exploited. While the victims payoff is the same (for $\beta > 2/3$) whether type is known or not, we remind that the attacker's payoff is lower when the victim's type is not known. This is because the attacker loses out from the low type not paying the ransom.

In our second example we set $k_N = 0.9$ while keeping everything else the same ($L = 1, V = 3, \alpha = 0.5, T_0 = 2, T_1 = 4, k_D = 0.1$). Imputing the parameter values into Theorems 8.3.1-8.3.3 it becomes apparent that there exists a pooling equilibrium with signaling for any value of β and does not exist a separating equilibrium or pooling equilibrium with no signal for any value of β. In Figure 8.3 we plot the corresponding payoffs. Again, we see that the attacker loses payoff from not knowing the type of the victim. This loss is maximal at the transition from equilibrium B1 to B2, given by $\alpha T_0 = \beta \alpha T_1 - (1 - \beta)(V - L)$ or equivalently

$$\beta = \frac{\alpha T_0 + V - L}{\alpha T_1 + V - L}. \tag{8.11}$$

For the parameters in our example this gives $\beta = 3/4$. If the type of the victim is unknown the expected payoff of the attacker is 2.5. If the type of the victim is known the ex-ante expected payoff of the attacker is 3.25. So, the attacker's payoff is 23.08% lower because it does not know the type of the victim.

The relative trade-offs for the victim are similar in the pooling example as the separating example. In particular, if the attacker sets the ransom for a victim of low type (equilibrium B1) then the victim gains from their type being private

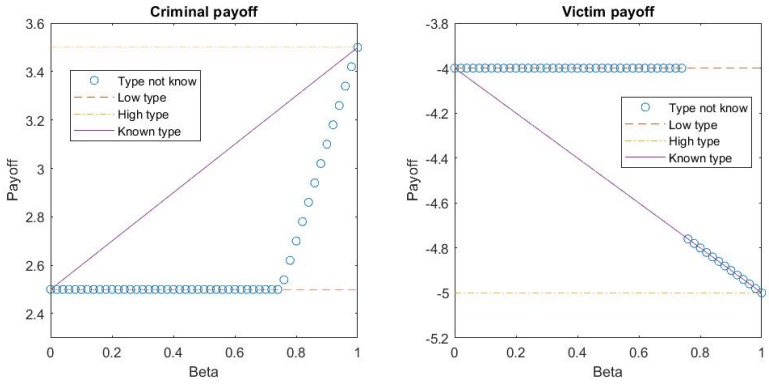


Figure 8.3: Expected payoff of the attacker and victim when $L = 1, V = 3, \alpha = 0.5, T_0 = 2, T_1 = 4, k_N = 0.9, k_D = 0.1$. An example of a pooling equilibrium with signalling.

if they are high type. If, however, the attacker sets the ransom for a victim of high type (equilibrium B2) then the victim does not gain from their type being unknown. In summary, the attacker loses payoff from not knowing the victim's type. The victim gains from their type being unknown in the case of equilibrium A1, B1 and also A3 and B3. The victim does not gain from the type being unknown in the case of equilibrium A2, A4, B2 and B4.

It is interesting to compare payoffs when $k_N = 0.9$ with those when $k_N = 6$ (for, say, $\beta = 2/3$). It can be seen from Figures 8.2 and 8.3 that the attackers expected payoff is higher when $k_N = 6$. This may seem counter-intuitive given that a high k_N means a higher cost from signalling. We highlight, however, that a high k_N results in a separating equilibrium that allows the type DE attacker to extract a high ransom because their signal of data exfiltration is credible. Specifically, when $k_N = 6$ the type DE sets ransom $R^*_{S0} = T_0 + V - L = 4$, while a type NDE sets ransom $R^*_{NS} = V - L = 2$. The expected payoff of the attacker is, therefore, $\alpha(R^*_{S0} - k_D) + (1 - \alpha)R^*_{NS} = 3.9\alpha + 2(1 - \alpha) = 2.95$.

By contrast, when $k_N = 0.9$ we obtain a pooling equilibrium in which the attacker's signal of data exfiltration is not sufficiently credible. This lowers the ransom the attacker can demand to $R^*_{P0} = \alpha T_0 + V - L = 3$. Consequently the type DE gets a lower payoff with the lower k_N (2.9 compared to 3.9). The type NDE, by contrast, has a higher payoff (2.1 compared to 2) because they are also able to demand ransom R^*_{P0}, although they incur cost k_N. The expected payoff

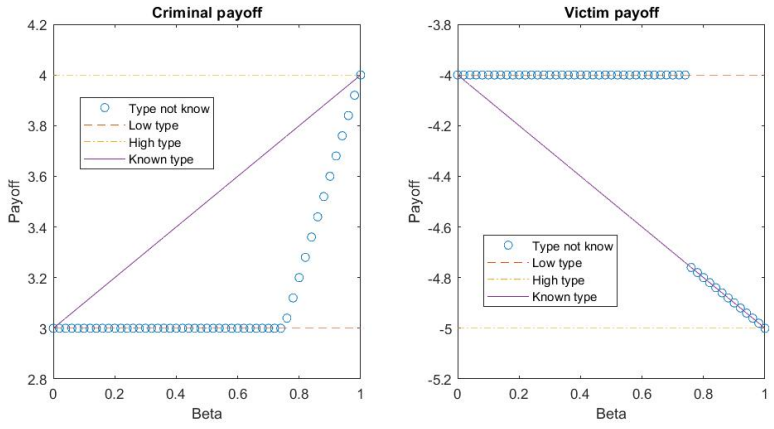


Figure 8.4: Expected payoff of the attacker and victim when $L = 1, V = 3, \alpha = 0.5, T_0 = 2, T_1 = 4, k_N = 6, k_D = 5$. An example of a pooling equilibrium with no signal.

of the attacker is $R^*_{P0} - 0.1\alpha - 0.9(1-\alpha) = 2.5$. Overall, therefore, the attacker has a lower expected payoff when $k_N = 0.9$ compared to $k_N = 6$ (2.5 compared to 2.95). This trade-off is apparent from the payoffs in Table 8.3, comparing A1 and B1.

For our final example we set we set $K_D = 5$ and $k_N = 6$ while keeping everything else the same ($L = 1, V = 3, \alpha = 0.5, T_0 = 2, T_1 = 4$). Imputing the parameter values into Theorems 8.3.1-8.3.3 it becomes apparent that there exists a pooling equilibrium with no signal for any value of β and does not exist a separating equilibrium or pooling equilibrium with signalling for any value of β. In Figure 8.4 we plot the corresponding payoffs. Again, we see that the attacker loses payoff from not knowing the type of the victim. This loss is maximal at the transition from equilibrium C2 to C4 given by $\Phi_P = 0$. This gives the same critical value of β as detailed in equation 8.11, which we know, for the parameters in our example, yields $\beta = 3/4$. If the type of the victim is unknown the expected payoff of the attacker for $\beta = 3/4$ is 3. If the type of the victim is known the ex-ante expected payoff of the attacker is 3.75. So, the attacker's payoff is 20% lower because it does not know the type of the victim.

Comparing Figures 8.3 and 8.4 you can see that the outcomes are very similar. Indeed, the victim payoff is exactly the same in the case of a pooling equilibrium

with signalling and no signalling. The criminal payoff is higher in the case of a pooling equilibrium with no signalling because they no long incur the cost of signalling. This reiterates the point that a high value of signalling, k_N and/or K_D, may not be an effective deterrent of ransomware because it can result in equilibria where criminals need not incur costs of signalling.

You can also see in Table 8.3 that the payoff of the victim is not directly impacted by k_N or k_D. This is because the criminal is able to extract the same surplus from the victim in equilibria A1, A3, B1, B3, C1 and C2. Generally, speaking, as would be expected, the loss to the victim is reduced by lowering T_0, T_1, V and β. The victim's payoff is also reduced by lowering α. Thus, reducing the losses from data exfiltration as well as reducing the probability of data exfiltration reduce the losses to the victim.

8.4 Conclusion

This paper provides a game-theoretic analysis of the double-sided information asymmetry in double-extortion ransomware attacks. We recognised that victims are typically unable to verify if data was exfiltrated or not, while attackers typically do not know the value of any data exfiltrated. We modeled the ransomware attack as a signaling game, where attackers could signal if data is exfiltrated and victims pay based on the ransom, signal and the value of information. Our key contribution is that, depending on the parameters of the game, private information of the victim (about the value of exfiltrated data) significantly lowers the profitability of the attack for the criminal. It is, therefore, in the interests of potential victims, businesses, organisations, and/or individuals, to retain and amplify the extent of their private information.

A further insight of our work is the subtle importance of signaling costs in determining equilibrium outcomes. If it is costly for criminals to signal data exfiltration, and they look to signal data exfiltration, then this can disrupt the business model by lowering profits. For instance, if the criminals need to search through files, or piece information together, to provide credible proof of data exfiltration then this is a barrier to extracting profitable ransoms. If, however, the costs of signaling become very high we may end up with an equilibrium in which there is data exfiltration but no signalling. This can increase criminal profit because they no longer need to incur the costs of signalling. A careful balancing act is, therefore, needed whereby it is costly for criminals to signal, but criminals are incentivized to reveal some information about data exfiltration via their actions. In terms of our model, this implies as high a cost of signaling

when data is exfiltrated (k_D) consistent with a separating equilibrium and a very high cost of signaling if data is not exfiltrated ($K_N >> k_D$).

There are various limitations in applying a game-theoretic framework to real-life situations. For example, the assumption of common knowledge of game parameters is strong: most probably there is little opportunity to learn of these parameters through repeated interaction between attacker and victim. Furthermore, it might be hard for victims to determine the value of the exfiltrated data, especially if it is uncertain which data is exfiltrated. Another limitation is the applicability of Bayesian equilibrium: while it describes an outcome in which no one wants to change their strategy, it does not predict the path towards an equilibrium. Therefore it is unknown, if there are multiple possible equilibria, which equilibrium will be reached.

Despite these limitation, we believe that a game-theoretic analysis gives useful insights about the interaction between attackers and victims during double-extortion ransomware. Beyond the two key insights mentioned above concerning the role of private information and signalling costs, according to our model, the most effective way to disrupt the attackers profitability is to: lower the probability of 'successful' data exfiltration, lower the probability the victim has files of high reputational cost, and lower the recovery cost from an attack. This would involve a mix of prevention (to lower the probability of data exfiltration and loss of sensitive data) as well as improved recovery options, such as back-ups.

These results align with preventive measures suggested by others [18, 28, 24, 20, 21]. Lee et al. [18] proposes a strategy to hide files from attackers. By considering real-world ransomware samples, there experiments show that this strategy is a cost-effective method to decrease the probability of valuable files being exfiltrated. Mundt and Baier [28] propose a strategy based on automated mitigation of attackers where data exfiltration takes place. This strategy is based on finding a fingerprint of data exfiltration in ransomware attacks and building monitoring systems which prevent data exfiltration to take place. Although their strategy is an efficient way to prevent the same type of attacks, it does not prevent new attacking patterns to be detected and prevented. Finally, Meurs et al. mention the use of canary files, which are files which alerts a monitoring systems if the files are moved, copied or edited. This strategy might be useful in preventing data exfiltation, but does depend on quick follow-up if a canary file alerts a monitoring system.

It would be beneficial for victims to take preventive measures. However, if data exfiltration has taken place, our study proposes a strategy to lower the impact of data exfiltration during ransomware attacks: victims should keep the value of the exfiltrated data as private as possible, as exposing this information

might increase the ransom. Finally, it is important to stress the following externality effect: the more victims safeguard their sensitive data the more that benefits other businesses, including those with vulnerable sensitive data. This is because it would revise downwards the beliefs of attackers about the ransoms they can reasonably expect victims to pay. This externality effect should be acknowledged by policy makers. In particular, it means businesses will under-invest in cyber security prevention and recovery compared to the social optimum. This can justify government support for cyber security investment.

Bibliography

[1] G. A. Akerlof. 'The market for "lemons": Quality uncertainty and the market mechanism'. *The Quarterly Journal of Economics* 84.3, 1970, pp. 488–500.

[2] R. P. Baksi and S. J. Upadhyaya. 'Game Theoretic Analysis of Ransomware: A Preliminary Study'. *ICISSP*. 2022, pp. 242–251.

[3] J. S. Banks and J. Sobel. 'Equilibrium selection in signaling games'. *Econometrica: Journal of the Econometric Society*, 1987, pp. 647–661.

[4] A. Cartwright, E. Cartwright, J. MacColl, G. Mott, S. Turner, J. Sullivan and J. R. Nurse. 'How cyber insurance influences the ransomware payment decision: theory and evidence'. *The Geneva Papers on Risk and Insurance - Issues and Practice* 48.2, 2023, pp. 300–331.

[5] E. Cartwright, J. H. Castro and A. Cartwright. 'To pay or not: game theoretic models of ransomware'. *Journal of Cybersecurity* 5.1, 2019.

[6] L. W. Cong, C. R. Harvey, D. Rabetti and Z. Wu. 'An anatomy of crypto-enabled cybercrimes', 2023.

[7] L. Y. Connolly, D. S. Wall, M. Lang and B. Oddson. 'An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability'. *Journal of Cybersecurity* 6.1, 2020, tyaa023.

[8] L. Y. Connolly and D. S. Wall. 'The rise of crypto-ransomware in a changing cybercrime landscape: Taxonomising countermeasures'. *Computers & Security* 87, 2019, p. 101568.

[9] S. R. Etesami and T. Başar. 'Dynamic games in cyber-physical security: An overview'. *Dynamic Games and Applications* 9.4, 2019, pp. 884–913.

[10] D. Fudenberg and J. Tirole. *Game theory*. MIT press, 1991.

[11] E. Galinkin. 'Winning the Ransomware Lottery: A Game-Theoretic Approach to Preventing Ransomware Attacks'. *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Proceedings*. Vol. 12. Springer International Publishing, 2021, pp. 195–207.

[12] J. C. Harsanyi. 'Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model'. *Management Science* 14.3, 1967, pp. 159–182.

[13] M. Humayun, N. Jhanjhi, A. Alsayat and V. Ponnusamy. 'Internet of things and ransomware: Evolution, mitigation and prevention'. *Egyptian Informatics Journal* 22.1, 2021, pp. 105–117.

[14] Q. Kerns, B. Payne and T. Abegaz. 'Double-Extortion Ransomware: A Technical Analysis of Maze Ransomware'. *Proceedings of the Future Technologies Conference (FTC) 2021*. Vol. 3. Springer International Publishing, 2022, pp. 82–94.

[15] D. M. Kreps and J. Sobel. 'Signalling'. *Handbook of game theory with economic applications* 2, 1994, pp. 849–867.

[16] M. Lang, L. Connolly, P. Taylor and P. J. Corner. 'The evolving menace of ransomware: A comparative analysis of pre-pandemic and mid-pandemic attacks'. *Digital Threats: Research and Practice* 4.4, 2023, pp. 1–22.

[17] A. Laszka, S. Farhang and J. Grossklags. 'On the economics of ransomware'. *Decision and Game Theory for Security: 8th International Conference, GameSec 2017, Proceedings*. Springer International Publishing, 2017, pp. 397–417.

[18] S. Lee, S. Lee, J. Park, K.-W. Kim and K. Lee. 'Hiding in the Crowd: Ransomware Protection by Adopting Camouflage and Hiding Strategy With the Link File'. *IEEE Access*, 2023.

[19] Z. Li and Q. Liao. 'Game Theory of Data-selling Ransomware'. *J. Cyber Secur. Mobil.* 10.1, 2021, pp. 65–96.

[20] Z. Li and Q. Liao. 'Preventive portfolio against data-selling ransomware—A game theory of encryption and deception'. *Computers & Security* 116, 2022, p. 102644.

[21] S. Liu and X. Chen. 'Mitigating Data Exfiltration Ransomware through Advanced Decoy File Strategies', 2023.

[22] M. Maschler, S. Zamir and E. Solan. *Game theory*. Cambridge University Press, 2020.

[23] S. R. Matthijsse, M. van't Hoff-de Goede and R. Leukfeldt. 'Your files have been encrypted: a crime script analysis of ransomware attacks'. *Trends in Organized Crime*, 2023, pp. 1–27.

[24] T. Meurs, E. Cartwright, A. Cartwright, M. Junger and A. Abhishta. 'Deception in Double Extortion Ransomware Attacks: An Analysis of Profitability and Credibility'. *Computers & Security*, 2023, p. 103670.

[25] T. Meurs, E. Cartwright, A. Cartwright, M. Junger, R. Hoheisel, E. Tews and A. Abhishta. 'Ransomware Economics: A Two-Step Approach To Model Ransom Paid'. *18th Symposium on Electronic Crime Research, eCrime*. 2023.

[26] T. Meurs, M. Junger, E. Tews and A. Abhishta. 'Ransomware: How attacker's effort, victim characteristics and context influence ransom requested, payment and financial loss'. *Symposium on Electronic Crime Research, eCrime*. 2022.

[27] G. Mott, S. Turner, J. R. Nurse, J. MacColl, J. Sullivan, A. Cartwright and E. Cartwright. 'Between a rock and a hard (ening) place: Cyber insurance in the ransomware era'. *Computers & Security* 128, 2023, p. 103162.

[28] M. Mundt and H. Baier. 'Threat-based simulation of data exfiltration toward mitigating multiple ransomware extortions'. *Digital Threats: Research and Practice* 4.4, 2023, pp. 1–23.

[29] K. Oosthoek, J. Cable and G. Smaragdakis. 'A Tale of Two Markets: Investigating the Ransomware Payments Economy'. *arXiv preprint arXiv:2205.05028*, 2022.

[30] M. J. Osborne. *An introduction to game theory*. 3rd ed. New York: Oxford University Press, 2004.

[31] H. Oz, A. Aris, A. Levi and A. S. Uluagac. 'A survey on ransomware: Evolution, taxonomy, and defense solutions'. *ACM Computing Surveys (CSUR)* 54.11s, 2022, pp. 1–37.

[32] P. Ryan, J. Fokker, S. Healy and A. Amann. 'Dynamics of targeted ransomware negotiation'. *IEEE Access* 10, 2022, pp. 32836–32844.

[33] B. Sabir, F. Ullah, M. A. Babar and R. Gaire. 'Machine learning for detecting data exfiltration: a review'. *ACM Computing Surveys (CSUR)* 54.3, 2021, pp. 1–47.

[34] F. Ullah, M. Edwards, R. Ramdhany, R. Chitchyan, M. A. Babar and A. Rashid. 'Data exfiltration: A review of external attack vectors and countermeasures'. *Journal of Network and Computer Applications* 101, 2018, pp. 18–54.

[35] I. Vakilinia, M. M. Khalili and M. Li. 'A Mechanism Design Approach to Solve Ransomware Dilemmas'. *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Proceedings*. Springer International Publishing, 2021, pp. 181–194.

[36] T. Yin, A. Sarabi and M. Liu. 'Deterrence, Backup, or Insurance: Game-Theoretic Modeling of Ransomware'. *Games* 14.2, 2023, p. 20.

[37] H. P. Young. *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press, 1998.

[38] Y. Zhao, Y. Ge and Q. Zhu. 'Combating Ransomware in Internet of Things: A Games-in-Games Approach for Cross-Layer Cyber Defense and Security Investment'. *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Proceedings*. Springer International Publishing, 2021, pp. 208–228.

Part IV

Law Enforcement Interventions

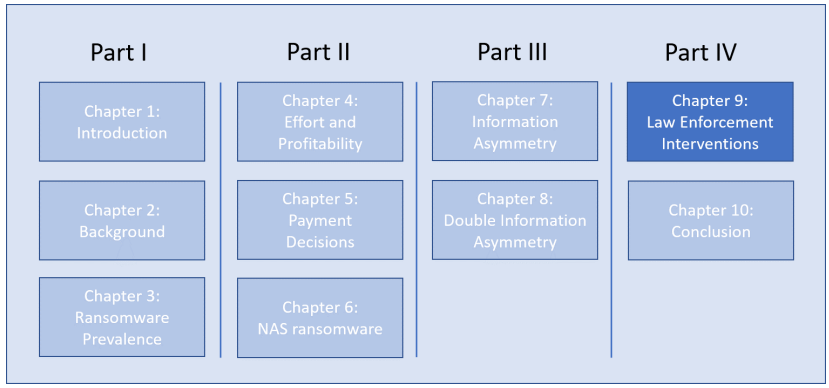
This page is intentionally left blank.

One explorer swore he saw a Tarahumara catch a deer with his bare hands, chasing the bounding animal until it finally dropped dead from exhaustion

~ Chris McDougall

Chapter 9

Evaluating Law Enforcement Interventions



Ransomware poses an increasing challenge to society, yet there is a notable gap in research on the effectiveness of law enforcement interventions. A key insight from this study is that the presence of victims' details on leak pages following double-extortion ransomware attacks offers a unique opportunity to evaluate these interventions. Analyzing a dataset containing victims published by ransomware groups, we assess the impact of five specific types of interventions: arresting group members, taking down leak page server infrastructure, freezing crypto assets, releasing decryptors, and imposing sanctions. From a collected list of interventions, we categorize ransomware groups' responses into three actions: ceasing operations, continuing operations, or rebranding under a new name.

9.1 Introduction

In recent years, ransomware has emerged as a significant societal concern [32, 34, 24, 5]. To our knowledge, systematic empirical research towards law enforcement interventions against ransomware are lacking [50]. We have identified a useful data source to evaluate the effectiveness of law enforcement (LE) interventions: victims published on leak pages during double-extortion ransomware attacks. This approach helps bridge the gap in understanding how LE interventions can disrupt or deter ransomware attacks effectively.

We examine five distinct types of law enforcement interventions: the arrest of ransomware group members, the takedown of leak page server infrastructure, the freezing of crypto assets, the release of decryptors, and the imposition of sanctions on ransomware group members. While these interventions are commonly employed, their efficacy has not been systematically evaluated [50].

To address this gap, we create metrics and identify data sources that enable an evaluation of these interventions. We measure the efficacy of LE interventions by considering the response of ransomware groups: ceasing operations, continuing operations, or rebranding under a new name. Additionally, we assess the characteristics of victims targeted by groups facing interventions compared to those who do not, as well as the changes in ransomware operations pre- and post-intervention for groups that continue their activities. These measures allow us to study the effectiveness of LE interventions against ransomware groups.

We analyse three data sources: a dataset of 12,250 ransomware victims posted by 134 ransomware groups, including characteristics of those victims such as country, number of employees, and sector; a constructed list of law enforcement interventions; and a list of rebranding occurrences. The theoretical foundation for evaluating the effectiveness of LE interventions is Situational Crime Prevention (SCP), a criminological theory that states crime occurs through favorable opportunities [22, 36, 38]. Thus, effective interventions should alter the cost-benefit trade-off of these circumstances[27].

The primary goal of this study is to assess the impact of law enforcement interventions on the operations of ransomware groups. To address our goal, we explore three sub-questions:

RQ 1: Which ransomware groups face a law enforcement intervention?

RQ 2: How do ransomware groups respond to law enforcement interventions?

RQ 3: How do ransomware operations compare prior and post-intervention for ransomware groups who continue after an intervention?

To answer these research questions, we analyzed data from 20 February 2020 till 4 March 2024 from 12,250 ransomware victims listed on leak pages by 134 ransomware groups, alongside 29 law enforcement interventions, and identified 19 groups that rebranded. We also conducted interviews with police officers, public prosecutors, and cyber security experts to validate our findings.

Our key contributions are:

1. Ransomware groups are more likely to face an intervention if they have a large number of victims, target many large companies, or maintain long uptime of their leak pages, indicating selective LE targeting.

2. Of the 17 groups with active leak pages during intervention, 8 ransomware groups continue with their operations, 7 cease operations and 2 rebrand after an intervention. Ceasing operations was most often associated with the takedown of a leak page server.

3. Crime displacement was limited. Rebranding occured only twice (N=17) post-intervention. Furthermore, groups continuing attacks after an intervention typically have less victims post-operation compared to prior.

The outline of this paper is as follows: In §9.2, we combine existing literature on SCP with a small pilot of interviews with law enforcement experts to state our propositions. Subsequently, in §9.3, we present our data and the methodology. Afterwards, §9.4 presents the results of the analysis of leak page data and law enforcement interventions. To conclude, we discuss our findings, limitations and outline implications for policy makers in §9.5, §9.6 and §9.7, respectively.

9.2 Related Works and Propositions

This section begins with an examination of double-extortion ransomware and Situational Crime Prevention (SCP) theory, first introduced by [18]. This will be combined with the results of a small pilot study. In this pilot study, we interviewed 13 experts who work in the criminal justice system, among which police officers, public prosecutors and cyber security experts. The goal was to explore the anticipated impact of law enforcement interventions of these experts and to validate intervention and rebranding lists used later in this chapter. For a full description of the pilot, please contact the lead author of the study.

The present study will not use hypothesis testing due to the limited number of interventions on which information is available. Instead, we will work with propositions and assess whether the empirical findings align with these propositions.

9.2.1 Situational Crime Prevention and LE Interventions

Ransomware is a type of malware that encrypts files and demands a ransom for access [73]. Double-extortion ransomware involves both data encryption and exfiltration [72, 70, 68]. Attackers threaten to publish exfiltrated data on their 'leak pages' if the ransom is not paid. Typically, if negotiations fail, the victim's name is listed on the leak page, followed by the publication of data after a delay. Stolen data may also be sold to other malicious actors, potentially for use in subsequent attacks [65, 71].

Meurs et al. [73] found that victims are willing to pay 5.5 times larger ransom amounts when data is exfiltrated, making double-extortion more lucrative than traditional ransomware [24, 11, 73]. Malicious actors, therefore, target victims who highly value their data, increasing the attack's cost [47].

In response to the global ransomware crisis, law enforcement agencies conduct various interventions, such as taking offline servers hosting leak pages to prevent data leakage. These interventions face challenges due to the international nature of ransomware crimes, information asymmetries, conflicting jurisdictions, and limited enforcement capabilities [66, 56]. Information asymmetry refers to the inconsistent enforcement of legal mandates for victims to share information about ransomware attacks. Conflicting jurisdictions occur when attackers reside in countries unlikely to prosecute them, often those not party to the Budapest Convention, which provides a unified legal framework for prosecuting cybercrime [21]. Many law enforcement agencies also suffer from a lack of personnel and technical resources, making it difficult to combat ransomware effectively. Forensic and diplomatic complications, such as difficulty attributing attacks to specific individuals, further hinder interventions [66]. While ransomware attacks can scale up easily, enhancing law enforcement responses is considerably more challenging.

Evaluating the impact of police interventions is crucial for combating ransomware. One main goal of these interventions, besides arresting attackers, is preventing subsequent attacks. Situational Crime Prevention (SCP) is an approach aimed at understanding and addressing crime prevention [18, 43, 50]. SCP focuses on the idea that malicious actors make rational choices based on favorable opportunities [27, 22, 36, 38]. Because specific types of crime differ in their modus operandi, SCP is usually 'crime specific': measures that prevent one type of crime may not prevent another [20, 26, 50].

SCP is distinguished by its focus on five general strategies: Increase the Effort, Increase the Risks, Reduce the Rewards, Reduce Provocations, and Remove Excuses [18, 19, 50, 43]. These strategies aim to deter potential offenders by mak-

ing crimes more difficult or less appealing. The effectiveness of SCP strategies in combating various crimes has been studied [8, 51, 50]. For extensive elaboration of these five principles, refer to [20, 26, 17, 50, 43].

Studies evaluating SCP measures against cybercrime are scarce [50]. A notable study includes Bada et al. [2], who evaluated SCP interventions like cease and desist letters, police visits, and workshops on cybercrime, finding a general decrease in self-reported offending. Another study conducted a meta-review of cybersecurity interventions, highlighting that implementation effectiveness drives intervention success more than the mere presence of controls [91].

Other studies mainly focused on SCP in their recommendations following a crime script analysis [64, 23, 79, 68]. A crime script describes all relevant aspects of a crime's modus operandi, from preparation to aftermath [20, 26, 7, 59, 30]. Because crimes differ in their modus operandi, SCP is usually 'crime specific'. Therefore, evaluating an intervention strategy to combat ransomware requires understanding the specific ransomware crime script.

A detailed account of the ransomware crime script involves recognizing the attack's lifecycle: infrastructure and malware development, network access, encryption, extortion through data exfiltration, ransom negotiation, data leakage for non-complying victims, and money laundering by the attackers [68, 72].

For this study, we operationalize ransomware by identifying ransomware variants typically by their file extensions post-encryption. Although multiple attackers might use the same variant, the associated leak page server is usually specific to a single group. We focus on leak pages, using 'ransomware group' to denote the group behind a leak page server of a specific ransomware strain, variant, or family.

We analyze five specific interventions targeting ransomware groups, within the control of law enforcement or other government-related agencies. For brevity, we define interventions by either law enforcement, government entities, or cybersecurity companies as law enforcement interventions. Each intervention is discussed, highlighting its importance and alignment with SCP strategies.

The five interventions used by law enforcement in this chapter include two that mainly increase the risks to attackers, two that decrease attackers' rewards, and one that increases the necessary effort.

The following interventions **increase the risks**:

Intervention 1: Arrests. Arrests is defined here as arrests of malicious actors associated with a specific ransomware group. This might be a 'low-level' malicious actor, like a money mule, but also a 'key player'. Obviously, when some of the attackers have been arrested, the perception of the risk of get-

ting caught might increase for the other attackers in a ransomware group. An important point is that it is unknown for those persons how much law enforcement knows.

Intervention 2: Sanctions. Another intervention consist of asset freeze or travel restrictions to a specific individual linked to ransomware [31]. Not only does this restrict the movement of the malicious actor, but also there is a name & shame element: the name of the malicious actor becomes known in public. This might increase the perceived risk for the malicious actor to continue his/her operations, and/or other malicious actors might be hesitant to work together with that person [31]. This clearly increases the risk for an attacker, since law enforcement knows who they are.

The following interventions **decrease the rewards**:

Intervention 3: Crypto-asset freezing. Crypto-asset freezing is the blocking of transactions of crypto-assets related to victims of a certain ransomware group. A crypto-exchange might block any transactions from a wallet, upon request from law enforcement. This means that an attacker cannot access his cryptocurrencies, thereby reduce the rewards of his malicious actor activities'.

Intervention 4: Decryptor release. Decryptor release is the release of a decryptor to victims by law enforcement. With the decryptor, the victim can regain access to files without paying the ransom. Often this is done through NoMoreRansom, an initiative to release decryptor keys safely to ransomware victims. The effect is that ransomware victims will not pay a ransom if a free decryptor is available. Subsequently, if they want to continue the attacks, ransomware groups would need to change their ransomware to make sure the victims could not recover without buying the decryption key.

The following intervention **increases the effort**:

Intervention 5: Takedown leak page server. Server takedowns relate to takedowns of leak page servers. Takedowns of other infrastructure of the malicious actors are outside the scope of this paper, since these are often not made public. While this would imply that attackers need to rebuild their infrastructure, this can also lead to an increased perceived risk. After a takedown, the group has to, which increases the effort. Furthermore, the ransomware group learns law enforcement has them in their crosshairs, increasing the perceived risks.

By comparing the interventions with the crime script, we observe that arrests and sanctions directly confront the attackers. Freezing crypto-assets disrupts the attack's monetization phase, while the release of decryptors intervenes in the file encryption process. Leak page server takedowns address the step where victims' data is exposed on leak pages. Law enforcement strategies targeting other steps of the crime script, like preventing malicious actors from gaining access to a victim's system, are outside the scope of the present study.

In addition to exploring the interventions and their effectiveness according to SCP, it is essential to determine which groups are targeted by these law enforcement interventions. This topic will be addressed in the following subsection.

9.2.2 Ransomware Groups Facing LE Interventions

Law enforcement agencies generally operate with limited resources [84] and face the challenge of more malicious actors than they can feasibly pursue. As a result, prioritization is essential in deciding which malicious actors to target [84, 46]. It seems reasonable that they will target malicious actors, according to certain selection criteria. For example, they will prioritize malicious actors who have many victims or high-value victims. As one police officer in the pilot mentioned:

> Police Officer 3: "*I think it might be smart in your study to only focus on the ransomware groups with more than, let's say, 20 victims. The smaller groups are not that interesting.*"

This leads to the following proposition:

Proposition 1: *Ransomware groups targeting a greater number and more significant victims are more likely to face law enforcement interventions.*

Likewise, the same reasoning would imply that countries experiencing a high number of victims may also be more frequently involved in interventions. Legal frameworks in most countries are built on the principles of subsidiarity and proportionality, which suggest that more aggressive interventions may be justified when the impact of ransomware crimes is relatively more significant. As one public prosecutor explains:

> Public Prosecutor: "*A takedown is not explicitly described in our legal code, so we must carefully examine the nature of the website/server, its location, and its technical aspects. Is one server sufficient, or is it a network of servers that needs to be addressed? This is then assessed*

> *within the legal framework, considering principles of proportionality and subsidiarity.*"

From this, we can infer that the extent of law enforcement interventions in a country may correlate with the number of ransomware victims it has encountered.

Proposition 2: *Countries with a higher incidence of ransomware victims are more likely to undertake law enforcement interventions than those with fewer victims.*

Having examined why certain ransomware groups might face an intervention, it is now important to explore how these groups respond to such actions.

9.2.3 Ransomware Groups Responding to LE Interventions

Interestingly, participants from our pilot study were less optimistic about the impact of interventions compared to empirical evidence from studies supporting the SCP principles [18, 19, 50, 43]. Participants expressed varying opinions about the effectiveness of arresting ransomware actors. Six participants believed arrests have a significant impact, four noted the impact depends on the malicious actor's role within the organization, and three felt arrests have no effect on ransomware activities or the effect is unknown.

> Cybersecurity Expert 5: "*The position of the individual is crucial during an arrest. Otherwise, it doesn't make much sense. You really need to apprehend the key figures. If you only go after small individuals, the big ones will just keep going.*"

With respect to 'increase the effort', 9 out of 13 participants believed that taking down leak page server would only have a symbolic impact.

> Police Officer 2: "*The effects of taking down leak page servers on ransomware attacks are mainly symbolic. It sparks a lot of discussion on online platforms and is considered extremely annoying for Ransomware-as-a-Service (RaaS) actors. It simply damages their reputation when their leak page is taken down.*"

Only four participants were confident that a combination of LE interventions involving both arrests and takedown of leak page servers was more effective than either intervention alone.

> Police Officer 6: "*A combined approach works better because taking down a website is less complex and therefore less impactful than actually apprehending someone. This also has a greater deterrent effect. The uncertainty of what law enforcement knows will have a deterrent effect on malicious actors.*"

Most participants (n=6) believed the effectiveness of an intervention depends on the role of the arrested malicious actor and whether the malicious actors have backups of the leak page server.

> SOC analyst: "*I don't expect that a combination of arrest and takedown will be significantly more effective in reducing the activity of leak page servers than just an arrest or takedown alone. People may have more difficulty regrouping and calming down after an arrest than after a takedown or a combination of both. This may lead to a temporary decrease in activity, but they often return, usually after a few months. I do not want to diminish the work of various law enforcement organizations around the world. They provide justice for the victims and show that ransomware actors are not untouchable. With these steps, we will eventually catch up with this form of criminality.*"

Previous research similarly suggests that police officers, with respect to offline crime, generally hold more negative views about the effectiveness of police interventions [42], which is not justified considering the evidence. Based on these insights, we propose the following:

Proposition 3: *After a law enforcement intervention, a significant amount of the ransomware groups cease ransomware operations.*

An important criticism of interventions based on SCP is that they may not stop crime but merely displace it. This issue will be the focus of the following section.

9.2.4 Crime Displacement

One important consideration of crime prevention techniques is that, first, it is necessary to show, possibly in experimental research, that there is a real crime reduction and, second, there should be no crime displacement. [89] considered the effect of the takedown of a darknet forum. To assess the effects of a darknet market takedown of 220 vendors migrating to a new darknet forum. They found that although some vendors reused their PGP-key, most malicious actors started

with a clean slate, which meant that they were erasing their past reputation completely. This meant they had to rebuilt their reputation of being 'a reliable' drug seller afresh. The authors concluded that a takedown is costly for malicious actors, even if there is some crime displacement[89].

Proposition 4: *Ransomware groups that continue after an intervention will target fewer and less significant victims than before the intervention.*

Rebranding is an important phenomenon in the ransomware ecosystem, where one strain disappears and another emerges, typically using the same infrastructure, part of the malware code, and operated by the same actors [87, 15]. It is believed that rebranding occurs for two main reasons: to obscure activities from law enforcement and/or to establish a new, more intimidating reputation [87]. According to [15], in 2022, the average lifespan of a ransomware strain was only 70 days, a significant decrease from 153 days in 2021 and 265 days in 2020. It could be argued that not all rebranding efforts are publicly acknowledged. However, there is an incentive for malicious actors to make their rebranding known publicly to avoid having to rebuild their reputation from scratch, which could lead to lower ransoms from victims who are unsure if the group will return the decryption key after payment or might demand additional payments [13].

LE experts interviewed in our pilot study believe that there is a lot of rebranding.

> Cyber security expert 1: "*Yes, there is often a connection between takedowns and the rebranding of ransomware groups. This can happen depending on the circumstances and the motives of the group. A takedown operation can prompt a group to rebrand, especially if sanctions have been imposed on the group due to alleged ties with a certain entity. In such cases, they might choose to continue their activities under a new name to evade legal consequences. On the other hand, rebranding can also be an initiative from the group itself. They might feel that they are attracting too much attention and have become too prominent. Initially, they may enjoy the security industry writing about them, but if the details become too intricate or the scrutiny too close, they might decide to change their name. An example of this was the case with GandCrab, which had about 150 active affiliates. When the REVIL group emerged, the most influential affiliates were taken away [by the coordinators of the group]. There were one or two versions [of the new malware] where affiliates were no longer involved, after which they decided to rebrand.*

However, they retained much of the same code and structure, and the admins chose to keep the most capable affiliates during the rebranding."

Consequently, it is assumed that following a law enforcement intervention, malicious actors are more likely to publicly disclose their rebranding efforts. This publicly disclosed rebranding is done to maintain their reputation as a 'reliable' ransomware group, one that returns the decryption key after receiving payment.

Proposition 5: *Following a law enforcement intervention, ransomware groups are more likely to rebrand compared to continuing or ceasing operations.*

Our dataset could reveal different forms of rebranding among ransomware groups. For example, groups aiming to build a more fierce reputation might maintain their old brand for a period to smoothly transition infrastructure and affiliates to the new group. This mitigation could result in overlapping active periods for both the old and new leak pages. Conversely, rebranding following law enforcement intervention might be more abrupt, potentially leading to no overlap in the uptime of leak page servers. Such interventions could also provoke internal disputes or paranoia within the group. This could result in a groups splitting up is two or more different ransomware groups. These observations suggest a distinction between normal rebranding processes and those triggered by law enforcement actions. Therefore, we propose the following:

Proposition 6: *Rebranding following an intervention is more likely to be combined with a split-up and no overlapping time periods of leak pages, compared to rebranding without intervention.*

The next section will outline the data, operationalization of variables, and methods utilized in this chapter.

9.3 Data and Methodology

Three datasets form the basis of our analysis:

- **Dataset 1: Leak page data.** The main dataset is a nested dataset in which ransomware groups, publish the names of the organisations that were a victim of ransomware: the victim's information is nested within the ransomware group. Besides the names of the victims, groups publish smaller or larger parts of the data if they managed to exfiltrated those from the victim's system. If parts of the data are published on the leak page, we assume that that specific victim did not pay. Additional information on the

Table 9.1: Variables in the Leak Page Dataset and Missing Values

Variables	Unit / Categories	Missing Values	%
Ransomware Group	Categorical (134 Groups)	0/12,250	0%
Country	Categorical (156 Countries)	137/12,250	1.1%
Sector	Categorical (309 Sectors)	1,010/12,250	8.2%
Data Leaked	Binary (Yes = 1 / No = 0)	6,098/12,250	49.8%
Number of Employees	Categorical (Small, medium, large)	1,767/12,250	14.4%
Victim First Seen	Date (YYYY-MM-DD)	0/12,250	0%
Victim Last Seen	Date (YYYY-MM-DD)	0/12,250	0%

organizations were manually added by ecrime.ch and provided to the researchers [29]. The dataset also indicates victim's first and last seen dates, with the initial 21 observations considered outliers until bulk observations started on December 4, 2020. The dataset spanned from December 20, 2019, to March 4, 2024. It includes ransomware group names (categorical), country of victim (categorical), sector of victim (categorical), data leakage status (binary), and employee count of victim (categorical). The dataset contained 12,250 unique victims.

- **Dataset 2: Intervention list.** The second dataset comprises 36 LE interventions, with some combined into single events, resulting in 29 unique interventions. After excluding groups that stopped or rebranded before the intervention, we identified 17 unique interventions. The complete list is provided in Appendix A (Table 9.5). To systematically explore the impact of interventions on ransomware groups' operations, we focused on groups that maintained leak pages from December 20, 2019, to March 4, 2024. Initially, we searched for relevant scientific articles using academic databases like Scopus and Web-of-Science, but this yielded no results. Consequently, we shifted our focus to cybersecurity company blogs. Using Google, we performed targeted searches with queries combining 'intervention type' and 'ransomware group name' for each intervention type and group, resulting in 670 queries (5 interventions x 134 groups). We restricted our search to the first three pages of Google results, assuming high-quality information is ranked highest. Each search result was reviewed for articles, reports, and mentions discussing the impact of interventions on ransomware groups or potential rebranding. Acknowledging potential lim-

itations associated with using the Google Search Engine [57], we adopted four measures to mitigate the possibility of having missed interventions.

1. Cross-referencing our interventions with the ransomware cartography developed by CERT Orange Cyberdefense [80].
2. Conducting pilot study interviews, which identified two missing arrests.
3. Querying the Wayback Machine of NoMoreRansom to find decryptors and their availability dates [75], which did not yield additional decryptors.
4. Checking EU and USA sanctions websites for additional sanctions, with no new sanctions found [31, 86].

Consequently, we believe we have a reasonably complete overview of LE interventions against the ransomware groups included in our study.

- **Dataset 3: Rebranding list.** The third dataset consists of a list of ransomware group rebrandings, which we compiled using the same search strategy as for identifying interventions. This resulted in 19 instances of rebranding, with the list provided in Appendix B. Using Google, we searched for 'rebranding' AND 'ransomware group name', generating 134 queries aimed at uncovering rebranding events in cybersecurity blogs. We cross-referenced our rebranding list with the ransomware cartography developed by CERT Orange Cyberdefense [80]. Further validation was conducted through interviews from our pilot study, which added one more rebranding event to our list. Given the clandestine nature of rebranding, we acknowledge that our list may not be exhaustive. However, we believe it provides valuable exploratory insights to understand displacement within the scope of this study.

Overall, our findings yielded a list of 36 interventions, with some combined interventions treated as single events, resulting in 29 unique interventions. Groups stopping or rebranding before the intervention were excluded from the study, resulting in 17 unique interventions. The complete list is provided in Appendix A, with Table 9.5 presenting interventions alongside corresponding malicious actor actions after the intervention. Similarly, we identified 19 instances of ransomware group rebranding, with a list available in Appendix B.

Next, we describe the variables used in this chapter. The two **dependent variables** in our study are (see Table 9.1):

1a. *Law Enforcement Intervention:* This categorical variable addresses propositions 1 and 6 by indicating whether the ransomware group experienced an intervention within our dataset. For propositions 2-5, it is also important to know the type of intervention. Therefore, we categorize the interventions as follows: 'arrest', 'sanction', 'crypto', 'decryptor', 'takedown', 'takedown+arrest', 'takedown+decryptor', and 'takedown+decryptor+arrest'. These interventions are described in Section 9.2.1.

1b. *Response to Intervention:* This categorical variable addresses propositions 3 and 5 by indicating the different responses of ransomware groups to an intervention. Timing is crucial for this variable since some groups might have stopped publishing victims before a law enforcement intervention, making it impossible to measure the intervention's effect. If the ransomware group stopped publishing victims before the intervention, we denote the response as 'BEFORE'. If no victims were published on leak pages after an intervention, we assume the group stopped all ransomware operations, denoted as 'STOP'. If new victims were published after an intervention, we assume ransomware operations continued, denoted as 'CONTINUE'. If the groups rebranded after the intervention, they are categorized as 'REBRAND'.

The **independent variables** in this chapter are (See Table 9.1):

2a. *Ransomware Group:* Names of the ransomware groups involved (categorical). In total 134 groups were found online and were included in the leak page dataset.

2b. *Country of Victim:* The country where the victim is located (categorical). There were 156 countries in the leak page dataset. Due to the prevalence of single or infrequent observations in countries and sectors, aggregation was performed. The top 10 most frequent countries were used, other countries were aggregated to category 'Other'.

2c. *Economic sector of Victim:* The economic sector in which the victim operates (categorical). The victims represented in the leak page dataset were active in 309 sectors. Due to the prevalence of single or infrequent observations aggregation was performed. Sectors were manually categorized as important or critical according to EU NIS2 legislation [74]. After aggregation 3,356 victims were considered critical, 2,415 victims were considered important and 5,463 victims were considered none of these. Additionally, sectors were aggregated based on technical intensity, measured through

sector-level R&D expenditure [39]. After aggregation 2,391 victims were considered from sectors with high technological intensity, 3,187 victims with medium technological intensity, and 2,391 victims with low technological intensity. For an overview see Table 9.1. See Table 9.1.

2d. *Data Leakage Status:* Victims who were listed on the leak pages did not always have data exfiltrated. Data leakage status indicates whether data from the victim was or was not leaked, that is, data were published on the leak page (data leaked, binary: yes = 1 / no = 0).

2e. *Employee Count of Victim:* The number of employees working for the victim (categorical). Employee counts were aggregated into small (1-50 employees), medium (51-500 employees), and large (501+ employees) companies, following definitions by [14].

The analyses were conducted using RStudio and R version 4.3.1, employing packages *ggplot*, and *dplyr*. Listwise deletion was applied to handle missing observations. This research has received approval by the Ethics Committee at the University of Twente, registered under number 240026. We aim to collect empirical evidence which might align with the propositions as stated in Section 9.2.

- **Proposition 1:** Logistic regression was used to determine if ransomware groups targeting a larger number of significant victims were more likely to face law enforcement interventions . In this context, "significant" refers to companies that are either critical according to the NIS directive, technologically intensive, based in the USA or elsewhere, or are large enterprises.
- **Proposition 2:** Due to many countries having a small number of attacks, or have victims listed by ransomware groups who faced an interventio n, a non-parametric Spearman's correlation test tested if countries with a many ransomware are more likely to conduct law enforcement interventions.
- **Propositions 3 and 5:** A binomial regression model tested to what extent ransomware groups cease operations, continue operations or rebrand after a LE intervention, compared to a baseline of zero.
- **Proposition 4:** A paired t-test and the non-parametric Wilcoxon signed-rank test were employed to compare the scale of operations—measured by the number and significance of victims—before and after interventions for groups that continued operations. Here, significance is defined as victims

from high technological sectors and/or critical infrastructure according to NIS2.

- **Proposition 6:** A multinomial logistic regression will assess the relationship between an intervention, uptime of a leakpage, the number of victims and possible rebranding either with or without split-up and with or without overlapping uptime of leak page servers of the original group and the rebranded group.

A p-value of 0.05 or lower indicates that the variable significantly predicts the dependent variable at a significance level of $\alpha = 0.05$. Given the limited number of observed interventions, the statistical power of these tests is likely to be low. While conducting these tests could provide explorative insights regarding our propositions, the results should be interpreted cautiously due to the increased risk of Type I errors (false positives) and Type II errors (false negatives).

9.4 Results of Analysis

In this section, we explore the group characteristics influencing the likelihood of a LE intervention. Subsequently, we outline the nature of the interventions carried out against the ransomware groups. Finally, we will conclude the section with an analysis of the reaction of the groups on the LE intervention.

9.4.1 Ransomware Groups Facing an LE Intervention

An overview of the descriptive statistics can be found in Table 9.11. Victims were reported across 156 countries and 309 sectors. Top sectors included Construction (662 observations), Law Practice (384 cases), and Hospitals and Health Care (378 victims). Most victims were from the United States, with 5,783 victims (47,7%). After normalizing for GDP [92], most countries appear to be relatively evenly affected, except for Canada (30.6%) and India (5.3%). Although the number of victims from critical (NIS) and technologically intensive (Tech) sectors is comparable to that of other countries, the percentage of data leaked on leak pages in the U.S. is lower at 37.2%, compared to 40-45% in other countries. This would suggests that the companies from the U.S. are more willing to pay, assuming that their data is less frequently published.

A summary of the results of the logistic regression analysis to address **Proposition 1** is shown in Table 9.3. The analysis revealed several key findings regarding the impact of various factors on the probability of a ransomware group being targeted by a LE intervention. Firstly, groups that attack a large number of victims and mainly target large companies among have a much higher

likelihood of facing a LE intervention then groups that make fewer victims and focus on smaller companies. Conversely, groups that attacked organisations belonging to Network Information Systems (NIS), to the technology sector, (Tech), whether or not they data was leaked and published on the leak page, and the amount of victims from the USA were not experiencing more LE interventions. Moreover, the total amount of time a group was active decreased the probability of intervention. Taken together, the evidence suggests support for **Proposition 1**, as a ransomware group's likelihood of facing intervention seems to rise with the number of victims, especially when those victims are substantial in size.

Although ransomware groups with many victims in the USA might not have a higher probability of facing interventions, US law enforcement could be more frequently involved with interventions compared to other countries. Table 9.11 shows that the USA is involved in 20 out of 29 interventions. Similarly, LE in other top 10 most frequently attacked countries, such as France, Germany, Canada, Spain, and the UK, is also very active against ransomware groups (Table 9.11). While Ukraine is not among the top 10 most targeted countries, it might be involved in many interventions since any arrests in Ukraine require the assistance of Ukrainian LE.

Table 9.2: Descriptive Statistics of Leak Page Dataset: Frequency of Attacks, Frequency over GDP, Sector Importance (NIS and Tech), Company Size, and Data Leaked.

Country	**Freq**	**Freq/GDP** $\times 10^5$	**% NIS**	**% Tech**	**% Large Companies**	**% Data Leaked**
USA	5783 (47.7%)	24.7	47.3	44.2	26.7	37.2
UK	696 (5.7%)	24.5	48.3	40.9	26.9	43.5
Canada	608 (5.0%)	30.6	42.3	42.9	26.6	37.3
Germany	508 (4.2%)	13.5	46.7	51.0	43.9	40.6
France	494 (4.1%)	19.3	41.9	42.1	36.1	40.9
Italy	416 (3.4%)	22.1	50.7	46.9	25.8	35.6
Spain	260 (2.1%)	19.9	48.1	49.2	29.1	45.4
Australia	255 (2.1%)	16.4	46.7	40.0	20.1	44.3
Brazil	224 (1.8%)	12.7	46.0	44.2	53.6	46.0
India	168 (1.4%)	5.3	67.3	64.9	67.1	42.3
Other	2701 (22.3%)	X	48.9	50.6	47.1	41.8

Table 9.3: Logistic Regression Analysis of the Likelihood of Ransomware Groups Facing an Intervention.

Variable	Estimate	Std. Error	z-value	Pr($>|z|$)
Intercept	-1.57	0.50	-3.13	0.002*
Total victims	0.19	0.09	2.12	0.034*
NIS count	-0.26	0.21	-1.22	0.224
Tech count	-0.22	0.19	-1.17	0.244
Uptime leakpage mean	-0.01	0.01	-2.44	0.015*
Data leak count	-0.13	0.07	-1.84	0.066
Large company count	0.45	0.20	2.27	0.023*
USA	-0.04	0.09	-0.43	0.667

To address **Proposition 2**, we conducted a Spearman's correlation test, which was also significant. The test reveals a moderate, positive correlation ($\rho = 0.437, p < 0.001$) between the number of victims and the number of interventions, supporting **Proposition 2** that countries that suffer a relatively high level of victimization correlates are also involved in more LE interventions.

9.4.2 Actions Of Ransomware Groups After Intervention

We begin this section by examining the interventions we identified during the data collection process and providing examples of exactly what happened. We refer to ransomware groups by their name, for example 'Cl0p', 'Doppelpaymer', etc.

Intervention 1: Arrests. The ransomware group 'Cl0p' faced arrest of six persons and equipment seized on June 1, 2021, allegedly involving the part of the group responsible for money laundering [10]. They continued operations until the end of our data collection period. The arrests begin when Cl0p breached four South Korean companies in 2019. The 'Doppelpaymer' group faced an arrest on February 28, 2023. The last victim that they put online on their leak page was in September 2021; the group allegedly rebranded before [12]. LE in Germany arrested one person, together with the Ukrainian police. Both police forces also seized equipment. In addition, arrest warrants for three important figures in the group were issued. The

Table 9.4: Summary of 29 interventions by country, with multiple countries involved in some interventions. Top 10 countries are shown; others are grouped as 'Other'.

Country	Arrest	Sanction	Crypto	Decryptor	Takedown	Multiple Interventions	Total
USA	6	5	1	1	2	5	20
UK	1	3	0	0	0	4	8
France	2	0	0	0	1	4	7
Germany	3	0	0	0	0	4	7
Netherlands	2	0	1	0	0	3	6
Ukraine	4	0	0	0	1	1	6
Sweden	1	0	0	0	0	3	4
Canada	1	0	0	0	0	2	3
Australia	1	1	0	0	0	1	3
Spain	0	0	0	0	0	3	3
Other	10	4	2	3	5	24	48

same with 'Grief' ransomware group, which faced an arrest on February 28, 2023, whereas the last victim appeared on their leak page on March 2022. They probably rebranded before to 'NoEscape' [12]. Finally, 'REvil' faced arrests twice, on November 4, 2021, and January 14, 2022, and also apparently rebranded (at least partially) before to 'Blogxx', 'Spectre', and 'Ransom Cartel' [80, 61, 40]. The last victim of REvil was in October 2021 after LE intervention. 'Egregor' stopped after affiliates were arrested on February 10, 2021, in a collaborative operation of Ukraine and France LE [82]. France LE started the investigation apparently after complaints from the public over the ransomware gang. Finally, Lockbit faced an arrest of an affiliate on June 15, 2023, but continued its activities until the end of our dataset period [9].

Intervention 2: Sanctions. Sanctions typically involve travel restrictions, asset freezes, and/or arrest warrants [**usasanctions** , 31]. It is important to note that these actions were all initiated by LE. However, they are often implemented with considerable delays, frequently occurring after the targeted ransomware group has already ceased operations or undergone rebranding. For instance, sanctions against 'BlogXX' and 'Babuk' were imposed

well after these groups had ceased publishing victims on leak pages, with more than a year passing before the individuals behind these operations were sanctioned [16, 83]. The imposition of sanctions against 'Babuk' may be linked to the public interview conducted with Babuk [16]. Similarly, sanctions against 'Conti' and 'REvil' were implemented after these groups had rebranded. In the case of 'Conti', sanctions were imposed half a year to one year after the group ceased operations [49, 35], while sanctions against REvil were enacted one month after the group stopped [28].

Intervention 3: Crypto-asset freezing. The freeze of crypto of the group 'DarkSide' occurred in the aftermath of the Colonial Pipeline attack, and after the group had already rebranded to 'BlackMatter' and/or 'BlackCat' [44]. Another crypto freeze involved the seizure of Cl0p assets in connection with the attack on Maastricht University [25]. It is worth mentioning that as a result of the freeze, Maastricht University received a refund of the ransom they had paid, and generated a significant profit due to the increased value of Bitcoin. 'Cl0p', however, continued their activities after this intervention.

Intervention 4: Decryptor release. 'Egregor' discontinued its ransomware operations before its creators distributed a decryptor, attributing the decision to the arrests of REvil members [60]. 'Avaddon' continued for 5 months after the decryptor became publicly available [58]. The 'BlackBasta' decryptor was known by December 30, 2023, but the group continued operations afterwards. Possibly, groups continuing operations after a decryptor becomes available change their ransomware malware [33]. After the 'REvil' decryptor became known on September 16, 2021, the USA assisted in its release [54]. A month later, on October 16, 2021, they released their last victim and rebranded [80]. While Bitdefender could not share details about how they obtained the master decryption key or the law enforcement agency involved, they informed BleepingComputer that it works for all 'REvil' victims encrypted before July 13th 2021. The Maze decryptor was released on February 9, 2022, while their latest victim was mentioned on the group's leak page on December 15, 2020. Allegedly, they published the decryptor released by their own makers, indicating a link to the 'REvil' arrests [60]. The Prometheus decryptor became known on August 1, 2021. A month later, on September 14, 2021, the last victim of the group was mentioned in the leak page data. The malware has a weak random number generator, which made a decryptor possible. Initially the Prometheus malware was based on Thanos ransomware, it later evolved into Spook, but they ceased operations on October 26, 2021 [6].

Intervention 5: Takedown of the leak page infrastructure. DarkSide experienced a takedown of their leak page infrastructure on May 13, 2021, although it remains unclear whether law enforcement was involved or if the group self-initiated the takedown to rebrand and mitigate the risk of law enforcement action [1]. Egregor was taken down on February 16, 2021, by the combined efforts of LE in the USA, France, and Ukraine. Following the takedown, the site remained offline, and associates deactivated their forum profiles [4]. Similarly, REvil was taken down on October 21, 2021. However, given that their last victim appeared on the group's leak page on October 16, this suggests that they already ceased their operations before the takedown. This takedown was initiated by the United States LE in response to REvil's significant Kaseya attack. Additionally, REvil's servers were reportedly hacked by the United States LE earlier in the same year [77]. Lastly, the takedown of Trigona was not conducted by LE but by the Ukrainian Cyber Alliance, an activist group targeting Russian hacker groups due to the Russian-Ukraine war [88].

6. Multiple interventions. There are five LE interventions that consisted of multiple actions . For instance, a takedown was combined with an arrest, decryptor, or both. AlphVM/Blackcat, which was the target of a joint operation involving LE of the USA, Germany, Denmark, Australia, UK, Spain, Switzerland, and Austria, underwent a takedown, followed by the subsequent release of a decryptor. Despite some fluctuations in website availability, the group continued its operations, with the last victim recorded on March 4, 2023 [85]. Similarly, Lockbit3.0 faced a takedown and decryptor release through coordinated efforts by LE of multiple countries including France, Germany, the Netherlands, Sweden, Australia, Canada, Japan, the UK, USA, and Switzerland [48]. Despite these actions, Lockbit3.0 persisted in publishing victims on their leak pages, remaining active throughout our observation period. The takedown and arrest of Netwalker and Ragnar Locker on January 27, 2021, and October 11, 2023, respectively, were successful, meaning that no further victims were reported on the groups leak page or on other security blogs post-intervention [53, 76]. Likewise, Hive, targeted on January 26, 2023, experienced a takedown, a decryptor release, and arrests through coordinated actions involving the LE of 13 countries [78]. Subsequently, there was no further activity from Hive on the leak page.

It is important to note that we encountered several events occurring across multiple groups, which could have potentially impacted a groups' decisions to

Table 9.5: The actions of ransomware groups in response to various interventions, which includes interventions occurring prior to group stopping (STOP BEFORE) or rebranding (REBRAND BEFORE).

Intervention	STOP BEFORE	REBRAND BEFORE	CONTINUE	STOP	REBRAND	Total
Arrest	0	4	2	1	0	7
Crypto Freeze	1	0	1	0	0	2
Decryptor	2	0	2	0	2	6
Sanction	1	3	0	1	0	5
Takedown	0	1	1	2	0	4
Takedown, Arrest	0	0	0	2	0	2
Takedown, Decryptor	0	0	1	0	0	1
Takedown, Decryptor, Arrest	0	0	1	1	0	2
Total	**4**	**8**	**8**	**7**	**2**	**29**

cease operations or undergo rebranding. These events included internal disputes, self-shutdowns, and public interviews. We describe these events below.

Two groups experienced an internal dispute that resulted in leaks of private communications. These leaks became known as the Conti leaks and the Yanluowang Leaks [37, 52]. The Conti leaks ensued after a conflict over a public statement indicating Conti's support for Russia in the Russia-Ukraine war [41]. Yanluowang, consisting of 18 members, of which 5 were active, and had chats that were exposed by a group member, revealing plans to target critical infrastructure, excluding those from the Soviet Union [37].

Some groups publicly announced they ceased operations, sometimes combined with the release of decryption keys. This is also labelled as 'self-shutdowns'. Sometimes this self-shutdown is combined with an 'exit scam', in which ransomware groups state they have been arrested by LE, with the hidden aim to keep the profit share of affiliates to themselves [63]. We observed self-shutdowns of File Leaks, AstroLocker, Ragnarok, BlackMatter, and Avaddon. Two groups, File Leaks and AstroLocker, underwent rebranding after a self-shutdown [69, 6]. Avaddon possibly rebranded after 2.5 years to NoEscape [40].

Furthermore, some ransomware actors grant interviews [45, 16], possibly driven by a desire to establish a reputation or a perception of invincibility against arrest [45, 55]. Typically, these interviews are conducted anonymously. One notable exception is the interview with Wazawaka, who provided insights into ransomware attacks of Babuk ransomware that only the perpetrator could pos-

sess [16]. This interview revealed the identity of Wazawaka and might therefore have an impact on the continuation of ransomware operations of Babuk.

There are several considerations regarding the labeling of interventions. Firstly, it's important to highlight that two groups, Hive and Netwalker, potentially underwent rebranding some time after the intervention, with Hive rebranding after nine months. According to the Hive operators, they sold their ransomware malware to Hunters International, but they kept operating independently as two separate groups. We decided that this incident is no rebranding event because there appear to be two different groups.

Secondly, the arrest of an actor associated with Doppelpaymer/Grief/Entropy is treated as a single intervention due to the multiple rebrandings preceding the arrests, indicating a complex scenario where all three groups were linked to the same attack.

Thirdly, out of the 25 groups targeted by interventions, seven experienced multiple interventions over time, with REvil facing the highest number of interventions (five in total) before rebranding. It's noteworthy that law enforcement's decryptor capabilities, as claimed in some cases like Lockbit3.0, appear to be relatively limited.

Reviewing the outcomes of various interventions (see Table 9.5), we observe that prior to any intervention, 4 groups stopped victim publication (STOP BEFORE), while 8 groups rebranded before an intervention (REBRAND BEFORE). 8 ransomware groups continued their activities post-intervention (CONTINUE). Additionally, 7 groups ceased operations post-intervention (STOP), and 2 groups rebranded after the intervention.

Table 9.6: Summary of Ransomware Group Statistics by Intervention Type.

Intervention Type	Freq	Mean Victims	Mean Intervention Time	Mean Uptime Leak Page	Δ Uptime - Intervention Time
Arrest	7	468	399	636	237
Sanction	5	425	727	505	-222
Crypto	2	316	379	671	292
Decryptor	6	243	316	348	32
Takedown	4	171	197	399	202
Multiple interventions	5	540	604	622	18
Interview, Dispute, Shutdown	10	373	442	551	109

To address **Proposition 3**, a binomial test was conducted to evaluate the effectiveness of interventions on ransomware groups ceasing operations. With 7 cases where groups ceased operations out of 17 total interventions where groups did not already stop or rebrand before the intervention. The test was significant ($p < 0.001$). However, rebranding was with 2 out of 17 cases not significantly different from 0 ($p = 0.2078$), which implies groups do not rebrand after an intervention, contradicting **Proposition 5**.

In addition to examining post-intervention behaviors of ransomware groups, it is interesting to better understand the dynamic between ransomware group characteristics and type of interventions. Table 9.6 offers an overview of statistics of intervention type and ransomware group characteristics. Notably, the analysis helps improve our understanding of intervention strategies and ransomware group actions. For example, sanctions (such as travel restrictions and/or assets freeze) and arrests typically occur after a group has ceased operations, reflecting the time it takes for law enforcement to identify suspects.Ransomware groups facing arrests and sanctions tend to have the largest average amount of victims, as observed among those subjected to multiple interventions. Taken together, these results imply that arrests and sanctions take more effort from law enforcement, and are used against ransomware groups that claim large of victims. Additionally, the decryptor intervention yields minimal differences between uptime and intervention time, suggesting that many ransomware groups rebrand or cease operations if a decryptor is available. Here, intervention time is the time between a group published there first victim and the LE intervention. Finally, takedowns is associated with an intervention time of approximately 197 days, despite groups continuing operations for roughly another 200 days thereafter, implying a relatively straightforward intervention process for law enforcement.

Table 9.7: Comparison of Victim Characteristics of Ransomware Groups Who Continue Before and After Interventions

Ransomware Group	Intervention	Victims Before Intervention	Victims After Intervention	% Large companies before Intervention	% Large companies after Intervention	% NIS before Intervention	% NIS after Intervention	% Tech before Intervention	% Tech after Intervention	% USA before Intervention	% USA after Intervention
AlphVM	Takedown, Decryptor	687	64	38.7	33.9	50.5	58.1	20.4	25.8	50.1	57.8
Avaddon	Decryptor	23	173	22.2	23.8	33.3	54.3	16.7	21.0	69.6	36.0
BlackBasta	Decryptor	382	42	34.5	22.5	43.0	40.0	20.7	15.0	57.5	53.7
CL0P	Crypto	131	399	40.3	63.7	60.5	64.5	24.8	31.4	62.6	61.0
CL0P	Arrest	66	464	66.7	56.8	70.8	62.5	26.2	30.3	50.0	63.0
LockBit 3.0	Takedown, Decryptor, Arrest	1574	19	28.1	47.4	50.0	47.4	20.4	21.1	35.9	63.2
LockBit 3.0	Arrest	902	691	29.4	27.2	49.2	50.9	20.9	19.9	33.2	40.2
REvil	Decryptor	312	5	33.9	40.0	45.1	20.0	17.5	20.0	60.5	20.0
Trigona	Takedown	35	13	18.2	18.2	42.4	50.0	15.2	25.0	45.7	33.3

9.4.3 Crime Displacement After Intervention

In this subsection we will first compare the scale of the operations of ransomware groups, before and after a LE intervention. Subsequently, we will examine the relationship between law enforcement interventions and specific types of rebranding.

Firstly, ransomware groups may alter their targeting strategy following an intervention, potentially opting to target fewer victims or shifting focus away from critical infrastructure to mitigate the risk of further interventions or Law Enforcement attention. See Table 9.7 for overview of groups who continue after an LE intervention.

To evaluate **Proposition 4**, we conducted a paired t-test, revealing no significant differences in the mean values of the number of victims ($p = 0.143$), $\%$ of large companies ($p = 0.378$), $\%$ NIS ($p = 0.856$), and $\%$ USA ($p = 0.492$). However, only for $\%$ tech companies, the paired t-test yielded a significant result ($p = 0.039$), meaning that percentage of tech companies before is larger than amount of tech companies after intervention. Additionally, to assess the robustness of these findings, we employed a non-parametric Wilcoxon signed-rank test. The results of this test resembled those of the paired t-test, indicating no significant differences, before and after the LE intervention in the number of victims ($p = 0.160$), $\%$ large companies ($p = 0.441$), $\%$ NIS ($p = 0.695$), and $\%$ USA ($p = 0.625$). Yet, for $\%$ tech companies, the Wilcoxon test is on the verge of statistical significance with p= ($p = 0.064$) with $\alpha = 0.05$. We conclude that, be-

Table 9.8: Names of Ransomware Groups Who Rebranded Categorized by Overlap and Split-up

Overlap	No split-up	Split-up
No Overlap	**Group 1** Avaddon, Cuba, Babuk, Darkside, Hive, RansomHouse, Nefilim, Prometheus	**Group 2** Conti
Overlap	**Group 3** DoppelPaymer, Haron, Lockbit1.0, Lockbit2.0, Vice Society	**Group 4** Maze

sides a decrease of tech companies after intervention, we do not have sufficient statistical evidence to support **Proposition 4**.

The second issue is rebranding. Ransomware groups may choose to rebrand and adopt a different strain, which could represent a form of crime displacement. We identify four types or groups of rebranding in our dataset, see Table 9.8.

To address **Proposition 6**, a multinomial logistic regression was performed to assess the relationship between interventions, leak page uptime, and the number of victims in relation to different types of rebranding: with or without split-up and with or without overlapping uptime. Despite the limited number of observations in the different groups (see Table 9.8), the model revealed a significant relationship between intervention and Group 1 (no split-up and no overlap) ($p < 0.001$). The other variables were not significant. This result is not congruent with **Proposition 6**, as Group 1 has no split-up. However, given that only two ransomware groups split up after rebranding, these findings may be attributed to the limited number of observations.

9.5 Discussion and Conclusion

In this chapter, our primary objective was to investigate the response of ransomware groups to law enforcement interventions. To achieve this, we formulated three research questions.

RQ1 aimed to understand the factors influencing the probability of a law enforcement intervention. We found that ransomware group characteristics such as total amount of victims, uptime of the leakpage, and the presence of large companies significantly impacted intervention probability. Additionally, law enforcement was more active in countries heavily affected by ransomware attacks. However, other factors like victim count of critical infrastructure, technological intensive sectors and data leakage did not affect the likelihood of ransomware groups being targeted by law enforcement.

RQ2 aimed to understand how ransomware actors respond to various law enforcement interventions, including arrests, sanctions, crypto-asset freezes, decryptors, and takedowns. Post-intervention, 8 out of 17 groups continued operations, 7 groups ceased operations, and 2 groups rebranded. We conclude that law enforcement interventions significantly impact ransomware operations, aligning with Situational Crime Prevention theory, where interventions increase efforts and risks while decreasing profits for ransomware groups.

RQ3 aimed to understand crime displacement. We found that ransomware groups typically do not rebrand after an intervention. However, there was limited evidence suggesting that groups continuing operations post-intervention change

the type or number of victims they target, including a decreased number of victims from technological intensive sector. Additionally, interventions were linked to rebranding characterized by no overlap between the old and new group's leak page uptime and no split-up into multiple new groups.

Our exploratory analysis suggests that arrests and sanctions may correlate with ransomware groups having a high victim count, with law enforcement taking longer to intervene from the time the first victim is published. Different kinds of interventions appeared to have specific consequences. The presence of a decryptor was linked to shorter leak page uptimes post-intervention. Takedowns of leak pages were associated with fewer victims and quicker intervention times. Considering that 2 out of 4 ransomware groups ceased activity following takedowns, this approach could be seen as a cost-effective intervention strategy against ransomware.

9.6 Limitations and Further Work

There are different limitations of this study:

1. **Causality.** Drawing causal conclusions from observational data presents challenges [91, 62, 90]. Ransomware groups may differ in various ways beyond facing interventions. Nonetheless, as emphasized by [91], an overly strict focus on the 'causation versus correlation distinction' can be limiting, as even randomized control experiments do not always provide watertight evidence. Furthermore, there are legal and ethical challenges conducting randomized trials with law enforcement interventions. Therefore, we argue this paper is a best effort of understanding the relationship between interventions and action of ransomware groups.

2. **Low sample size.** Due to low sample size of interventions, it is hard to draw definitive conclusions because the statistical tests that were used do not have that much statistical power. Furthermore, it makes an analysis more prone to measurement errors, to the volatility or special circumstances of specific groups.

3. **Biased intervention list.** Our list of interventions may be biased due to the 'searchlight effect' [91], wherein interventions are more likely to be found in areas that are actively searched, potentially overlooking others. For example, the absence of Chinese or Japanese-speaking authors may hinder the identification of interventions from these regions. Additionally, some law enforcement interventions may not be publicly disclosed. To mitigate

this bias, future research could explore alternative search engines from different countries and continents.

Further research could explore the costs associated with different types of law enforcement and government agency interventions, both material and immaterial [67]. This would enable cost-benefit analyses to evaluate the effectiveness and efficiency of various intervention strategies and compare them with preventive interventions, which might be more cost-effective [8, 51, 50, 81]. Additionally, investigating the impact of perceived attacker reputation on ransomware groups' decisions to rebrand or cease operations could provide valuable insights. Attacker reputation usually refers to the perceived likelihood of receiving a decryption key after payment [13]. Examining how law enforcement interventions affect attacker reputation from both the victim's and affiliate's perspectives could offer a broader understanding of intervention effectiveness.

In conclusion, this study is the first to evaluate ransomware interventions using data from victims published on leak pages after double-extortion ransomware attacks. Despite its limitations, we believe this study represents a step in the right direction for policymakers and law enforcement agencies worldwide to make more evidence-based decisions regarding law enforcement interventions.

9.7 Policy Recommendations

This study provides insights for policymakers and law enforcement on the effectiveness of ransomware interventions. The results suggest the following:

Policy Implication 1: Emphasize Frequency over Scale. Based on our findings, increasing the frequency of interventions might be disruptive. Smaller, frequent actions might significantly pressure malicious actors, contrary to the expert belief that only major takedowns or arrests are effective.

Policy Implication 2: Maintain Unpredictability. Vary and randomize interventions to counter ransomware groups' adaptive methods. Use the different types of interventions discussed in this chapter as inspiration. Focusing on implementing Situational Crime Prevention principles—such as Increasing Effort, Increasing Risks, Reducing Rewards, Reducing Provocations, and Removing Excuses—can enhance effectiveness [2, 20, 26, 17, 50, 43].

While our study indicates smaller interventions can be effective, more controlled studies are needed. The discrepancy between our findings and expert opinions underscores the need for further research to refine these recommendations.

9.8 Ethics

We follow the principles from Menlo Report [3] to justify the ethical considerations made in this chapter:

Respect for Persons: Prioritizing privacy and confidentiality, data was aggregated at country and sector levels to safeguard the privacy of victims.

Beneficence: While there is a possibility that providing information about interventions may aid criminals in altering their actions, we believe our approach ultimately aids law enforcement in combating ransomware. We estimate that the overall impact of our study is positive.

Justice: All ransomware attacks included in the study were afforded equal opportunity, without bias towards specific entities. Selection criteria were based solely on the presence of ransomware-related keywords.

Respect for Law and Public Interest: Information pertaining to law enforcement operations and government interventions was handled discreetly. Our study aims to offer valuable insights into the effectiveness and cost-benefit of law enforcement interventions against ransomware, thereby assisting law enforcement in making well-informed decisions when planning interventions.

Appendix A

This appendix provides detailed information about the ransomware strains, their interventions, and the frequency of ransomware incidents by country. The following tables summarize the key aspects of the dataset used in this chapter Table 9.9, 9.10, and 9.11.

Table 9.9: List of Ransomware Strains and Interventions

ID	Strain	Event	Date intervention	Date last victim
1	AlphVM	Takedown, Decryptor	19/12/2023	01/03/2024
2	Darkside	Takedown	13/05/2021	13/05/2021
3	Egregor	Takedown	16/02/2021	10/02/2021
4	HiveLeaks	Takedown, Decryptor, Arrest	26/01/2023	26/01/2023
5	NetWalker	Takedown, Arrest	27/01/2021	27/01/2021
6	RagnarLocker	Takedown, Arrest	16/10/2023	11/10/2023
7	REvil	Takedown	21/10/2021	16/10/2021
8	Trigona	Takedown	21/10/2023	01/03/2024
9	LockBit 3.0	Takedown, Decryptor, Arrest	20/02/2024	01/03/2024
10	CL0P	Arrest	01/06/2021	26/02/2024
11	DoppelPaymer	Arrest	28/02/2023	17/09/2021
12	Egregor	Arrest	10/02/2021	10/02/2021
13	Grief	Arrest	28/02/2023	24/03/2022
14	LockBit 3.0	Arrest	15/06/2023	01/03/2024
15	REvil	Arrest	04/11/2021	16/10/2021
16	REvil	Arrest	14/01/2022	16/10/2021
17	Avaddon	Shutdown	11/06/2021	11/06/2021
18	BABUK	PublicInterview	26/08/2022	26/02/2021
19	BlackBasta	Decryptor	30/12/2023	01/03/2024
20	BlackMatter	Shutdown	01/11/2021	04/11/2021
21	Conti	InternalDispute	27/02/2022	22/06/2022
22	Darkside	Crypto	07/06/2021	13/05/2021
23	MAZE	Decryptor	09/02/2022	15/12/2020
24	Prometheus	Decryptor	01/08/2021	14/09/2021
25	Ragnarok	Shutdown	26/08/2021	26/08/2021
26	REvil	Decryptor	16/09/2021	16/10/2021
27	Avaddon	Decryptor	15/01/2021	11/06/2021
28	BABUK	Sanction	16/05/2023	26/02/2021
29	Conti	Sanction	09/02/2023	22/06/2022
30	Conti	Sanction	07/09/2023	22/06/2022
31	REvil	Sanction	08/11/2021	16/10/2021
32	Yanluowang	InternalDispute	31/10/2022	31/10/2022
33	File Leaks (SYNack)	Shutdown	15/08/2021	15/08/2021
34	AstroLocker	Shutdown	04/07/2022	09/06/2021
35	BlogXX	Sanction	23/01/2024	06/01/2023
36	Egregor	Decryptor	09/02/2022	10/02/2021
37	CL0P	Crypto	02/07/2022	01/03/2024

Table 9.10: Strain Rebranding

Initial Strain	**Rebrand strain 1.0**	**Rebrand strain 2.0**
Avaddon	NoEscape	
Cuba	IndustrialSpy	
Babuk	Payload.bin	
Conti	3AM, Akira, Blackbastsa, BlackByte, MountLocker, Karakurt	
	Royal	Blacksuit
	XingLocker	Quantum
Darkside	Blackmatter	Blackcat
Doppelpaymer	Grief	
Haron	Midas	
Hive	Hunters International	
Ransomhouse	8Base	
Lockbit1.0	Lockbit2.0	Lockbit3.0
Revil	LV	
Nefilim	Nokoyawa	
Prometheus	Spook	
Maze	Suncrypt, Egregor	
Vice Society	Rhysida	

Table 9.11: Summary of country frequencies for ransomware victims

ID	Country	Freq	ID	Country	Freq	ID	Country	Freq
1	United States	5783	54	Hungary	14	107	Iran, Islamic Republic of	2
2	United Kingdom	696	55	Puerto Rico	14	108	Isle of Man	2
3	Canada	608	56	Venezuela	14	109	Madagascar	2
4	Germany	508	57	Dominican Republic	13	110	Maldives	2
5	France	494	58	Ecuador	13	111	Monaco	2
6	Italy	416	59	Finland	13	112	Myanmar	2
7	Spain	260	60	Guatemala	13	113	North Macedonia	2
8	Australia	255	61	Kenya	12	114	Saint Kitts and Nevis	2
9	Brazil	224	62	Angola	10	115	Seychelles	2
10	India	168	63	Jamaica	10	116	Ukraine	2
11	Switzerland	141	64	Morocco	10	117	Virgin Islands, U.S.	2
12	Unknown	137	65	Pakistan	10	118	Zimbabwe	2
13	Netherlands	130	66	Panama	10	119	Antigua and Barbuda	1
14	Mexico	117	67	Slovakia	10	120	Belize	1
15	Belgium	109	68	Bangladesh	9	121	Bermuda	1
16	Japan	104	69	Croatia	8	122	Brunei	1
17	Thailand	91	70	Cyprus	8	123	Burkina Faso	1
18	Austria	88	71	Nigeria	8	124	Cayman Islands	1
19	Taiwan	86	72	Trinidad and Tobago	8	125	Curacao	1
20	China	85	73	Uruguay	8	126	Democratic Republic of the Congo	1
21	United Arab Emirates	79	74	Iran	7	127	Ethiopia	1
22	South Africa	73	75	Oman	7	128	French Guiana	1
23	Argentina	70	76	Tunisia	7	129	Gambia	1
24	Israel	70	77	Jordan	6	130	Ghana	1
25	Sweden	69	78	Lithuania	6	131	Gibraltar	1
26	Hong Kong	65	79	Serbia	6	132	Greenland	1
27	Singapore	63	80	Sri Lanka	6	133	Guernsey	1
28	Turkey	62	81	Bahrain	5	134	Guyana	1
29	Indonesia	58	82	Namibia	5	135	Honduras	1
30	Portugal	54	83	Nicaragua	5	136	Iceland	1
31	Colombia	51	84	Senegal	5	137	Iraq	1
32	Malaysia	47	85	Bahamas	4	138	Jersey	1
33	Poland	38	86	Barbados	4	139	Kazakhstan	1
34	Philippines	37	87	Bolivia, Plurinational State of	4	140	Libya	1
35	Denmark	36	88	Bosnia and Herzegovina	4	141	Liechtenstein	1
36	Peru	35	89	Botswana	4	142	Macedonia, Republic of	1
37	Chile	33	90	Estonia	4	143	Mali	1
38	New Zealand	33	91	Slovenia	4	144	Malta	1
39	Saudi Arabia	32	92	Tanzania	4	145	Moldova	1
40	South Korea	32	93	Algeria	3	146	Mongolia	1
41	Czech Republic	31	94	Cameroon	3	147	Montenegro	1
42	Vietnam	30	95	Cuba	3	148	Palestine	1
43	Egypt	27	96	El Salvador	3	149	Papua New Guinea	1
44	Ireland	27	97	Fiji	3	150	Russia	1
45	Norway	27	98	Haiti	3	151	Sint Maarten	1
46	Romania	27	99	Ivory Coast	3	152	Syria	1
47	Bulgaria	23	100	Latvia	3	153	Tonga	1
48	Greece	23	101	Mauritius	3	154	Uzbekistan	1
49	Kuwait	18	102	Paraguay	3	155	Vanuatu	1
50	Luxembourg	18	103	Uganda	3	156	Virgin Islands, British	1
51	Costa Rica	16	104	Albania	2	157	Zambia	1
52	Lebanon	16	105	Czechia	2			
53	Qatar	16	106	Gabon	2			

Bibliography

[1] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. V. Eeten, M. Levi and S. Savage. 'Measuring the cost of cybercrime'. *The Economics of Information Security and Privacy*. Berlin, Heidelberg: Springer, 2013, pp. 265–300.

[2] M. Bada, A. Hutchings, Y. Papadodimitraki and R. Clayton. *An evaluation of police interventions for cybercrime prevention*. Tech. rep. UCAM-CL-TR-983. University of Cambridge, Computer Laboratory, 2023.

[3] M. Bailey, D. Dittrich, E. Kenneally and D. Maughan. 'The Menlo Report'. *IEEE Security & Privacy* 10.2, 2012, pp. 71–75.

[4] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak and M. K. Khan. 'Ransomware: Recent advances, analysis, challenges and future research directions'. *Computers & Security* 111, 2021, p. 102490.

[5] J. Blatchly. 'The Impact of Ransomware–A Comparison of Worldwide Governmental Policies and Recommendations for Future Directives'. PhD thesis. Utica University, 2023.

[6] H. Borrion and L. Y. Connolly. *Your money or your business: Decision-making processes in ransomware attacks*. 2020.

[7] H. Borrion, H. Dehghanniri and Y. Li. 'Comparative analysis of crime scripts: One CCTV footage—twenty-one scripts'. *Proceedings of the 2017 European Intelligence and Security Informatics Conference (EISIC)*. 2017.

[8] R. Brewer, M. de Vel-Palumbo, A. Hutchings, T. Holt, A. Goldsmith, D. Maimon and D. Maimon. 'Situational crime prevention'. *Cybercrime Prevention: Theory and Applications*. 2019, pp. 17–33.

[9] S. V. Buuren and K. Groothuis-Oudshoorn. 'mice: Multivariate imputation by chained equations in R'. *Journal of Statistical Software* 45, 2011, pp. 1–67.

[10] A. Cartwright and E. Cartwright. 'Ransomware and reputation'. *Games* 10.2, 2019, p. 26.

[11] A. Cartwright, E. Cartwright, J. MacColl, G. Mott, S. Turner, J. Sullivan and J. R. Nurse. 'How cyber insurance influences the ransomware payment decision: theory and evidence'. *The Geneva Papers on Risk and Insurance-Issues and Practice* 48.2, 2023, pp. 300–331.

[12] A. Cartwright, E. Cartwright, L. Xue and J. Hernandez-Castro. 'An investigation of individual willingness to pay ransomware'. *Journal of Financial Crime*, 2022. Ahead-of-print.

[13] E. Cartwright, J. H. Castro and A. Cartwright. 'To pay or not: game theoretic models of ransomware'. *Journal of Cybersecurity* 5.1, 2019, tyz009.

[14] Centraal Bureau voor de Statistiek. *Meer bedrijven met bedrijfsopleidingen*. `https://www.cbs.nl/nl-nl/nieuws/2017/28/meer-bedrijven-met-bedrijfsopleidingen/bedrijfsgrootte`. Accessed: 2024-03-27. 2017.

[15] Chainalysis. *Crypto Crime Report 2024*. `https://go.chainalysis.com/2024/crypto-crime-report.html`. Accessed: 2024-03-27. 2024.

[16] F. Changyong, W. Hongyue, L. Naiji, C. Tian, H. Hua and L. Ying. 'Log-transformation and its implications for data analysis'. *Shanghai Archives of Psychiatry* 26.2, 2014, p. 105.

[17] R. V. Clarke. 'Situational crime prevention'. *Environmental Criminology and Crime Analysis*. Ed. by R. Wortley and L. Mazerolle. London, UK: Willan, 2008, pp. 178–194.

[18] R. V. Clarke. 'Situational crime prevention: Its theoretical basis and practical scope'. *Crime and justice* 4, 1983, pp. 225–256.

[19] R. V. Clarke. 'Situational crime prevention: Theoretical background and current practice'. *Handbook on crime and deviance*. Ed. by M. D. Krohn, A. J. Lizotte and G. P. Hall. New York, NY: Springer New York, 2009, pp. 259–276.

[20] R. V. Clarke and J. E. Eck. *Crime Analysis for Problem Solvers in 60 Small Steps*. Retrieved from Washington D.C.: `http://www.popcenter.org/library/reading/PDFs/60steps.pdf`. 2005.

[21] J. Clough. 'A world of difference: the Budapest convention on cybercrime and the challenges of harmonisation'. *Monash University Law Review* 40.3, 2014, pp. 698–736.

[22] L. E. Cohen and M. Felson. 'Social change and crime rate trends: A routine activity approach'. *American Sociological Review* 44.4, 1979, pp. 588–608.

[23] L. Coles-Kemp and M. Theoharidou. 'Insider threat and information security management'. *Insider threats in cyber security*. Ed. by M. Bishop and A. N. Tehrani. Boston, MA: Springer US, 2010, pp. 45–71.

[24] A. Y. Connolly and H. Borrion. 'Reducing ransomware crime: analysis of victims' payment decisions'. *Computers & Security* 119, 2022, p. 102760.

[25] L. Y. Connolly, M. Lang, P. Taylor and P. Corner. *The Evolving Threat of Ransomware: From Extortion to Blackmail*. 2021.

[26] D. B. Cornish. 'Proceedings of the International Seminar on Environmental Criminology and Crime Analysis'. *Proceedings of the International Seminar on Environmental Criminology and Crime Analysis*. Tallahassee, FL: Florida Statistical Analysis Center, Florida Criminal Justice Executive Institute and Florida Department of law enforcement, 1994, pp. 30–45.

[27] D. B. Cornish and R. V. Clarke. *The reasoning criminal: Rational choice perspectives on offending*. Transaction Publishers, 2014.

[28] D. Cornish and R. Clarke. 'Crime specialisation, crime displacement and rational choice theory'. *Criminal behavior and the justice system*. Berlin, Heidelberg: Springer, 1989, pp. 103–117.

[29] C. Cosin. *Ecrime*. Retrieved March 1, 2023. 2022. URL: https://ecrime.ch/.

[30] H. Dehghanniri and H. Borrion. 'Crime scripting: a systematic review'. *European Journal of Criminology*, 2019.

[31] European Sanctions Map. https://sanctionsmap.eu/#/main. Accessed: 2024-03-20. n.d.

[32] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2021*. Tech. rep. Retrieved August 31, 2022. Luxembourg, 2021. URL: https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021.

[33] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2021*. Publications Office of the European Union, Luxembourg. Retrieved August 31, 2022, from https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021. 2021.

[34] Europol. *Internet Organised Crime Threat Assessment (IOCTA) 2023*. Tech. rep. Retrieved August 31, 2023. Luxembourg, 2023. URL: https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-assessment-iocta-2023.

[35] G. Falk. 'The influence of the seasons on the crime rate'. *J. Crim. L. Criminology & Police Science* 43, 1952, p. 199.

[36] M. Felson. 'Linking criminal choices, routine activities, informal control, and criminal outcomes'. *The reasoning criminal*. Routledge, 2017, pp. 119–128.

[37] M. Felson and R. V. Clarke. *Opportunity makes the thief*. Police Research Series, Paper 98. Pages 1-36. 1998.

[38] M. Felson and R. V. Clarke. *Opportunity makes the thief*. 1998.

[39] F. Galindo-Rueda and F. Verger. *OECD Taxonomy of Economic Activities Based on R&D Intensity*. Tech. rep. 2016/04. Paris: OECD Publishing, 2016.

[40] J. V. Ginkel, M. Linting, R. Rippe and A. van der Voort. 'Rebutting existing misconceptions about multiple imputation as a method for handling missing data'. *Journal of Personality Assessment* 102.3, 2020, pp. 297–308.

[41] I. W. Gray, J. Cable, B. Brown, V. Cuiujuclu and D. McCoy. 'Money Over Morals: A Business Analysis of Conti Ransomware'. *2022 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2022, pp. 1–12.

[42] R. T. Guerette and K. J. Bowers. 'Assessing the extent of crime displacement and diffusion of benefits: a review of situational crime prevention evaluations'. *Criminology* 47.4, 2009, pp. 1331–1368.

[43] P. H. Hartel, M. Junger and R. J. Wieringa. *Cyber-crime science = crime science + information security*. Tech. rep. TR-CTIT-10-34. CTIT, University of Twente, 2010.

[44] N. A. Hassan. *Ransomware revealed: a beginner's guide to protecting and recovering from ransomware attacks*. Apress, 2019.

[45] A. Haymore. *We Wait, Because We Know You. Inside the Ransomware Negotiation Economics*. Retrieved from https://research.nccgroup.com/2021/11/12/we-wait-because-we-know-you-inside-the-ransomware-negotiation-economics/. 2021.

[46] R. Heaton and S. Tong. 'Evidence-based policing: from effectiveness to cost-effectiveness'. *Policing: a journal of policy and practice* 10.1, 2016, pp. 60–70.

[47] J. Hernandez-Castro, A. Cartwright and E. Cartwright. 'An economic analysis of ransomware and its welfare consequences'. *Royal Society Open Science* 7.3, 2020, p. 190023.

[48] J. Hernandez-Castro, A. Cartwright and E. Cartwright. 'An economic analysis of ransomware and its welfare consequences'. *Royal Society Open Science* 7.3, 2020, p. 190023.

[49] S. Hilt and F. Merces. *Backing Your Backup*. 2022.

[50] H. Ho, R. Ko and L. Mazerolle. 'Situational Crime Prevention (SCP) techniques to prevent and control cybercrimes: A focused systematic review'. *Computers & Security* 115, 2022, p. 102611.

[51] T. Hodgkinson and G. Farrell. 'Situational crime prevention and Public Safety Canada's crime-prevention programme'. *Security Journal* 31, 2018, pp. 325–342.

[52] T. Hofmann. 'How organisations can ethically negotiate ransomware payments'. *Network Security* 2020.10, 2020, pp. 13–17.

[53] M. Humayun, N. Z. Jhanjhi, A. Alsayat and V. Ponnusamy. 'Internet of things and ransomware: Evolution, mitigation and prevention'. *Egyptian Informatics Journal* 22.1, 2021, pp. 105–117.

[54] A. Hutchings and T. J. Holt. 'A crime script analysis of the online stolen data market'. *British Journal of Criminology* 55.3, 2015, pp. 596–614.

[55] E. M. Hutchins, M. J. Cloppert and R. Amin. 'Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains'. *Leading Issues in Information Warfare & Security Research* 1, 2011, p. 80.

[56] A. S. Irwin and C. Dawson. 'Following the cyber money trail: Global challenges when investigating ransomware attacks and how regulation can help'. *Journal of money laundering control* 22.1, 2019, pp. 110–131.

[57] H. R. Jamali and S. Asadi. 'Google and the scholar: the role of Google in scientists' information-seeking behaviour'. *Online information review* 34.2, 2010, pp. 282–294.

[58] M. Junger, V. Wang and M. Schlömer. 'Fraud against businesses both online and offline: Crime scripts, business characteristics, efforts, and benefits'. *Crime Science* 9.1, 2020, pp. 1–15.

[59] D. Keatley. 'Crime Script Analysis'. *Pathways in Crime: An Introduction to Behaviour Sequence Analysis*. Cham: Springer International Publishing, 2018, pp. 125–136.

[60] M. Keshavarzi and H. Ghaffary. 'I2CE3: A dedicated and separated attack chain for ransomware offenses as the most infamous cyber extortion'. *Computer Science Review* 36, 2020, p. 100233.

[61] E. Kontopantelis, I. White, M. Sperrin and I. Buchan. 'Outcome-sensitive multiple imputation: a simulation study'. *BMC Medical Research Methodology* 17.1, 2017, pp. 1–13.

[62] C. C. Lanfear, R. L. Matsueda and L. R. Beach. 'Broken windows, informal social control, and crime: Assessing causality in empirical studies'. *Annual review of criminology* 3, 2020, pp. 97–120.

[63] P. Leo, Ö. Işik and F. Muhly. 'The Ransomware Dilemma'. *MIT Sloan Management Review*, 2022.

[64] R. Leukfeldt and E. E. Kleemans. 'Cybercrime, money mules and situational crime prevention: Recruitment, motives and involvement mechanisms'. *Criminal networks and law enforcement*. Routledge, 2019, pp. 75–89.

[65] Z. Li and Q. Liao. 'Ransomware 2.0: to sell, or not to sell a game-theoretical model of data-selling ransomware'. *Proceedings of the 15th International Conference on Availability, Reliability and Security*. 2020, pp. 1–9.

[66] A. Lubin. 'The Law and Politics of Ransomware'. *Vand. J. Transnat'l L.* 55, 2022, p. 1177.

[67] M. Manning, G. T. Wong, T. Graham, T. Ranbaduge, P. Christen, K. Taylor and P. Skorich. 'Towards a 'smart' cost–benefit tool: using machine learning to predict the costs of criminal justice policy interventions'. *Crime Science* 7, 2018, pp. 1–13.

[68] S. R. Matthijsse, M. S. van 't Hoff-de Goede and E. R. Leukfeldt. 'Your files have been encrypted: A crime script analysis of ransomware attacks'. *Trends in Organized Crime*, 2023, pp. 1–27.

[69] P. H. Meland, Y. F. F. Bayoumy and G. Sindre. 'The Ransomware-as-a-Service economy within the darknet'. *Computers & Security* 92, 2020, p. 101762.

[70] T. Meurs and L. Holterman. *Whitepaper data-exfiltratie bij een ransomware-aanval*. Retrieved from https://executivefinance.nl/wp-content/uploads/2023/01/VCNL-Whitepaper-Exfiltratie.pdf. 2022.

[71] T. Meurs, M. Junger, A. Abhishta, E. Tews and E. Ratia. 'COORDINATE: A model to analyse the benefits and costs of coordinating cybercrime'. *JISIS* 12.4, 2022.

[72] T. Meurs, M. Junger, E. Tews and A. Abhishta. 'Ransomware: How Attacker's Effort, Victim Characteristics and Context Influence Ransom Requested, Payment and Financial Loss'. *2022 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2022, pp. 1–13.

[73] T. Meurs, E. Cartwright, A. Cartwright, M. Junger, R. Hoheisel, E. Tews and A. Abhishta. 'Ransomware Economics: A Two-Step Approach To Model Ransom Paid'. *18th Symposium on Electronic Crime Research, eCrime 2023*. 2023.

[74] NCSC. *Amendment: NIS2 Directive Protects Network Information Systems*. `https://business.gov.nl/amendment/nis2-directive-protects-network-information-systems/`. Accessed: 2024-03-27. 2024.

[75] No More Ransom. `https://web.archive.org/web/20240000000000*/www.nomoreransom.org`. Accessed: 2024-03-25. n.d.

[76] K. Oosthoek, J. Cable and G. Smaragdakis. 'A Tale of Two Markets: Investigating the Ransomware Payments Economy'. *arXiv preprint arXiv:2205.05028*, 2022.

[77] O. Owolafe and A. Thompson. 'Analysis of Crypto-Ransomware Using Network Traffic'. *Journal of Information Security and Cybercrimes Research* 5.1, 2022, pp. 72–79.

[78] H. Oz, A. Aris, A. Levi and A. S. Uluagac. 'A survey on ransomware: Evolution, taxonomy, and defense solutions'. *ACM Computing Surveys (CSUR)*, 2021.

[79] K. Padayachee. 'A framework of opportunity-reducing techniques to mitigate the insider threat'. *2015 Information Security for South Africa (ISSA)*. IEEE, 2015, pp. 1–8.

[80] M. Pichon. *Global CERT Orange CyberDefense - World Watch team's ransomware ecosystem map*. `https://github.com/cert-orangecyberdefense/ransomware_map`. Version 26, March 2024. 2024.

[81] E. L. Piza. 'The effect of various police enforcement actions on violent crime: Evidence from a saturation foot-patrol intervention'. *Criminal Justice Policy Review* 29.6-7, 2018, pp. 611–629.

[82] C. P. Research. *Behind the curtains of the ransomware economy - the victims and the Cybercriminals*. Retrieved July 12, 2022, from https://research.checkpoint.com/2022/behind-the-curtains-of-the-ransomware-economy-the-victims-and-the-cybercriminals/. 2022.

[83] C. P. Research. *Leaks of Conti Ransomware Group Paint Picture of a Surprisingly Normal Tech Start-Up*. Retrieved August 31, 2022, from https://research.checkpoint.com/2022/leaks-of-conti-ransomware-group-paint-picture-of-a-surprisingly-normal-tech-start-up-sort-of/. 2022.

[84] L. W. Sherman. 'The rise of evidence-based policing: Targeting, testing, and tracking'. *Crime and justice* 42.1, 2013, pp. 377–451.

[85] C. Simoiu, J. Bonneau, C. Gates and S. Goel. '"I was told to buy a software or lose my computer. I ignored it": A study of ransomware'. *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. 2019, pp. 155–174.

[86] U.S. Department of the Treasury. *Sanctions List Search*. `https://sanctionssearch.ofac.treas.gov/`. Accessed: 2024-03-25. n.d.

[87] D. S. Wall. 'The transnational cybercrime extortion landscape and the pandemic: Changes in ransomware offender tactics, attack scalability and the organisation of offending'. *Special Issue 5 Eur. L. Enf't Rsch. Bull.* 45, 2022.

[88] K. Wang, J. Pang, D. Chen, Y. Zhao, D. Huang, C. Chen and W. Han. 'A large-scale empirical analysis of ransomware activities in bitcoin'. *ACM Transactions on the Web (TWEB)* 16.2, 2021, pp. 1–29.

[89] R. V. Wegberg and T. Verburgh. 'Lost in the Dream? Measuring the effects of Operation Bayonet on vendors migrating to Dream Market'. *WebSci'18: Evolution of the Darknet*. Association for Computing Machinery (ACM), 2018, pp. 1–5.

[90] C. Winship and S. L. Morgan. 'The estimation of causal effects from observational data'. *Annual review of sociology* 25.1, 1999, pp. 659–706.

[91] D. W. Woods and S. Seymour. 'Evidence-based cybersecurity policy? A meta-review of security control effectiveness'. *Journal of Cyber Policy*, 2024, pp. 1–19.

[92] World Bank. *World Development Indicators*. `https://databank.worldbank.org/reports.aspx?source=2`. 2024.

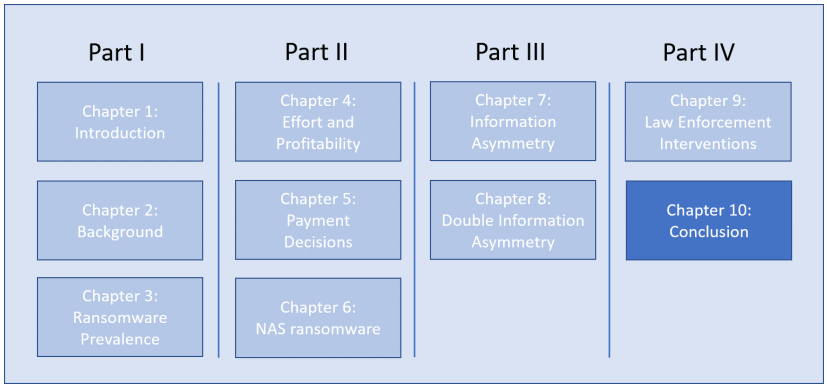
This page is intentionally left blank.

There is no real ending. It's just the place where you stop the story

~ Frank Herbert

Chapter 10

Conclusions



In this chapter, we summarize the main findings of this thesis. We begin by examining ransomware prevalence, followed by a discussion of the implications for Rational Choice Theory. Subsequently, we revisit each sub-question and provide answers informed by the preceding chapters. Afterwards, we address the main research question outlined in Chapter 1. Finally, we offer our perspective on future research directions.

10.1 Prevalence

Assessing the prevalence of ransomware is important for ensuring government accountability, informing public awareness, guiding resource allocation, and supporting research efforts [14]. This dissertation examines the prevalence of ransomware attacks on both individuals and companies, using police reports, incident response records, and leakpage data.

For individuals, our focus was on ransomware targeting NAS devices (Chapter 6). Between 2019 and 2022, 502 successful ransomware attacks were reported to the Dutch police, excluding 23 unsuccessful attempts. Of these, 104 (20.7%) targeted NAS devices, which are more automated and less complex. Shodan data showed 604 DeadBolt infections in the Netherlands by August 2022, while police estimated 1,100 infected devices by October 2022 [11, 5]. However, only 31 DeadBolt attacks were reported to the police, reflecting a low reporting rate of 5.1% based on Shodan data and 2.8% based on police estimates. These figures are consistent with previous research indicating that only 5-10% of cybercrime incidents are reported [16].

For companies, ransomware prevalence varies across datasets and company sizes (Chapters 3, 4, 5, and 9). From 2019 to 2022, we estimate that 138 large companies (95% CI: 55.1 to 122.6), 219 medium companies (95% CI: 98.3 to 190.1), and 2,706 small companies (95% CI: 1272.6 to 7057.2) in the Netherlands experienced ransomware attacks. While the estimates for large and medium companies are reliable, the figures for small businesses contain more uncertainty due to wider confidence intervals.

Underreporting remains an important issue across all company sizes. Approximately 41.4% of attacks on large companies were reported (58.6% unreported). For medium companies, 40.2% of incidents were captured (59.8% unreported). However, ransomware reporting rates (around 40%) appear higher than other online crime, such as online fraud, which typically sees reporting rates of 11.5-14% across different countries [13, 10, 7, 8].

Our models estimate the annual ransomware risk at 1.3% for large companies and 0.6% for medium companies, which aligns with Cybersecurity Monitor (CBS) figures—4.0% for large companies in 2021 (dropping to 2.3% in 2022) and 2.3% for medium companies in 2021 (falling to 1.4% in 2022) [4]. Our confidence intervals (CI) for large companies (2.1% to 4.7%) and medium companies (1.0% to 1.9%) closely match these CBS estimates, reinforcing the robustness of our findings.

Several limitations affected the accuracy of our ransomware prevalence estimates. Our data focuses on direct victims and excludes indirect impacts on busi-

nesses connected through supply chains. Unsuccessful ransomware attempts, which may be underreported, were also not considered. Additionally, company size data was only available for 2021, and extrapolating for other years introduced some uncertainty [4]. Despite these limitations, the consistency between our estimates and CBS's data suggests that our findings are reliable, as two different methods of measuring the same phenomenon produced similar results, indicating a robust measurement.

This dissertation highlights underreporting of ransomware, particularly among individuals and small companies. For businesses, medium and large companies experience a lower risk of ransomware attacks compared to smaller companies; however, the reported financial losses are higher (Chapter 4). Larger businesses also demonstrate a greater willingness to report incidents, engaging more frequently with law enforcement and/or incident response companies. These findings illustrate the impact of (double-extortion) ransomware in the Netherlands between 2019 and 2022.

10.2 Implications for Rational Choice Theory

Throughout this dissertation, we applied Rational Choice Theory (RCT) as a central framework to explain offender decision-making in empirical studies of double-extortion ransomware attacks. The findings from our research indicate that the existence of different types of ransomware attacks and their modus operandi can be well explained by the trade-offs offenders make between profitability, effort, and risk. Additionally, we explored how other theories related to RCT — Routine Activity Theory (RAT), signaling games, and Situational Crime Prevention (SCP) — complement RCT in explaining empirical observations in our studies: RAT explains how ransomware offenders target specific victims, signaling games address information asymmetry in ransomware negotiations, and SCP highlights the effect of law enforcement interventions. In this section we will give a description how the different chapters are related to the RCT elements profitability, effort, and risk.

Firstly, we analyzed crime chains. Based on a systematic literature review, we developed a theoretical framework (Chapter 2): the Cybercrime Coordination Model (COORDINATE). The framework shows that longer, coordinated chains—such as those in double-extortion ransomware—enhance profitability. Furthermore, COORDINATE implies that double-extortion ransomware is facilitated by the fifth crime chain mechanism, "setup," where an encryption-only ransomware attack creates the opportunity to also steal sensitive data [6].

An important insight from the COORDINATE framework is that specialized roles enhance the skills of individual offenders, enabling them to target a wider range of victims. As each offender focuses on a specific aspect of the ransomware attack, such as negotiating ransoms or breaching networks, they become more proficient, allowing them to breach even highly secured systems. This specialization makes organizations with robust IT protection measures more vulnerable, as offenders are now more skilled in circumventing advanced security defenses. This implies that profitability potentially increases, since companies with robust IT protection might manage more sensitive data.

Additionally, according to the COORDINATE framework, coordinated attacks benefit offenders by reducing their individual exposure to risk. Since tasks are divided across multiple actors, each offender leaves a smaller digital footprint, making it more difficult for law enforcement to trace and arrest those involved. Offenders engaged in less visible roles, such as initial access or data exfiltration, may evade detection, while those directly responsible for encryption are often the primary focus of law enforcement efforts.

Evidence of coordination and specialization can be seen in multiple chapters of this dissertation. For instance, we found that specialization in roles like data exfiltration and ransom negotiation increases the ransom requested (Chapters 4 and 5). Furthermore, we found that offenders, through automated attacks on vulnerable NAS devices, further reduce effort while maintaining profitability (Chapter 6). Lastly, Chapter 9 highlights how dividing roles allows offenders to avoid detection, with law enforcement primarily focusing on the more visible actors involved in encryption, leaving other roles with a smaller footprint and lower risk of arrest. This coordinated approach in double-extortion ransomware demonstrates that offenders make deliberate, rational choices to exploit vulnerabilities, while minimizing their own exposure.

Secondly, we examined the profitability of double-extortion ransomware. Based on the analysis of 353 ransomware attacks reported to the Dutch Police, our findings show that double-extortion ransomware is more profitable than encryption-only ransomware (Chapter 4). Factors positively influencing ransom requested included data exfiltration, offender's affiliation with a RaaS group, and blackmail. Offenders also seem to assess the financial profile of their victims, adjusting ransom demands based on factors such as yearly revenue and industry (Chapters 4 & 5). These adjustments indicate a clear trade-off between increased effort and profitability. For example, data exfiltration might require additional effort, such as identifying, collecting, and exfiltrating valuable files. Additionally, offenders rationally target victims with a higher capacity to pay larger sums. However, these victims might have better IT protection measures in

place. Both examples illustrate the trade-off between effort and profit, aligning with RCT.

The trade-off between increased effort and profitability is further supported by the analysis of a combined dataset of 481 ransomware attacks from police and an Incident Response (IR) company, where ransom amounts increased by 5.4 times when data exfiltration occurred (Chapter 5). Additionally, ransom amount increased 2.8 times if victims had insurance coverage, despite not being more likely to pay in general. In line with RCT and RAT, offenders demonstrate a rational decision-making process by adjusting ransom amounts according to the victim's capacity to pay. Furthermore, victims with recoverable backups are less likely to pay ransom, highlighting the importance of victim preparedness in influencing the success of ransomware offenders. However, when proper backups are lacking, victims are more inclined to pay. Many offenders, therefore, invest effort in locating and removing backups, increasing the likelihood of payment.

Furthermore, we found that information asymmetry of data exfiltration might increase offender's profitability (Chapters 7 & 8). In many ransomware cases, victims struggle to determine whether their data has been exfiltrated (Chapters 4 % 5). Some victims may possess logs that help identify accessed files, while others are left uncertain about the extent of the breach. This uncertainty provides offenders with the opportunity to claim data exfiltration even when it has not occurred, inflating ransom requested. Offenders may offer 'evidence' of exfiltration, or victims may request it, creating a strategic interaction.

To explore this interaction, we applied a game-theoretic framework, focusing on signaling in data exfiltration (Chapter 7). Our analysis revealed five distinct equilibria, representing different strategies offenders use to signal data exfiltration, from always signaling, to never signaling, or signaling only when theft occurred. The interaction between the offender's signal and the victim's response forms a strategic game where both navigate the uncertainty of exfiltration. Offenders often provide ambiguous or unverifiable signals, pushing victims to pay more than they otherwise would.

We expanded on the signaling game by introducing a double-sided information asymmetry model (Chapter 8), in which offenders are unaware of the true value of the exfiltrated data, while victims possess this knowledge. Private information, such as the actual value of the stolen data, affects the profitability of the attack. In our model the losses to the offender ranged from 0% to over 20% when the true value of stolen data is unknown. These findings emphasize the importance of victims withholding information about the value of exfiltrated data to prevent offenders from extracting more surplus through bluffing.

The results of these studies align closely with RCT. The strategic manipulation of proof of data exfiltration through bluffing and signaling highlights how offenders exploit information asymmetry to increase profits from the threat of data exfiltration without the effort of conducting the exfiltration.

Thirdly, we explored the differences between ransomware targeting NAS devices versus other types of ransomware. We focused on differences between their modus operandi, victim profiles, and trends (Chapter 6). We analysed 502 ransomware attacks reported to the Dutch Police, of which 104 (20.7%) targeted NAS devices. Our findings indicate that NAS ransomware relies on automation and scalability, demanding smaller ransoms through standardized ransom notes, whereas regular ransomware involves more complex, tailored attacks with higher demands. NAS ransomware attacks are simpler, typically involving only reconnaissance and encryption, while regular ransomware includes multiple steps like lateral movement, data exfiltration, and negotiations.

Furthermore, NAS ransomware seems to target individuals, with 66% of victims being citizens, while regular ransomware primarily targets businesses (91%). The average ransom for NAS attacks is €1,404, compared to €727,544 for regular ransomware. NAS victims often experience non-financial losses, such as the loss of personal data, whereas businesses hit by regular ransomware report financial damage. Moreover, 70% of NAS victims had no backups, compared to 38% for regular ransomware victims, indicating offenders' focus on less prepared targets.

Additionally, NAS ransomware is strongly linked to the discovery of vulnerabilities in NAS devices, with major attack spikes following newly identified weaknesses. This contrasts with the steady rise in regular ransomware attacks, which are not tied to specific (NAS) vulnerabilities.

In conclusion, NAS ransomware differs from regular ransomware, which are more automated and less complex than regular ransomware, leading to lower ransom demands since victims are primarily individuals. Offenders might conduct NAS ransomware attacks due to its low effort and risk, while still earning profits, aligning with RCT. The vulnerabilities in NAS devices also highlight the need for preventive interventions, as emphasized by SCP.

Finally, we evaluated the impact of law enforcement interventions. We analysed three datasets (Chapter 9): a list of 12,250 ransomware victims published on leak pages, a list of 35 law enforcement interventions and a list of 19 instances of ransomware groups rebranding. The interventions list consisted of interventions such as arrests, leak page server takedowns, crypto-asset freezes, sanctions, and the use of decryptors (Chapter 9). Interventions like arrests and sanctions increased the risk for ransomware offenders, interventions such as

server takedowns increased the effort required for ransomware groups to continue their operations, and freezing crypto-assets after ransom payments and releasing decryptors directly impacted the profitability of ransomware attacks. Decryptors enabled victims to recover their data without paying ransoms. Almost half of the ransomware groups ceased operations after interventions, with minimal evidence of crime displacement.

The effect of law enforcement interventions is further supported with the DeadBolt operation, where the Dutch Police retrieved decryption keys without payment for victims of DeadBolt ransomware (Chapter 6). After the intervention there was a a decrease in DeadBolt ransomware attacks.

Overall, the interventions examined in our dissertation disrupted ransomware operations by altering the balance of risk, effort, and profitability. Consistent with RCT, and therefore SCP, these actions increased risks, raised operational costs, and reduced potential rewards, effectively deterring many ransomware groups from continuing their activities.

10.3 Revisiting the sub-questions

In this section, we revisit the individual sub-questions addressed throughout the dissertation. These questions explored different aspects of offender decision-making in double-extortion ransomware attacks, each contributing to the understanding of profitability, effort, and risk as outlined by Rational Choice Theory.

Research Question 1 (RQ1): In what ways do crime chains affect the profitability, effort, and risks of attacks?

Crime chains enhance the efficiency of ransomware attacks by dividing tasks among different offenders, allowing them to increase profitability while managing effort and risk. Double-extortion ransomware, in particular, benefits from crime chains by combining data encryption and exfiltration, which boosts ransom demands with relatively low additional effort (Chapter 2).

Research Question 2 (RQ2): How does combining data encryption with data exfiltration alter the profitability of ransomware attacks?

Combining data encryption with exfiltration increases the profitability of ransomware attacks. Double-extortion ransomware results in ransom demands that are several times higher than those of encryption-only attacks. The added complexity of exfiltration is justified by the substantial increase in potential financial gains (Chapters 4 & 5).

Research Question 3 (RQ3): How does the profitability of ransomware attacks targeting NAS devices compare to the profitability of regular ransomware attacks?

Ransomware attacks targeting NAS devices typically demand lower ransoms due to the financial limitations of individual victims, but their high level of automation makes them profitable at scale. These attacks involve less effort and risk, and while the ransoms are smaller, the volume of attacks compensates for this (Chapter 6).

Research Question 4 (RQ4): How does the uncertainty of data exfiltration affect the dynamics between offenders and victims in double-extortion ransomware attacks?

Uncertainty about whether data has been exfiltrated allows offenders to bluff, inflating ransom demands and manipulating victims. This information asymmetry increases the likelihood of higher payouts, as victims are often unable to verify whether their sensitive data has been stolen (Chapters 7). However, the offender often does not know the value of the exfiltrated data, which decreases the profitability for offenders (Chapter 8).

Research Question 5 (RQ5): What intervention strategies could law enforcement employ to combat double-extortion ransomware attacks?

Law enforcement interventions—such as arrests, leakpage server takedowns, and cryptocurrency freezes—are effective in disrupting ransomware operations. These actions raise the perceived risk for offenders, increase the effort of conducting attacks and decrease profitability of ransomware attacks, leading to reduced ransomware activity (Chapters 6 & 9).

10.4 Main conclusions

Building on the answers to the sub-questions in the previous section, we now address the main research question:

> **Main Research Question: How do double-extortion ransomware attacks influence profitability, effort, and risks for offenders?**

Our research focused on understanding why offenders choose to conduct double-extortion ransomware attacks by analyzing how profitability, effort, and risks compare to encryption-only attacks. Below, we summarize our main findings to answer the main research question:

Profitability: Double-extortion ransomware increases profitability. By threatening both data encryption and public exposure, offenders exploit both the confidentiality and availability aspects of data, pressuring victims into paying larger ransoms. In contrast, encryption-only attacks target only data availability, which limits the leverage on victims. Additionally, the mere

claim of data exfiltration, whether real or fabricated, adds further pressure, increasing the ransom amounts paid. Moreover, we observed that triple- and quadruple-extortion ransomware, where offenders use additional tactics like DDoS-attacks or contacting employees/customers, are rare. While these methods may influence victim payment decisions, the extra effort does not seem to justify the profits, explaining their low prevalence. In contrast, double-extortion ransomware strikes a favorable balance between effort and profit, leading to its higher prevalence.

Effort: Double-extortion attacks demand greater effort compared to encryption-only attacks. Offenders must not only encrypt the data but also set up infrastructure for data exfiltration, locate, collect, and exfiltrate sensitive information, all while remaining undetected. During negotiations, they must credibly signal the exfiltration to increase pressure on the victim. Our findings suggest that this extra effort, while substantial, is more than compensated by the increased profits from larger ransom payments. For example, NAS ransomware, which involves fewer steps and is more automated, typically results in lower ransoms, underscoring the connection between complexity, effort, and financial rewards. Moreover, the need for credible signaling during negotiations adds another layer of complexity for the offenders, who must convince victims of the actual or potential data exfiltration.

Risks: Our research suggests that double-extortion ransomware does not increase the risk of offenders getting caught compared to encryption-only attacks. While these attacks may draw more attention from law enforcement due to their broader societal impact, we found no evidence that publishing stolen data increases the likelihood of law enforcement intervention. The primary risk for offenders lies in the possibility of early detection by the victim during the data exfiltration phase. If victims detect the attack early, they may implement defensive measures to prevent encryption, reducing the offender's leverage. However, our study shows that early detection is uncommon, and the majority of double-extortion attacks proceed undetected until the ransom demand. Therefore, while double-extortion increases the complexity and effort for offenders, it does not proportionally raise their risk of being caught.

In conclusion, our findings strongly support Rational Choice Theory (RCT), showing a clear link between profits, efforts, and risks. Double-extortion ransomware increases both effort and profitability without affecting risk compared to

encryption-only attacks. Not all ransomware groups engage in this tactic, suggesting different valuations of effort, risk, and profit. However, given its rising prevalence, the profit of double-extortion appears to outweigh the additional effort for most groups.

10.5 Future research

This section outlines potential directions for future research, focusing on novel approaches to refine the application of Rational Choice Theory (RCT) to cybercrime. First, RCT could be enhanced by incorporating concepts from behavioral economics to better understand how offenders value RCT elements. Second, we propose methods to improve the measurement of RCT elements. These approaches aim to deepen our understanding of both offender and victim decision-making, as well as their interactions within the broader cybercrime ecosystem.

10.5.1 Integrating Behavioral Economics and RCT

Economic Models of Offender Behavior. Future research should investigate whether concepts from behavioral economics—such as risk aversion, bounded rationality, and prospect theory—could better explain empirical observations of offender decision-making than the models used in this dissertation. While RCT assumes rational actors who carefully weigh risks, effort, and profits, behavioral economics suggests that offenders may exhibit cognitive biases, leading them to make decisions that deviate from strict rationality. For instance, risk-averse offenders might avoid high-risk, high-reward attacks, while others may misjudge potential profits or underestimate risks due to bounded rationality. Integrating these concepts with RCT could provide a more nuanced view of how offenders perceive and respond to risks and rewards in the context of ransomware, leading to more targeted intervention strategies.

Victim Decision-Making and Strategic Interaction with Offenders. Victim decision-making before and during ransomware attacks appears to involve a strategic interaction with the offender's RCT. Future research could explore victim's decision-making before the attack: is there an optimal level of cybersecurity investment that influences an offender's perception of profitability, effort, and risk? For example, if victims invest too little, attacks become more profitable for offenders; conversely, if investments are too high, the effort for offenders increases, making the attack less attractive. This implies the existence of an "optimum" level of investment

by victims that could strategically shift the offenders' decision-making under RCT.

Empirical research could focus on collecting data from companies with varying levels of cybersecurity investments to study how these investments influence ransomware outcomes. [17] may point in the right direction by examining how businesses make trade-offs between security costs and potential losses. Additionally, data from cyber insurance providers, breach reports, and cybersecurity audits could be valuable in understanding how investment levels impact both the frequency and success of ransomware attacks. This type of research could help identify the optimal balance of companies investments in cybersecurity that alter offenders' RCT-based trade-offs, thereby reducing the likelihood and profitability of attacks.

Refining Risk Perception in Cyber Offenders. Risk perception in cybercrime may differ from offline crime, and future research should examine how this affects offender decision-making. For instance, if online crimes are perceived to have a lower probability of arrest, do online interventions have a larger impact due to the lower baseline risk? This aligns with behavioral economics, where risk valuation depends on context. Understanding how offenders perceive risks from interventions like arrests, server takedowns, or cryptocurrency freezes can help develop more effective strategies. [3] analyzed Reddit communications after law enforcement interventions, offering a potential starting point for studying how interventions influence perceived risks and decision-making, and whether increased perceived risk alters the effort-reward balance for offenders, ultimately reducing crime.

Effort Optimization and Mistrust in Ransomware Attacks. Effort optimization in ransomware attacks might be constrained by the nature of online offender collaborations, where trust between offenders is low, and agreements lack formal contracts [12]. This mistrust, combined with the risk of being scammed by collaborators, may limit the ability of offenders to efficiently outsource tasks or reduce effort [9]. If offenders are unpredictable and prone to self-interest, as behavioral economics suggests, optimizing the profits-effort trade-off may be difficult in practice. Future research could test this hypothesis by analyzing communications within offender networks, such as on dark web marketplaces and forums, to identify instances of collaboration breakdowns or scams. By studying how mistrust shapes the structure of online crime chains, researchers can explore whether increasing this mistrust—through tactics like law enforcement in-

filtration or disinformation campaigns—might lead to reduced efficiency in offender networks and, consequently, a reduction in ransomware attacks.

10.5.2 Measurement of RCT elements

Improve Accuracy of Measuring Effort, Risks, and Profits. More precise measurement of RCT elements—effort, risks, and profits—can enhance our understanding of cybercrime [2, 18]. Future research can improve these metrics in two ways. First, combining different data sources, as done in this dissertation, can increase the accuracy of measuring RCT elements. For example, integrating data from victims, leak pages, and darknet forums could provide a clearer picture of ransomware profits. Second, new technological advancements, such as large language models (LLMs), can help extract insights from unstructured data, such as dark web communications, to better quantify RCT elements.

Tracking Effort, Risks, and Profits Over Time. Longitudinal studies are essential to track how the balance between effort, risks, and profits changes over time [1]. Offenders continuously adapt their strategies in response to shifting victim defenses, technological advancements, and law enforcement actions. Future research should investigate how interventions like arrests or leak page server takedowns affect these elements over time and whether offenders shift to lower-risk, lower-reward attacks. Additionally, it would be valuable to explore whether changes in RCT elements occur independently of external circumstances or are driven by specific contextual factors, such as the war between Russia and Ukraine. Emerging technologies, like LLMs, are likely to reduce the effort required for some cybercrimes, potentially increasing their prevalence [15]. Comparing the adoption of new technologies with cybercrime trends could provide valuable insights into the factors influencing cybercrime rates.

This page is intentionally left blank.

Bibliography

[1] L. Allodi, F. Massacci and J. Williams. 'The work-averse cyberattacker model: theory and evidence from two million attack signatures'. *Risk Analysis* 42.8, 2022, pp. 1623–1642.

[2] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. V. Eeten, M. Levi, T. Moore and S. Savage. 'Measuring the cost of cybercrime'. *The economics of information security and privacy*. 2013, pp. 265–300.

[3] C. Bradley. 'On the resilience of the Dark Net Market ecosystem to law enforcement intervention'. PhD thesis. UCL (University College London), 2019.

[4] Centraal Bureau voor de Statistiek. *Online veiligheid en criminaliteit 2022*. https://www.cbs.nl/nl-nl/publicatie/2023/19/online-veiligheid-en-criminaliteit-2022. Accessed: 2024-06-19. 2023.

[5] Chainalysis Team. *Deadbolt Ransomware Strain Tricked into Giving Up Decryption Keys*. Accessed: 2024-08-08. 2023. URL: https://www.chainalysis.com/blog/deadbolt-ransomware-strain-tricked-into-giving-up-decryption-keys/.

[6] M. Felson and M. A. Eckert. *Crime and everyday life: A brief introduction*. Sage Publications, 2018.

[7] C. Fonseca, S. Moreira and I. Guedes. 'Online consumer fraud victimization and reporting: A quantitative study of the predictors and motives'. *Victims & Offenders* 17.5, 2022, pp. 756–780.

[8] L. Koning, M. Junger and B. P. Veldkamp. 'Reporting fraud victimization to the police: factors that affect why victims do not report'. *Psychology, Crime and Law*, 2023. in press.

[9] T. Meurs, R. Hoheisel, M. Junger, A. Abhishta and D. McCoy. 'What To Do Against Ransomware? Evaluating Law Enforcement Interventions'. *2024 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2024, pp. 1–13.

[10] R. E. Morgan. *Financial Fraud in the United States, 2017*. Tech. rep. NCJ 255817. Washington DC: Bureau of Justice Statistics, 2021. URL: https://bjs.ojp.gov/redirect-legacy/content/pub/pdf/ffus17.pdf.

[11] R. Nieuws. *Unieke actie: politie bevrijdt gegijzelde computers dankzij truc met bitcoin*. Retrieved November 11, 2022, from https://www.rtlnieuws.nl/nieuws/nederland/artikel/5337913/unieke-actie-politie-probeert-gegijzelde-computers-met-slimme-truc. 2022.

[12] L. Norbutas, S. Ruiter and R. Corten. 'Reputation transferability across contexts: Maintaining cooperation among anonymous cryptomarket actors when moving between markets'. *International Journal of Drug Policy* 76, 2020, p. 102635.

[13] ONS. *Nature of Fraud and Computer Misuse in England and Wales: Year Ending March 2022*. Tech. rep. 2022. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/natureoffraudandcomputermisuseinenglandandwales/yearendingmarch2022.

[14] A. Smith. *Crime Statistics: An Independent Review*. Tech. rep. London, UK: Home Office, 2006. URL: http://rds.homeoffice.gov.uk/rds/pdfs06/crime-statistics-independent-review-06.pdf.

[15] P. Treleaven, J. Barnett, D. Brown, A. Bud, E. Fenoglio, C. Kerrigan, A. Koshiyama, S. Sfeir-Tait and M. Schoernig. *The future of cybercrime: AI and emerging technologies are creating a cybercrime tsunami*. SSRN. 2023.

[16] S. van de Weijer, R. Leukfeldt and S. van der Zee. 'Reporting cybercrime victimization: determinants, motives, and previous experiences'. *Policing: An International Journal*, 2020.

[17] D. W. Woods and S. Seymour. 'Evidence-based cybersecurity policy? A meta-review of security control effectiveness'. *Journal of Cyber Policy*, 2024, pp. 1–19.

[18] D. W. Woods and L. Walter. 'Reviewing estimates of cybercrime victimisation and cyber risk likelihood'. *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2022, pp. 150–162.

Police Collaboration

This PhD project represents a special collaboration between the Dutch police and the University of Twente. The project is based on placing two PhD candidates in a team of detectives, which is to my knowledge highly unusual. Over the past four years, this partnership has proven to be a unique and fruitful experience. In my opinion, the project has provided valuable insights about ransomware and created an environment of mutual benefit for both academic research and law enforcement operations:

Understanding Real-World Cybercrime: Working within the police environment provided valuable insights into law enforcement operations. Observing the challenges of tracking ransomware groups, the effort required to find mistakes, and the deanonymization of internet traffic helped me better understand the balance between profits, effort, and risks for ransomware offenders. Although not all insights could be published, this experience enriched the academic value of the research by offering a nuanced perspective on cybercrime.

Quick Integration of Research and Police Practice: Being part of the police team allowed for the rapid integration of research findings into practice. Regular 'lunch-presentations' enabled me to share my work with detectives and receive real-time feedback based on their field experience. This collaboration shaped the research direction, such as focusing on ransomware negotiation dynamics.

Deeper Exploration of Police Challenges: Beyond knowledge exchange, the collaboration enabled a more in-depth exploration of cybercrime issues than is typically feasible in police operations. This led to new insights,

such as strategies for identifying Dutch criminals on darknet forums. The project provided direct benefits to the police, including improved disruption strategies and a better understanding of criminal behavior.

Direct Access to Academic Expertise: The collaboration with the university gave the police immediate access to academic knowledge. This partnership facilitated cooperation with field specialists, making it easier to study specific communication protocols with expert input. This integration of academic insights into police work ensured practical application without delays often caused by formal communication channels.

Alignment of Ethical Considerations: The project underscored the importance of ethics, especially concerning individual privacy. Both the university and the police adhered to strict ethical guidelines, ensuring responsible research. Aligning these procedures strengthened the ethical foundation of the project, balancing academic rigor with respect for privacy.

In my opinion, the collaboration between the Dutch police and the University of Twente has been highly successful, offering numerous benefits for both parties. It has bridged the gap between theoretical research and practical application, provided unique insights into offenders' decision-making, and facilitated ethical and effective research practices. The findings from this research have the potential to enhance police operations and methodologies. Continued funding for PhD positions within detective teams could yield further valuable results, benefiting both academic research and practical law enforcement efforts. By supporting such projects, the police can ensure ongoing advancements in their operations while promoting useful academic research.

Ethics

This dissertation adheres to the ethical guidelines of the Menlo Report, which provides a framework for ICT research ethics. Special considerations were necessary due to our collaboration with law enforcement and the handling of sensitive data from victims and offenders.

Respect for Persons: Participation as a research subject is voluntary and based on informed consent. However, due to the nature of the police organization and the context of cybercrime research, obtaining informed consent from all participants, particularly victims and offenders, was not feasible. We took extra care to ensure that no personally identifiable information (PII) was disclosed. The decision to use aggregated results was made in collaboration with the police to maximize societal benefits by improving strategies to combat cybercrime effectively.

Beneficence: The principle of "do no harm" was a guiding force in this research. We systematically assessed the risks and benefits of the study, aiming to maximize probable benefits while minimizing potential harms. We were aware of the ethical concern that studying criminal decision-making might inadvertently assist criminals in making better decisions. Our primary goal was to derive actionable insights to better defend victims rather than educate criminals. Despite this risk, we believe that the overall benefit to society outweighs the potential harm, as the research aims to enhance cybercrime prevention and mitigation strategies.

Justice: We ensured that each participant was treated with equal consideration. The selection of subjects was fair, and the benefits of the research were distributed equitably. We did not give special attention to high-profile

victims; each incident was evaluated on its own merits. However, we acknowledge the possibility of selection bias, as large companies might be more inclined to report incidents to the police. This potential bias is taken into account when generalizing from our specific sample and formulating defensive strategies.

Respect for Law and Public Interest: Legal due diligence and transparency in methods and results were paramount. We engaged with the ethical committee of the university and the privacy officers from the police to ensure that our research complied with legal and ethical standards. No PII of victims was included in the research. Additionally, we were careful not to disclose sensitive information about the police's modus operandi, as this could inadvertently aid criminals. Final drafts of papers were screened by police officers to ensure that no critical operational details were revealed.

In conclusion, in our research we conducted a best effort to balance ethical standards with the pursuit of valuable research insights, with special focus on the protection of individuals' privacy and societal interests. We hope the above considerations might be useful for further researchers conducting empirical research in collaboration with Law Enforcement.

About the Author

Tom Meurs was born in Utrecht on January 23rd, 1992. He grew up in Ermelo, where he completed his secondary education at Christelijk College Nassau Veluwe in Harderwijk. After graduating in 2011, he took a gap year to fully focus on his main passion: chess. Subsequently, he pursued a bachelor's degree in Psychology, graduating in 2016 with specializations in methodology and psychometrics, as well as clinical psychology. Additionally, he completed a major in Econometrics and Operations Research.

Following his studies, Tom began his career at TNO (Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek; English: Dutch Organization for Applied Scientific Research), an independent research organization in the Netherlands that focuses on applied science. At TNO, he worked in the Military Operations department, conducting projects for the Dutch Armed Forces, Dutch Police, and RIEC.

After TNO, Tom started his PhD at the Dutch Police and the University of Twente. His research primarily focuses on ransomware, using his background both in statistics, methodology, and social sciences to understand crime from two perspectives. Tom's research has also supported the Dutch Police in the development of the Ransomware TaskForce, which was established to enhance the information position on ransomware and to effectively intervene in high-priority targets.

In addition to his academic pursuits, Tom enjoys chess, Crossfit, running, and theatersport. He is married to Sylvia Meurs-Drijver and resides in Amersfoort. They are expecting a baby boy.

List of Publications

The list of peer-reviewed publications (in chronological order) co-authored by Tom Meurs during his doctoral research are as follows:

- Meurs, T., Junger, M., Cruyff, M., & van der Heijden, P. G. M. (2024). Estimating the Number of Unobserved Ransomware Attacks. *Available at SSRN 4942706*.
- Meurs, T., Hoheisel, R., Junger, M., Abhishta, A., & McCoy, D. (2024). What to do against ransomware? Evaluating law enforcement interventions. In *2024 APWG Symposium on Electronic Crime Research (eCrime)* (pp. 1–13). IEEE.
- Hoheisel, R., Meurs, T., Junger, M., Tews, E., & Abhishta, A. (2024). Dark web dialogues: Analyzing communication platform choices of underground forum users. In *2024 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE.
- Meurs, T., Hoheisel, R., Junger, M., Abhishta, A., & McCoy, D. (2024). What to do against ransomware? Evaluating law enforcement interventions. In *2024 APWG Symposium on Electronic Crime Research (eCrime)* (pp. 1–13). IEEE.
- Meurs, T., Cartwright, E., Cartwright, A., Junger, M., & Abhishta, A. (2024). Deception in double extortion ransomware attacks: An analysis of profitability and credibility. *Computers & Security*, 138, 103670.
- Meurs, T., Cartwright, E., Cartwright, A., Junger, M., Hoheisel, R., Tews, E., & Abhishta, A. (2023). Ransomware economics: a two-step approach

to model ransom paid. In *18th Symposium on Electronic Crime Research, eCrime 2023*.

- Meurs, T., Cartwright, E., & Cartwright, A. (2023). Double-sided information asymmetry in double extortion ransomware. In *International Conference on Decision and Game Theory for Security* (pp. 311-328). Cham: Springer Nature Switzerland.
- Meurs, T., Junger, M., Tews, E., & Abhishta, A. (2022). NAS-ransomware: hoe ransomware-aanvallen tegen NAS-apparaten verschillen van reguliere ransomware-aanvallen. *Tijdschrift voor veiligheid*, 21(3-4), 69-88.
- Meurs, T., Junger, M., Tews, E., & Abhishta, A. (2022). Ransomware: How attacker's effort, victim characteristics and context influence ransom requested, payment and financial loss. In *2022 APWG symposium on electronic crime research (eCrime)* (pp. 1-13). IEEE.
- Meurs, T., Junger, M., Abhishta, A., Tews, E., & Ratia, E. (2022). COORDINATE: A model to analyse the benefits and costs of coordinating cybercrime. *Journal of internet services and information security*, 12(4).
- Meurs, T., Junger, M., Abhishta, A., & Tews, E. (2022). POSTER: How Attackers Determine the Ransom in Ransomware Attacks. In *7th IEEE European Symposium on Security and Privacy*.

Bachelors: *I'm going to cure cancer*
Masters: *I'm going helping discover new biochemical pathways which are crucial to our understanding of cancer!*
First year PhD: *This gene plays an important role in this cycle. If we inhibit it by...*
Fourth year PhD: *I'm opening a bakery*

~ @PhDemetri, Twitter