TI 2025-052/VII
Tinbergen Institute Discussion Paper

The Ransomware Pricing Paradox: An Empirical Study of the Six Stages of Ransomware Negotiations

Tom Meurs[1]
Anna Cartwright[2]
Edward Cartwright[3]
Harold Houba[4]
Daniel Woods[5]

1 Dutch Police

2 Independent researcher

3 De Montfort University

4 Vrije Universiteit Amsterdam, Tinbergen Institute

5 University of Edinburgh

The Ransomware Pricing Paradox: An Empirical Study of the Six Stages of Ransomware Negotiations

Tom Meurs[*] Anna Cartwright[†] Edward Cartwright[‡]

Harold Houba[§] Daniel Woods[¶]

August 2025

[*] Dutch Police, Apeldoorn, Netherlands, Email:tom.meurs@politie.nl, ORCID: 0000-0003-0963-5232.

[†] Independent researcher, Kent, United Kingdom, Email: a.cartwright11@icloud.com, ORCID: 0000-0003-1965-842X.

[‡] Department of Accounting, Finance and Economics, De Montfort University, Leicester, LE1 9BH, United Kingdom, Email: edward.cartwright@dmu.ac.uk, ORCID: 0000-0003-0194-9368.

[§] School of Business & Economics and Tinbergen Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, Email: harold.houba@vu.nl, ORCID: 0000-0001-9085-7339.

[¶] School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, United Kingdom, Email: daniel.woods@ed.ac.uk, ORCID: 0000-0002-8569-1917.

Abstract

Ransomware has become the most common cyber risk for businesses. The rise is not driven by attackers using innovative attacks, but instead by deteriorating negotiation outcomes. The average payment grew by almost 20,000% since 2018. However, it remains unclear why attackers can demand ever higher ransoms. Our study explores potential explanations: lack of backups, cyber insurance, access to incident response (IR) firms, data exfiltration, and negotiating style. We model negotiation as a six-stage model: attacker intent, victim engagement, discount offer, discount magnitude, payment decision, and re-extortion. We test hypothetical explanations for ransom outcomes using two datasets: (1) 481 police-reported incidents (2019–2023); and (2) 237 negotiation transcripts from 23 ransomware groups.

We discover a pricing paradox: victims are more likely to pay after high initial demands, followed by large discounts, than after low fixed-price demands. Stage-level regression resolves this paradox: progression through stages is shaped by backup status, victim revenue, IR involvement, and negotiation duration. Fully recoverable backups sharply reduce payment rates and discount offers; higher revenue increases engagement and discount likelihood; and longer negotiations reduce payment. We find no evidence that insurance increases payment rates, that discount size matters once interaction is accounted for, or that re-extortion is common. These results position ransomware as a market-driven crime shaped by selection effects and signaling.

1 Introduction

Ransomware has evolved from automated software that demands a fixed amount into a sophisticated criminal enterprise in which attackers engage in selective targeting, dynamic pricing, and strategic negotiation [46, 51, 72, 5]. The change in business model coincided with a sharp increase in the frequency and impact of ransomware incidents. Ransomware comprised 40% of publicly reported incidents impacting businesses in 2024, up from around 1% in 2016 [30]. In terms of impact, one professional negotiator reports the average ransom payment rose from $6k in Q3 2018 [16] to $1.13m in Q2 2025 [17]. These rapid changes destablized the cyber insurance market [48, 67] and led politicians to declare national emergencies related to specific ransomware incidents [27].

The deterioration cannot be explained by ransomware actors using new attacks. Industry reports find the majority of incidents begin with stolen credentials [17, 65, 61], obtained via well-established cybercrime techniques like phishing [47, 23] or cybercrime marketplaces [22, 39]. When ransomware groups exploit software vulnerabilities, they typically rely on longstanding issues like memory safety [18, 54]. The lack of technical innovation suggests an alternative lens is needed to understand the rise of ransomware.

Poor negotiation is a possible explanation. Ransomware victims consciously bargain with attackers while the crime is taking place. By contrast, data breach victims do not interact with stealthy attackers, and phishing and fraud victims unconsciously interact with deceptive attackers. Negotiation allows attackers to exploit the victim's wish to avoid business interruption or publication of confidential data [11].

It is unclear why victims negotiate poorly. Some argue any payment is misguided because it encourages re-extortion of the original victim [41] and also future victims by allowing criminals to profit from crime [4]. IR firms argue negotiation by amateurs leads to outsized payments. Another problem is victims lacking reliable backups, a view that is complicated by the rise of threats to leak ex-filtrated data [41]. Looking beyond the victim, the insurance industry has been accused of incentivizing payment of ransoms [36].

These flaws in negotiation have motivated various proposals like taxing payments [49], banning ransom insurance [36], and the UK Government's recent proposal to ban CNI operators from paying ransoms [4]. These initiatives must be guided by empirical evidence, which motivates asking:

RQ1: Which factors influence ransomware negotiations?

RQ2: How does negotiation strategy impact discounts?

RQ3: What problems do victims face post-payment?

We answer **RQ1** by quantitatively modeling 481 ransomware incidents (2019–2023) as a six stage sequence. To answer **RQ2**, we analyze the meta data of 237 negotiation transcripts across 23 ransomware groups. We conduct a qualitative case-study of post-payment outcomes to answer **RQ3**.

Contributions. We find:

RQ1 A seeming paradox: cases with high initial ransom demands show higher payment rates than low fixed-price demands. Our 6-stage model resolves the paradox: high-demand, high-discount cases reaching payment are a filtered subset of incidents with characteristics that make payment more likely, not random variation in pricing.

RQ2 Sustained engagement, measured by message volume, is the strongest predictor of receiving a discount. Payment rates rise with message volume but plateau once negotiation frequency is high (100+ messages).

RQ3 Reinfections affected 3.6% of all cases. Among paying victims, 2.5% never received a decryption key and 6.7% experienced partial, delayed, or initially unusable decryption keys. Post-payment negotiation occurred 7.6% of the time.

Section 2 presents our conceptual contribution, a six-stage model of ransomware negotiation. Section 3 shows prior studies have focused on individual stages, not the entire process. Section 4 outlines our data sources and hypotheses. Section 5 presents the analyses. Section 6 discusses the findings. Section 7 concludes.

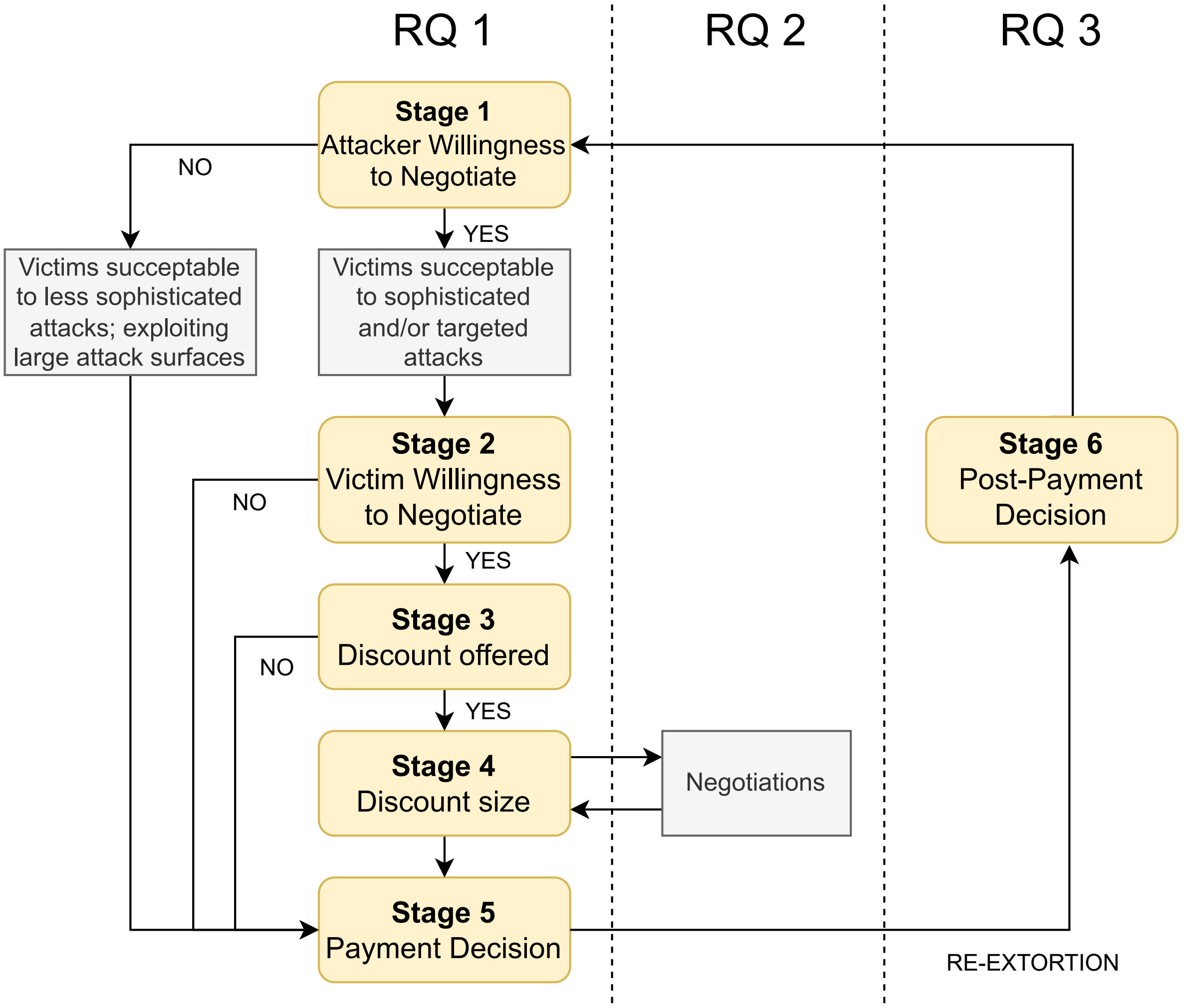
RQ 1
RQ 2
RQ 3
Stage 1
Attacker Willingness to Negotiate
NO
YES
Victims succeptable to less sophisticated attacks; exploiting large attack surfaces
Victims succeptable to sophisticated and/or targeted attacks
Stage 2
Victim Willingness to Negotiate
NO
YES
Stage 3
Discount offered
NO
YES
Stage 4
Discount size
Negotiations
Stage 5
Payment Decision
Stage 6
Post-Payment Decision
RE-EXTORTION

Figure 1: Six-stage model of ransomware negotiations, showing the sequence of decision points and the corresponding research questions (RQ1–RQ3) addressed in this study.

2 A Six-Stage Ransomware Negotiation Model

Ransomware payment is typically modeled as a binary event, namely pay or not [12]. In practice, negotiations must progress through a series of stages that necessitate engagement and trust from both sides.

We draw on three theoretical foundations. First, from the economics of crime and signaling theory, we posit that both attacker and victim use observable actions (e.g., ransom anchoring, message frequency, discounting) to infer the counterparty's willingness and constraints. Second, we borrow from industrial organization literature: attackers segment their "market" using forms of second-degree price discrimination, adjusting behavior based on expected willingness to pay (WTP) rather than offering fixed prices. Third, we recognize that ransomware negotiations defy the classic Coase Conjecture [7]: information asymmetries and time constraints mean the attacker can profit from sequentially lowering the ransom over time. This leads to a multi-stage bargaining rational.

This paper formalizes the bargaining process as a six-stage model. Initial compromise can be considered Stage 0, which is outside our scope. The stages are illustrated in Figure 1 and can be summarized as follows:

Stage 1: Attacker's Willingness to Negotiate. Negotiation requires the attacker to expend resources engaging with victims. Rational attackers will offer fixed-ransoms when the victim's expected ransom payment is lower than the cost of negotiating. This typically pccirs when attackers believe the victim lacks financial resources and/or the skills to facilitate payment via cryptocurrency. For high value victims, rational attackers provide contact details to create the opportunity for negotiation. This represents a form of market segmentation distinguishing victims susceptible to less sophisticated attacks versus 'high value' victims.

Stage 2: Victim's Willingness to Negotiate. When presented with contact details, some victims initiate communication, whereas others remain silent. This stage reflects the victim's assessment of expected losses, recovery prospects, and the informational value of dialogue. Thus, victim negotiation reflects cost-benefit reasoning under uncertainty: the direct and indirect costs of delay versus the informational value of communication. Firms may access expert negotiators—either in-house or via incident response firms—to inform the decision to negotiate.

Stage 3: Attacker Offers Discount. If the victim makes contact, the attacker may offer a reduction on the initial ransom demand. Attackers must assess the victim's perceived willingness to pay, often by observing message tone, urgency, or payment feasibility.

Industry lore warns victims that failure to pay within 48 hours doubles the ransom, but experienced negotiators know this is rarely enforced [43].

Stage 4: Attacker Discount Size. If the attacker is willing to offer a discount they next need to decide how much of a discount to offer. Larger discounts are sometimes offered in high-ransom cases or after prolonged engagement. However, attackers may adopt rigid pricing policies, especially in Ransomware-as-a-Service (RaaS) models, to avoid setting expectations of large discounts in future attacks. If the attackers holds information about insurance coverage and revenue size this may also anchor attackers' expectations.

Stage 5: Victim Payment Decision. After receiving an offer, the victim must make the decision of whether to pay the ransom. This includes weighing sanctions risk [63], the cost of recovering without a decryption key [21], the threat actor's reputation, and so on. Not all negotiated cases result in payments [43]. Potentially, the victim may request a further discount leading to repeated rounds of bargaining, a process we study in **RQ2**.

Stage 6: Post-Payment Decision. The final stage captures any further interactions between victim and threat actor. Attackers may exploit failed decryption or non-deletion of exfiltrated data to demand further payment. The trade-off is that further demands may reduce trust among IR companies, insurers, and negotiators. This may have negative consequences for future negotiations [60].

Our model reinterprets ransomware, not as a coercive ultimatum, but as a bargaining game shaped by anchoring, signaling, and iterative filtering. It also allows for empirical differentiation between one-stage "take-it-or-leave-it" attacks and multiple-stage bargaining processes.

Table 1: Ransomware studies mapped to our six-stage model.

				Investigated Stages					
Article	**Year**	**Data Type**	**n**	**1**	**2**	**3**	**4**	**5**	**6**
[12]	2019	Theory	-					✓	
[21]	2021	Decryptors	78						✓
[15]	2022	Cases	41					✓	
[46]	2022	Cases	453	✓				✓	
[43]	2023	Cases	382					✓	
[11]	2023	Survey	1,798					✓	
[10]	2023	Theory	-					✓	
[24]	2024	Survey	800					✓	
[40]	2024	Survey	445					✓	
[72]	2025	Theory	-			✓	✓	✓	
Us	2026	Cases	718	✓	✓	✓	✓	✓	✓

3 Related Work

This section maps prior research according to which stage of our model was studied. Technical research typically focuses on initial compromise, which cannot be mapped to our model. For research into negotiations, Table 1 shows we are the first to study all six stages of our model.

Stage 0. Computer security research into ransomware has mostly focused on how ransomware actors gain access and encrypt data. This involves proposing novel attacks [71, 52], detection heuristics [31, 73, 66, 75], and defensive measures [58, 34, 29]. These kinds of questions (how to attack, detect and prevent) are familiar to computer scientists. However, this line of research does not speak to the unique aspect of ransomware, active negotiation with criminals.

Stage 1—4. Until around 2016, ransomware actors largely offered fixed ransom demands [57], which means stages 2—4 were not explored. A 2016 study [59] found “only a very small portion of the victims actually pays the attacker”, largely because victims distrust attackers and lack the skills to facilitate cryptocurrency payments. It appears ransomware actors learned that willingness to negotiate led to greater profits, as evidenced by communications between attacker and defender influencing the initial demand [46]. Connolly and Borrion [15] provide examples of firms initiating negotiation. Anecdotal evidence suggests discounts (Stage 3 and 4) are common, offering a 50–90% reduction on the initial demand [5]. Zhang and Luo [72] proposed a theoretical bargaining model to study how negotiation influences ransom demands. We are not aware of work that studies all stages empirically or theoretically. Instead most studies focus on stage 5 in isolation.

Stage 5. Tracking ransom payments across blockchain networks has been an evergreen topic [53]. Huang et al. [28] tracked “$16 million USD in likely ransom payments made by 19,750 potential victims” covering ten ransomware families, which was just $850 per victim. A 2022 study found $80m of payments associated with one ransomware family (Conti), highlighting the success of later groups [25]. In a 2024 study, Cable et al. [8] introduced novel methods to identify $900m of ransom payments. Ransom payments are also tracked by for-profit firms like Chainanalysis, who traced $1.1 billion of payments in 2023 [13]. These aggregate estimates track the aggregate impact of ransomware, but not which factors increase whether and how much ransom is paid.

In practice, victims must weigh the payment decision by considering various legal obligations, such as contractual terms in cyber insurance [68], data breach notification obligations [26], and sanctions compliance [63]. Game theoretical studies have almost exclusively focused on Stage 5 and the payment decision (and stage 0 prevention measures) [12, 70] exploring issues such as data exfiltration [35, 42] and insurance [10]. Empirical surveys of ransom decision making have also focused on Stage 5 exploring the characteristics of individuals [11, 24] and small business owners/managers [40] willing to pay a ransom.

Stage 6. We are not aware of any studies that estimate re-extortion prevalence, even though it is frequently discussed [41]. It is also unclear how often decryption tools fail for victims, which is likely to be common given a 2021 study found that half of these tools failed in a lab setting [21]. Our article fills this gap by estimating the prevalence of both re-extortion and decryptor failure.

Frequency and Impact. Although influenced by negotiation outcomes, victimization studies are also outside our model. These studies show ransomware remains mercifully rare among individuals, with annualized frequency estimated at just 0.1% in a nationally representative survey [6]. Victimization rates among businesses range from around 1% to 60% depending on the study [69]. Recent capture–recapture estimates for the Netherlands suggest an annual risk of 1.3% for large companies and 0.6% for medium companies, with reporting rates to police around 40% [45]. In terms of impact, the average ransomware incident cost is estimated to cost over $1 million in a study of public incidents [30]. The mean size mean value of cyber insurance claims resulting from ransomware was $292k in 2024 [14]. These quantitative estimates do not account for impacts like emotional stress and support networks disruptions [38, 74].

4 Methodology

Studying ransomware negotiations is challenging because they are rare and occur in private channels. Section 4.1 explains how we collected a convenience sample of historic ransomware cases from three data sources: the police; an incident response firm; and an open-source repository. Section 4.2 describes how we analyzed the data. Section 4.3 identifies ethical considerations in studying victims of crime.

4.1 Data Collection

We obtained anonymized data from the police and an IR firm. Both organizations keep detailed records of historic incidents, which we could reconcile to create a combined dataset (see Section 4.1.1). To study negotiations, we collected anonymized negotiation transcripts from *ransomware.live* (see Section 4.1.2).

4.1.1 Incident Data

The dataset consists of a combination of 525 ransomware attacks reported to the Police between 1 January 2019 and 1 January 2023 and a dataset including 116 ransomware incidents handled by an IR company between 20 February 2020 and 1 January 2023. By combining law enforcement and private sector sources, we help mitigate reporting biases that may arise when victims are reluctant to report to authorities.

From these datasets, we excluded attempted attacks and incidents involving individuals, retaining only successful attacks targeting companies. After removing duplicates between the datasets, we constructed a combined dataset of 481 unique ransomware attacks on organizations.

The variables used in the analysis are as follows. The numbering reflects the stage of the six-stage negotiation model at which each variable is iteratively added to the regression framework.

- **1a. Revenue (log).** Logarithm (base 10) of the estimated revenue of the victim organization in euro, used as a proxy for organizational size and financial capability.
- **1b. Ransomware-as-a-Service (RaaS).** Indicates whether the attack was conducted by a known Ransomware-as-a-Service group (yes/no).
- **1c. Backups: Unrecoverable.** Indicates whether the victim's backups were rendered unusable due to the attack.
- **1d. Partial**. Indicates whether the victim had partially functioning backups at the time of the attack.
- **1e. Recoverable.** Indicates whether the victim had fully functioning and restorable backups during the incident.
- **1f. Data exfiltration.** Whether data exfiltration occurred as part of the ransomware attack (yes/no).

- **2a. Cyber insurance.** Indicates whether the victim had cyber insurance that covered ransomware (yes/no).
- **2b. Incident Response (IR) firm involvement.** Whether an incident response firm assisted the victim during the attack (yes/no).
- **3. Negotiation duration (log).** Logarithm of the total number of days in which negotiation between attacker and victim took place.
- **4a. Initial ransom demand (log).** The initial ransom demand made by the attacker in euro, expressed as a base-10 logarithm.
- **4b. Negotiation (attacker).** Binary indicator for whether the attacker engaged in negotiation.
- **4c. Negotiation (victim).** Binary indicator for whether the victim engaged in negotiation.
- **5a. Discount offered (binary).** Indicates whether the attacker offered a discount on the initial ransom amount during negotiation (yes/no).
- **5b. Discount size (log).** The size of the discount in euro, expressed as a base-10 logarithm.
- **6. Payment (yes/no).** Indicates whether the victim ultimately paid the ransom.

4.1.2 Negotiation Transcripts

To test the middle stages of our model, we analyzed a corpus of 237 negotiation transcripts from 23 ransomware groups collected from *ransomware.live*. We collected a snapshot measurement on July 31, 2025. Victims are located from around the world. For each transcript, we extracted metadata that map directly to Stages 3 to 5 of the model. This leads to the following variables, which are tested in a separate regression to the data from the previous subsection:

- **7a. Message volume.** A proxy for negotiation depth and progression
- **7b. Initial ransom demand (log).**
- **7c. Discount offered (binary).**
- **7d. Discount size (proportion ransom).**
- **7e. Payment (yes/no).**

The following extract illustrates the transition from opening pressure to anchoring and then bargaining:

> **Akira:** Hello. You've reached Akira support chat. Currently, we are preparing the list of data we took from your network. For now, you have to know that dealing with us is the best possible way to settle this quick and cheap. Keep in touch and be patient with us. Do you have permission to conduct a negotiation on behalf of your organization? Once we get your reply, you will be provided with all the details.
>
> **Victim:** Yes. What to do?
>
> **Akira:** So, we didn't take your data. We are the ones who can properly decrypt your data and restore your infrastructure in a short period of time. [...] The price is $1,000,000. To prove we can decrypt your data, you can upload 2–3 encrypted files up to 10MB.
>
> **Victim:** Alright, is it negotiable? As you can see, our true finance register. Also, let us know what all files you have of ours?
>
> **Akira:** We do not have your files. Do you have a counteroffer for me?

Consistent with the model, the exchange begins with a generic pressure message, followed by a request that the victim signal willingness to engage in negotiations. Substantive pricing appears only after that signal (Stage 1 and 2).

4.2 Analysis

The ransomware bargaining process is conceptualized as a six-stage model, reflecting sequential decisions made by both offenders and victims. We draw on economic theory and bargaining game theory, as well as existing research on ransomware negotiations, to derive testable hypotheses.

Using forward induction we start by looking at Stage 1. Negotiation is costly for attackers in terms of time, expertise required, and the opportunity cost, of foregone profit, by exploiting fewer victims at any one time. The attacker will, therefore, go the route of indicating a willingness to negotiate only if the expected ransom they can earn is substantially higher. We consequently predict market segmentation with low ransom demands for automated attacks with no negotiation and substantially higher initial ransom demands when the attacker is willing to negotiate. Moreover, we predict higher initial ransom demands if the attacker has more leverage, as proxied by corrupted backups and/or data exfiltration.

The market segmentation in Stage 1 means that Stages 2, 3 and 4 will take place between an attacker who has invested in the attack and a victim that is likely to be high worth. In Stage 2 the victim's strategy will rationally informed by their willingness to pay. This will depend on their readiness and preparedness for an attack in terms of back ups.

Hypothesis 1. Victim preparedness reduces the likelihood of negotiation.

The presence of insurance and access to an IR firm may encourage negotiation as it means the victim has access to relevant services for negotiation.

Hypothesis 2a. Cyber insurance and access to IR firm increases likelihood of negotiation.

In Stages 3 and 4 the attacker is in negotiation with the victim. Negotiation allows the attacker to update their belief about the willingness to pay of the victim. For instance, victims who are more prepared for an attack in terms of, for example, having functioning back ups can hold out for a lower ransom.

Hypothesis 3. Victim preparedness increases the likelihood of a ransom discount and the size of discount.

A victim's willingness to be patient in negotiations is also a signal of lower willingness to pay.

Hypothesis 4a. Longer negotiations lead to larger discount amounts.

Further, access to experts may facilitate negotiation.

Hypothesis 2b. Insurance coverage and IR firm involvement increase the likelihood of a ransom discount and the size of discount.

Data exfiltration allows the attacker to offer a 'menu of ransom options' covering encryption and/or exfiltration. This may result in larger discounts being offered, such as when the victim is only willing to pay to cover data exfiltration.

Hypothesis 5. Data exfiltration increases the likelihood of negotiation and discount in negotiations.

In Stage 5 the victim will rationally pay if their willingness to pay is higher than the final ransom demand. This will depend on factors like preparedness for attack, the presence of insurance, and IR firm involvement. This has been studied in previous work [43]. The novelty in our six-stage model is to capture the stages of negotiation. Bargaining theory suggests that if negotiation has taken place, and discounts offered, then it is more likely the victim and attacker will agree on a mutually beneficial ransom amount.

Hypothesis 6. A willingness of the attacker and/or victim to negotiate and ransom discounts increase the likelihood of ransom payment.

Longer negotiations can be a signal of a failure to reach agreement because the ransom the victims are willing to pay is below the minimum amount the attackers are willing to accept (potentially because of reputation reasons).

Hypothesis 4b. Longer negotiations decrease the probability of payment.

For the sixth stage (post-payment outcomes), there were too few incidents to test the hypotheses quantitatively, so we instead describe the outcomes. Stage 6 of the model is part of a 'larger game' in which the attackers trade-off the potential gains from re-extortion or re-infection with the reputational loss that will come from doing so. It is in the interests of attackers to build a reputation for being 'trustworthy' post payment if that will increase future ransom payments [9, 60]. We would, therefore, hypothesize that unworkable decryption keys, re-extortion and re-infection are relatively rare. In our data we analyzed cases that involved post-payment interaction with the threat actor (**RQ3**). These were identified through manual review of case records, focusing on incidents where victims reported faulty decryptors, renewed demands, or reinfection after payment.

Statistical Tests For the first five stages of the model (M1–M5), we ran regression analyses. We used probit models for binary outcomes (e.g., attacker negotiation in M1, victim negotiation in M2, discount offered in M3, and payment in M5) and ordinary least squares (OLS) for continuous outcomes (e.g., discount size in M4). All models control for basic victim and attacker characteristics like yearly revenue and RaaS ransomware groups) as controls. We iteratively add predictors at each stage according to the variable numbering described above. Goodness-of-fit metrics (pseudo-R^2 for probit models and adjusted R^2 for OLS models) were also calculated; detailed values are reported in the Appendix. The primary purpose of these models is to identify associations across stages rather than to maximize predictive accuracy.

4.3 Ethics

This study combines structured incident-level data with publicly available ransomware negotiation transcripts. The structured dataset was obtained from law enforcement and a commercial incident response partner under data-sharing agreements. All records were anonymized before analysis and we focused on incidents where the victim was an organization. None of the records contained no personally identifiable information (PII) about individuals involved in the incident. The public negotiation data were scraped from the `ransomware.live` platform, which aggregates chat logs already published by ransomware groups on leak sites. Although these data are publicly accessible, we further anonymized details of the incident to reduce the risk of reputation damage and re-victimization. For example, our case studies (**RQ3**) do not mention the ransomware group.

The study did not involve direct interaction with human subjects, and no identifiable or sensitive organizational data were collected without prior anonymization. Ethical approval was therefore not required under the law or university policy. Nevertheless, we adhered to the MENLO principles of responsible research [3], including minimizing potential harm and avoiding the reinforcement of criminal behavior. We acknowledge that studying offender decision-making may, in theory, expose operational weaknesses or biases that could be exploited. However, we judge that the primary insights generated here are of far greater value to defenders than to offenders, and thus contribute to harm reduction.

We recognize the ethical tension inherent in studying ransomware payments, like prior work [63]. We do not endorse ransom payments, instead we see payment as, in some cases, the least bad option for organizations facing severe operational and legal pressure. Our analysis aims to inform victim strategies, not to legitimize payment decisions.

5 Results

We present the results for each research question in turn.

5.1 Factors Influencing Negotiations (RQ1)

5.1.1 Descriptive Results

Figure 2 shows that cases with higher initial ransom demands are more likely to end in payment than cases with low, fixed demands and no negotiation. This pattern appears paradoxical: why would victims comply more readily when the starting price is higher and

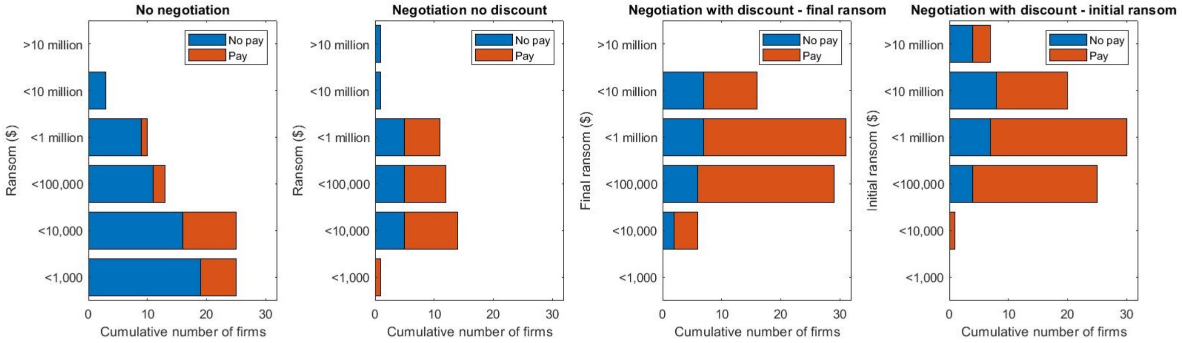

Figure 2: Negotiation outcomes by ransom amount distinguishing cases with no negotiation and negotiation. In the case of negotiation with discount we provide data for both the final ransom demand and initial ransom demand.

the final amount remains substantial?

Our staged model helps explain this. Each step in the process functions as a filtering mechanism, progressively narrowing the set of incidents toward those most likely to result in payment. The subset of cases that reaches Stage 4 (payment) is not randomly distributed across all ransom amounts and strategies—it is shaped by prior strategic interaction.

In opportunistic attacks (typically characterized by Stage 1 = No), the attacker sets a fixed, low ransom without negotiation. These cases resemble a one-stage ultimatum game, where victims must decide whether the ransom is worth paying based solely on immediate cost. As shown in Figure 2, such cases have low payment rates, especially at ransom levels below €10,000.

In contrast, cases that proceed through multiple negotiation stages (typically characterized by Stage 1 = Yes) follow a dynamic game structure. The attacker opens with a high initial ransom demand (Stage 1), the victim engages (Stage 2), and a discount is offered (Stage 3 and 4). Although the final ransom is higher than in typical opportunistic cases, the process itself may reshape victim perception. If the victim's experience is anchored on the initial high price, a discount may appear generous or as good-faith behavior. In the terminology of Thaler [62] the victim receives transaction utility from a 'good' or 'fair' deal. This process also appears less coercive when the victim is given agency to negotiate and explain their limitations in paying the high initial demand.

Table 2: Negotiation Outcome by Initial Ransom and Discount Percentage. Cell values represent counts, shaded by intensity.

Initial Ransom	**0–25%**		**26–50%**		**51–75%**		**76–100%**	
(euro)	No Pay	Pay	No Pay	Pay	No Pay	Pay	No Pay	Pay
<1,000	0	0	0	0	0	0	0	0
<10,000	0	1	0	0	0	0	0	0
<100,000	0	5	4	7	0	8	1	2
<1 million	0	5	3	12	1	3	4	4
<10 million	2	3	3	7	2	0	1	2
>10 million	0	0	1	0	2	1	1	2

Table 3: Negotiation outcomes per ransomware group (source: ransomware.live, N=237)

Group	*n*	Init. %	Paid %	Negot. %	Med. Init	Mean Init	Med. Neg	Mean Neg	Med. Paid	Mean Paid	Mean Disc. % (Med.)
Akira	61	67.2	32.8	44.3	390k	948k	150k	248k	150k	256k	22.4 (0.0)
Avaddon	7	85.7	14.3	57.1	190k	304k	90k	103k	1.3k	1.3k	32.3 (33.3)
Avos	1	100.0	0.0	100.0	150k	150k	85k	85k	–	–	43.3 (43.3)
Babuk	2	100.0	0.0	100.0	700k	700k	293k	293k	–	–	64.4 (64.4)
BlackBasta	5	80.0	80.0	100.0	600k	1.09M	300k	632k	225k	275k	31.2 (28.6)
BlackMatter	2	100.0	0.0	100.0	10.0M	10.0M	14.3M	14.3M	–	–	-95.0 (-95.0)
Cloak	2	0.0	0.0	0.0	–	–	–	–	–	–	0.0 (0.0)
Conti	32	93.8	31.2	65.6	1.04M	7.93M	325k	1.68M	171k	224k	40.9 (40.0)
Darkside	5	80.0	60.0	60.0	1.50M	3.40M	250k	283k	250k	283k	24.7 (0.0)
Dragonforce	14	50.0	0.0	21.4	1.01M	859k	124k	464k	–	–	5.8 (0.0)
Hive	8	100.0	0.0	62.5	1.10M	3.78M	276k	3.98M	–	–	21.4 (15.4)
Hunters International	1	100.0	0.0	0.0	10.0M	10.0M	–	–	–	–	0.0 (0.0)
Mallox	3	66.7	33.3	100.0	28.8k	28.8k	20k	23.3k	19.9k	19.9k	6.7 (0.1)
NoEscape	2	50.0	0.0	0.0	80k	80k	–	–	–	–	0.0 (0.0)
Qilin	2	0.0	0.0	0.0	–	–	–	–	–	–	0.0 (0.0)
REvil	20	70.0	35.0	50.0	2.75M	3.12M	290k	425k	280k	348k	34.9 (20.6)
RansomHub	1	0.0	0.0	0.0	–	–	–	–	–	–	0.0 (0.0)
Ranzy	2	0.0	0.0	0.0	–	–	–	–	–	–	0.0 (0.0)
RunSomeWares	1	0.0	0.0	0.0	–	–	–	–	–	–	0.0 (0.0)
fog	6	100.0	50.0	50.0	245k	329k	150k	917k	150k	917k	-3.7 (12.5)
lockbit3.0	45	91.1	6.7	31.1	1.40M	5.80M	1.00M	3.11M	40k	40k	7.7 (0.0)
mount-locker	1	100.0	0.0	100.0	9.00M	9.00M	4.12M	4.12M	–	–	54.3 (54.3)
trinity	14	42.9	7.1	0.0	30k	139k	–	–	15k	15k	0.0 (0.0)

A complementary explanation is that victims with sufficient financial capacity or professional guidance, such as via IR firms, tend to reach the later stages. This induces a selection effect: higher-value victims are both more likely to be targeted for negotiation and more likely to pay, conditional on progressing through the funnel. The result is a strategic equilibrium in which high initial demands, followed by discounts, yield the most profitable outcomes for attackers.

The amount of the discount itself does not show a linear relationship with payment rates (Table 2). This supports the view that discounting functions not purely as price

Table 4: Direction and magnitude of significant predictors across first five stages of ransomware negotiation and payment.

Predictor	M1: Attacker negotiation	M2: Victim negotiation	M3: Discount given	M4: Discount size	M5: Payment
1a. Revenue (log)	ns	+	ns	-	ns
1b. Ransomware-as-a-Service (RaaS)	ns	ns	ns	-	ns
1c. Backups: Unrecoverable (vs. no backups)	+	ns	ns	ns	ns
1d. Backups: Partial (vs. no backups)	+	ns	ns	ns	ns
1e. Backups: Recoverable (vs. no backups)	+	--	--	ns	--
1f. Data exfiltration (yes/no)	+	ns	ns	-	ns
2a. Cyber insurance (yes/no)	.	ns	ns	-	ns
2b. IR firm involvement (yes/no)	.	++	++	ns	ns
3. Negotiation duration (log)	.	.	+	ns	--
4a. Initial ransom demand (log)	.	.	.	++	ns
4b. Negotiation (attacker)	.	.	.	ns	ns
4c. Negotiation (victim)	.	.	.	ns	++
5a. Discount offered (yes/no)	.	.	.	.	++
5b. Discount size (log)	.	.	.	.	ns

Legend: ns = non-significant; ++ = significant, $\beta > 1$; + = significant, $0 < \beta < 1$; -- = significant, $\beta < -1$; - = significant, $-1 < \beta < 0$; . = predictor not applicable for that stage.

reduction but as a signal of flexibility or legitimacy. In other words, it is the presence of a discount conditional on prior interaction, and not the amount, that drives payment.

In summary, both attackers and victims strategically engage in negotiations. However, analysis of outcomes cannot explain how these decisions are shaped by the profile of the victim or the attacker. The next set of results explore how the victim's profile shapes these outcomes, and Table 3 speaks to how these outcomes vary across threat actors.

5.1.2 Regression Results

The full regression results, including estimated β-coefficients and p-values, are presented in the Appendix (see Table A1). Table 4 summarizes the direction and magnitude of significant predictors for each stage with $\alpha = 0.05$.

Stage 1. Consistent with Hypothesis 5, attackers are more likely to open negotiations if they have exfiltrated data (see Table 4). We found no evidence that attackers are more likely to negotiate with victim's with higher revenue, possibly because this is not clear to attackers until they begin negotiating.

The presence of any form of backups (unrecoverable, partial, or recoverable) is associated with increased likelihood that the attacker initiates negotiation. This appears counter-intuitive given that restoring from backups provides an alternative to paying the ransom. However, it likely results from a selection effect: organizations that maintain backups may also hold more valuable or complex data, making them worthwhile targets for attackers to engage with more deliberately. The existence of backups may correlate with other attributes (such as larger infrastructure or higher-value data) that justify attacker investment in a negotiation platform.

Stage 2. Victims with higher revenues are more likely to negotiate with the attacker. Potential explanations include greater capacity to respond and/or having more to lose from an extended outage. Consistent with Hypothesis 2a, IR involvement was also strongly predictive, as evidenced by a larger effect size ($\beta > 1$). The presence of experts facilitates or recommends negotiation as part of incident handling. Consistent with Hypothesis 1, victims with recoverable backups were much less likely to negotiate, likely because they had a viable alternative to paying the ransom.

Stage 3. The likelihood of a discount being offered by the attacker was significantly associated with the victim's revenue, negotiation duration, and IR involvement (see Table 4). This is consistent with Hypotheses 2b and 4a. Longer negotiation duration had a particularly strong effect, indicating that discounts often emerge as part of prolonged back-and-forth interaction. This supports the interpretation that discounts function as dynamic price discrimination: high-revenue victims may be more capable of paying, but also more demanding, prompting attackers to lower their price to close a deal.

Stage 4. The size of the discount was associated with several variables. As expected, higher initial ransom demands allowed more room for larger discounts. However, discount size decreased with higher revenue, RaaS involvement, and cyber insurance. This suggests that attackers operating under a RaaS Model use more rigid pricing strategies, possibly to maintain reputation or enforce affiliate compliance. Insurance coverage may also anchor expectations, reducing the need for aggressive discounting. Interestingly, negotiation duration showed a positive but only marginally relationship with discount size, indicating that extended interactions may influence not just the decision to discount but also how much.

Table 5: Negotiation outcomes by message volume bucket (source: ransomware.live, N=237)

Bucket # msg	*n*	Init. %	Paid %	Negot. %	Med. Init ($)	Mean Init ($)	Med. Neg ($)	Mean Neg ($)	Mean Disc. % (Median)
0–10	43	41.9	0.0	0.0	1.35M	8.82M	0	0	0.0 (0.0)
11–25	44	65.9	4.5	22.7	0.40M	0.79M	225k	391k	23.3 (23.3)
26–50	65	80.0	20.0	46.2	0.80M	2.44M	285k	1.24M	42.2 (42.0)
51–100	63	92.1	44.4	73.0	0.53M	3.48M	160k	1.44M	49.5 (58.3)
100+	22	90.9	45.5	86.4	0.95M	7.38M	250k	2.32M	65.1 (65.8)

Stage 5. Consistent with Hypothesis 6, payment was more likely when the victim engaged in negotiation and when a discount was offered. Longer negotiations were negatively associated with payment, consistent with Hypothesis 4b. Prolonged negotiations may indicate tactical stalling, disagreement, mistrust, or internal delays on the victim's side (e.g., legal or executive hesitations) or a failure to reach agreement. Victims with fully recoverable backups were less likely to pay, consistent with being less likely to negotiate. The initial ransom amount had a negative but statistically insignificant impact (p = 0.068; see Appendix A).

5.2 Public Ransomware Negotiations (RQ2)

The negotiation dataset spans various ransomware groups (see Table 3). Four ransomware groups represent the majority (58%) of the negotiations: Akira (n=61), Lockbit 3.0 (n=45), Conti (n=32), and REvil (n=20). Akira, REvil and Conti all had a similar payment rate (31–35%), with comparable average discounts of around 20–40%. By contrast, Lockbit 3.0 appears to be less flexible with a mean discount of just 7.7%, but also much lower payment rates (6.7%) suggesting a sub-optimal strategy.

Analyzing the data in aggregate, Table 5 shows that negotiation length strongly correlates with the probability and size of a ransom discount, and the payment rate. In very short exchanges (0–10 messages), no discounts were observed and no victims paid. Once talks extended to 11–25 messages, 23% of cases saw a discount (median 23%), with payment

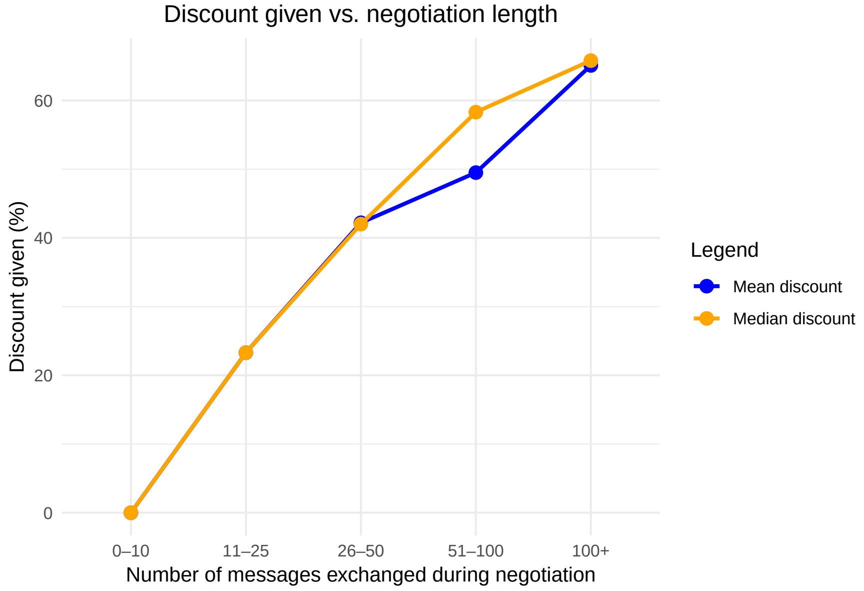

Figure 3: Relationship between message volume and likelihood of discount.

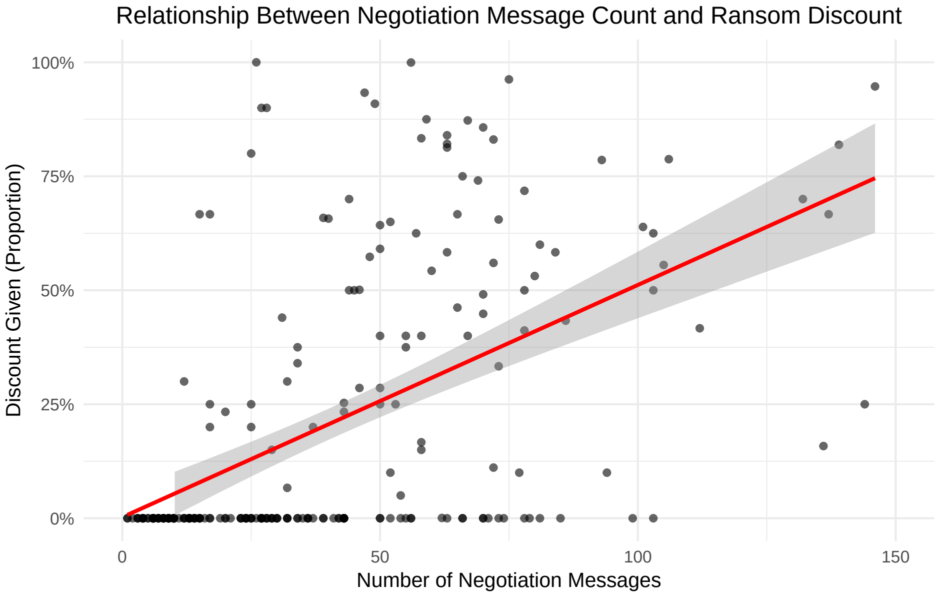

Figure 4: Relationship between message volume and likelihood of discount.

rates still under 5%. Between 26 and 50 messages, nearly half of cases received a discount (median 42%) and one in five victims paid. The longest negotiations (100+ messages) yielded discounts in 86% of cases (median 66%) and had payment rates approaching 46%.

Figure 3 visualizes these stepwise increases by message-count bucket, showing near-linear growth in both mean and median discounts. Figure 4 plots discount percentage against raw message count, with a fitted line and confidence interval. While some short exchanges still yield large discount, the overall trend is that more messages lead to larger discounts, consistent with Hypothesis 4a.

To formalize these patterns, we estimated two separate logistic regression models:

Regression 1. Discount (Stage 3): The dependent variable is whether a discount was given (yes/no). Independent variables are: (a) log-transformed message count, and (b) log-transformed initial ransom amount. Message count is a strong positive predictor (OR $\approx$ 52.8, 95% CI: 15.5–216.2, $p < .001$), while initial ransom size is not statistically significant ($p = .12$).

Regression 2. Payment (Stage 5): The dependent variable is whether the victim paid (yes/no). Independent variables are: (a) log-transformed message count, (b) discount size as a proportion, and (c) log-transformed initial ransom amount. Message count is again the dominant predictor (OR $\approx$ 50.6, 95% CI: 5.5–500.6, $p < .001$). The discount variable is not significant after accounting for message count, and larger initial demands significantly reduce the likelihood of payment (OR $\approx$ 0.27, $p < .001$).

Taken together, the results indicate that sustained engagement rather than the size of the discount is the main driver of settlement. Larger initial demands suppress payment and shorten the negotiation, while brief exchanges almost never end in agreement. These results reinforce the importance of analyzing all the stages of negotiation and are consistent with the results in Table 4: prolonged bargaining is associated with discounts, but resilience measures and the size of the initial ransom ultimately determine whether payment occurs.

5.3 Post-Payment Outcomes (RQ3)

So far, we examined 481 ransomware incidents, of which 119 victims paid (24.7%). Stage 6 captures post-payment interactions with threat actors. Three post-payment failure modes emerged:

1. The decryptor was not delivered or functioned poorly.
2. The attacker renegotiated for additional payment after an initial settlement.
3. The victim was reinfected by the same or another ransomware strain.

These outcomes were relatively rare, which suggests an orderly system where ransom contracts are mostly respected.

Decryptor Failures. Among the 119 victims who paid, 3 cases (2.5%) involved no decryptor being provided, and 8 cases (6.7%) involved a decryptor that was initially unusable or only partially effective. The observed problems included keys that decrypted only certain file types, restored files without making systems bootable, and waiting several days before a working key was supplied.

In one manufacturing case (2020), the attacker initially demanded €38k[1], which was then negotiated down to €22k (40% discount) in a single day. Backups had been deleted during the attack, and only one of three decryption keys was provided at first. The victim persuaded the attacker to release the remaining keys without additional payment by stressing the reputational damage that non-delivery would cause. In a service-sector case (2020), the victim paid €300k after a two-hour negotiation with no discount, yet waited two more days for a functional key and experienced nine days of total downtime.

The three "no key" incidents affected one SME and two large organizations. In all cases, backups ranged from partial to non-existent, and the attacker cut off contact immediately after receiving payment. These figures are likely lower bounds given incomplete post-incident reporting.

Renegotiation. At least 9 of the 119 paying victims (7.6%) faced additional ransom demands after making the first payment. This typically occurred when the attacker claimed the first payment covered only part of the encrypted assets (e.g., certain servers) or when

[1] Throughout we perturb the values for anonymity.

additional “fees” were invented after payment. Notably, this did not involve repeat demands following non-deletion of exfiltrated data.

In one private sector case (2021), the attacker accepted an initial €10,000 payment but then demanded approximately €1.3million after claiming to charge €10k per server. The organization, with annual revenue between €100 million and €1 billion and no backups, incurred €2.5–3.5 million in recovery costs instead of paying the ransom.

In another manufacturing case in 2020, a €6k payment to a ransomware-as-a-service variant was immediately followed by a request for €4,500 more; the second payment finally unlocked all files. A third case began with a €24k payment for three servers, escalated to €32k for the remaining four, and ended with a final settlement of €10k more. In the renegotiation cases, backups were almost always absent or unrecoverable, leaving victims with little leverage to resist further demands.

Reinfection. Seventeen victims (3.5% of all 481 incidents) experienced another ransomware incident after a previous one. These fell into four categories:

Category 1: Immediate reinfection occurred in two cases, when victims restored backups without fully cleaning systems. Backups enabled recovery, but rushed restoration prolonged the outage. No incident response (IR) was initially involved.

Category 2.1: Delayed reinfection, no payment in the second incident accounted for 7 cases. In at least two incident, the victim disclosed that the earlier incident had prompted investment in improved backup strategies (including immutable backups), allowing rapid recovery without payment.

Category 2.2: Delayed reinfection, payment in the second incident comprised 7 cases. Most victims in this category disclosed that their IT infrastructure lacked network segmentation, enabling attackers to locate and encrypt or delete backups in the later incident. The gap between incidents was at least one year. No clear link between repeated attacks could be established.

Category 3: No successful reinfection occurred in 1 case (0.2%), about 18 months after the first attack. The intrusion was detected early enough to prevent encryption.

Across all categories, no victims paid in the first incident and then experienced a second infection within one year. Immediate reinfections occurred only when the victim chose not to pay, typically where backups were restored too quickly without fully cleaning compromised systems. This suggests threat actors can largely be trusted to honor the ransom agreement, at least in our sample. Payment itself was likely not a determining factor for reinfection, as some victims in categories 2.1 and 2.2 had paid in the first incident, while others had not.

5.4 Summary

The six-stage model reframes the paradox: victims do not pay because the ransom is low. Instead victims pay because the negotiation structure leads them to believe that payment is the least bad outcome. Section 5.1 shows how high initial demands enable this structure, while discounts facilitate closure. Extended message exchanges during negotiations gives victims a sense of agency as the discount increases, as evidenced by Section 5.2. Further, the relative infrequency of post-payment complications in Section 5.3 showed that attackers mostly honor the agreement, thereby increasing trust and managing their reputation.

6 Discussion

We now discuss the implications, limitations and recommendations implied by the empirical analysis of our six-stage ransomware negotiation model.

6.1 Implications

Our results relate directly to common misconceptions in the ransomware literature and practice community, which are discussed in turn.

Technical Controls. The majority of security controls are designed to prevent ransomware (and other security incidents). By contrast, backups do not help to prevent or detect incidents. However, our results show they are valuable because they reduce the victim's willingness to negotiate and pay (see Table 4). This explains why backups are one of the only two controls that appear in the cybersecurity guidance of all 41 countries in one study [56], given policymakers are so focused on ransom payment [4]. However, the government guidance is inconsistent regarding how backups should be maintained [56].

This matters because victims are doing something wrong given the prevalence of partial or non-functioning backups, which provide no benefit over no backups (see Table 4). Unfortunately, our data sources did not identify why these backups failed. Potential explanations include: (i) attackers destroying backups; (ii) misconfigured or corrupted backups; (iii) out-of-sync backups; and (iv) delays transferring data. Offsite backups are widely recommended [56], largely because the physical separation is a reliable form of segmentation that prevents type (i) backup failures. Counter to this advice, one industry study found offsite backups had the highest failure rate (55% compared to 80% for cloud) [32], possibly because offsite backups are harder to maintain, leading to type (ii—iv) failures. Exploring why backups fail is a promising area of future work, most likely for usable security given the problems are likely rooted in organizational processes.

Discounts. Game-theoretic models treat discounts not as unconditional generosity but as signals in a screening game [12, 72]. In these models, attackers use early or large discounts to extract information about willingness to pay. Our staged results align with this theory by showing that discounts are common, correlate with negotiation length, and help close deals. However, the size of the discount is a weak predictor once preparedness and initial pricing are accounted for. The misconception that "larger discounts cause payment" is not supported. Preparedness and initial price remain the dominant levers, whereas discounts are better understood as persuasion and closure devices emerging in structured, multi-round bargaining.

In terms of victim psychology, discounts are known to "increase purchase satisfaction" [19]. This motivates regulatory scrutiny to ensure they are not deceptive or manipulative, which is reflected in regulatory guidelines [50, 20]. Unfortunately, ransomware operates within the black market. This lack of oversight means ransomware actors can employ any persuasive tactics they choose, including arbitrary or misleading discounts, without being held accountable for consumer harm. Increased data sharing between victims, law enforcement, IR companies and insurers would allow better understanding of 'typical' ransom amounts to judge whether 'discounts' are genuine.

Insurance. The insurance industry has faced criticism that cyber coverage encourages payments [33], a view that is supported by evidence ransomware groups search for cyber insurance policies post-infection [1]. Cartwright et al. [10] model and empirically show that insurance can raise payment probability in some settings, mainly through liquidity provision and claims handling, but the effect is conditional on exclusions and the timing of insurer involvement. In our six-stage model, insurance did not predict whether victims engaged or paid. However, insurance pushes victims to hire IR professionals [68], which raises engagement and the chance of receiving a discount. The latter effect is offset by insurance being associated with smaller discounts (Stage 4).

The null effect on payment can be explained by insurance also covering lost revenue due to the outage [55]. Business interruption coverage means victims can incur longer outages, offsetting the incentive to pay a ransom partly covered by the insurer. Thus, insurance does not affect the likelihood of paying a ransom because it symmetrically affects both sides of the calculus. However, both types of coverage increase the magnitude of cost that can be incurred as either payment or lost revenue. This partly explains why insurance is associated with higher ransom amounts [42]. Another explanation is selection effects, whereby organizations with greater IT dependency are more likely to buy cyber insurance.

Table 6: Summary of post-payment outcomes. Frequencies for decryptor failures and renegotiations are relative to 119 paying victims; reinfection frequencies are relative to all 481 incidents. Observations are derived from case study analysis.

Category	Subtype	Frequency (%)	Key Observations from Case Studies
1. Decryptor failure	No key delivered	3/119 (2.5%)	Actor ceased contact after payment; affected one SME and two large organizations; backups ranged from none to partial.
	Partial or delayed or initially unusable	8/119 (6.7%)	Keys worked only for some file types or were delayed; in some cases systems remained unusable after decryption; recovery often extended by several days despite payment.
2. Renegotiation	Aggregated	9/119 (7.6%)	Attackers demanded further payment after initial settlement, often claiming first payment covered only a subset of servers or data; price hikes ranged from small to over 100×; backups generally absent or destroyed.
3. Reinfection	Category 1: Immediate reinfection	2/481 (0.4%)	Both involved restoring backups without cleaning infected systems; reinfection occurred within days, but recovery still possible from backups.
	Category 2.1: Delayed reinfection, no payment in second attack	7/481 (1.5%)	Earlier incident sometimes prompted investment in stronger backups (including immutable backups) which made later recovery easier.
	Category 2.2: Delayed reinfection, payment in second attack	7/481 (1.5%)	Lack of segmentation allowed attackers to delete or encrypt backups in second incident; most gaps between incidents exceeded one year.
	Category 3: No successful reinfection	1/481 (0.2%)	Later intrusion detected about 18 months after first; encryption prevented in time.

Reputation & Re-extortion. Theoretical reputation models predict that systematic re-extortion—failing to deliver decryption keys, demanding more after payment—damages attacker credibility and thus future profits [9]. Our data supports the theoretical prediction that while some failures occur, repeat cheating should be rare. Many repeats appear linked to technical remediation gaps rather than deliberate re-extortion policy (see Table 6). This is further supported by prior work showing that certain attacker TTPs (e.g., maintaining decryption reliability) are re-used because they serve as a credible signal of professionalism [63].

This suggests both attackers and IR firms are engaged in not just a multi-stage game within each incident as in Figure 1, but a multi-stage game across incidents where bad actors preserve reputation. While this creates predictable outcomes for individual actors, it is unclear whether an orderly criminal ecosystem is optimal for society. For comparison, researchers recommend disrupting trust networks in other areas of cybercrime [37, 2].

Take-down operations largely avoid these problems by striking the operational capacity of ransomware groups [44, 64].

6.2 Limitations

This study has several limitations.

Representativeness and scope of data sources. The combination of law enforcement files, one incident response firm's reports, and negotiation transcripts from leak sites introduces selection biases. Police case files tend to over-represent larger or regulated organisations, and under-represent victims who knowingly violated legal obligations during the incident (e.g. paying sanctioned entities [63]). Incident response firm data reflects the characteristics of its own client base, likely well-resourced victims who can afford to pay for professional help. Leak site transcripts capture only high-profile ransomware groups that choose to publish victim data [45]. As a result, the findings are most applicable to well-resourced victims and prominent ransomware operations.

The dataset covers a specific time period (2019—2023) and is weighted toward European and North American jurisdictions. Changes in regulation or different geopolitical contexts could alter bargaining patterns. It is also possible that ransomware business models will evolve again, after all this paper's focus on negotiation only became relevant since 2016 when ransomware actors moved away from fixed price demands [57]. Another such change would reduce the applicability of our results.

Data Access. Only one researcher had full access to sensitive police and incident response reports, and those results cannot be shared externally. This limitation undermines the reproduction of these results, a key part of open science. However, widely sharing victim data undermines privacy and risks facilitating re-victimization if the data was leaked. This kind of highly limited, data sharing arrangement may be the only way to conduct this kind of study.

Omitted variables and causality. Important factors such as attacker reputation, decryption credibility, and industry-specific ransom norms could not be systematically measured. These may correlate with over-represented attributes in the dataset, such as company size. While multiple controls were included to reduce the risk of confounding variables, this risk cannot be eliminated. The six-stage model assumes a sequential flow, yet the causal direction might be reversed: some victims might expect a discount for certain ransomware groups, and therefore be willing to negotiate. As an observational analysis, the model identifies stage-specific associations but does not confirm causal relationships.

Missing psycho-social dimension of negotiation chats. We did not analyze the content of ransomware negotiations chats. We therefore missed insights into persuasion tactics, cultural bargaining strategies, and other factors rooted in psychology or sociology. Omitting these factors may add noise, but it is unlikely to reverse the direction of significant effects. For example, negotiation duration and discount size could be interpreted as proxies for broader psychological and sociological factors.

6.3 Recommendations

Our results have several implications for policy and practice.

Victims. In addition to investing in Stage 0 preventative controls, organizations must maintain functioning backups via regular testing and backup cadence. More detailed and consistent guidance is required [56]. This must be guided by evidence and not conventional wisdom, which is contradicted evidence that offsite backups had the highest failure rate [32]. Academics should conduct user studies on how backup maintenance.

In terms of who to call in a crisis, our findings suggest IR firms help with securing a discount, but not the size of the discount. More generally, the skill of negotiators is questionable given ransom payments have soared by over 10,000% since 2018 [16, 17]. Instead IR firms likely provide value in helping victims manage uncertainty, and enforcing the reputation system that means post-payment complications are relatively rare (see Table 6).

Government Policy & Insurance. Our six-stage model argues we should move away from the understanding of ransomware as question of pay vs not. This binary framing motivates policy options like banning payment [4], a nuclear option that imposes a ruinous cost on unprepared victims who must rebuild from scratch. It may also increase data privacy violations, as criminals seek alternative monetization strategies. Instead of a binary, policymakers could use a policy lever like a ransom tax or a ransom cap [36] to gradually tilt the calculus away from paying ransoms. For example, a cap on ransom payments improves the negotiation discipline of individual victims who can argue they cannot pay above the cap by law. Such policies exploit attackers' willingness to negotiate (Stage 2–4), thereby reducing the size of payments ensures less funds flow to criminals. Crucially, the gradual approach preserves optionality for unprepared victims.

Insurers can introduce a similar policy, but within each contract. Cyber policies could make higher limits available for business interruption than is available for extortion payments. Alternatively, insurers could make extortion payments subject to co-insurance, so the victim faces a greater cost for paying. This tilts victims away from paying ransoms, without imposing ruin on victims who simply need to pay.

Data Collection & Research. Governments and industry should strengthen systems for collecting and sharing structured ransomware data between law enforcement and private partners. Improved monitoring would make it possible to detect shifts in offender tactics early and adjust interventions before re-extortion becomes a widespread source of attacker income. A specific recommendation is for IR firms to collect more fine-grained information on why backups failed.

7 Conclusion

This paper introduces and advances a six-stage model of ransomware negotiation and validates it using both incident-level data and public negotiation transcripts. Our findings show that outcomes are shaped by sequential filtering rather than a one-off payment decision. Each stage (victim engagement, discount offering, discount size, and final payment) narrows the set of incidents and alters the bargaining position of both sides.

Across both structured models and public negotiation transcripts, the strongest predictors of payment are preparedness and initial pricing, not discount size. Functional backups sharply reduce the likelihood of payment, while insurance primarily operates indirectly by enabling incident response involvement, which in turn increases victims willingness to negotiate. The message exchanges during negotiation increase the probability of discount offers, yet the size of those discounts does not, on its own, raise payment probability.

Our findings also clarify three common misconceptions. Insurance does not mechanically drive ransom payment but influences it through incident handling; discounts are better understood as persuasion tools emerging in longer negotiations, not simple price reductions; and systematic re-extortion is rare, likely constrained by attackers preserving reputation.

Taken together, these findings highlight the value of viewing ransomware as a staged bargaining process, providing a stronger basis for both future research and evidence-based interventions for victims and policy-makers alike.

References

[1] Lawrence Abrams. Conti ransomware prioritizes revenue and cyberinsurance data theft. In *https://www.bleepingcomputer.com/news/security/conti-ransomware-prioritizes-revenue-and-cyberinsurance-data-theft/*, 2021.

[2] Sadia Afroz, Vaibhav Garg, Damon McCoy, and Rachel Greenstadt. Honor among thieves: A common's analysis of cybercrime economies. In *2013 APWG eCrime Researchers Summit*, pages 1–11. IEEE, 2013.

[3] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The menlo report. *IEEE Security & Privacy*, 10(2):71–75, 2012.

[4] Robert Booth. Uk government to ban public bodies from paying ransoms to hackers. In *The Guardian*, 2023.

[5] Simona Boticiu and Fabian Teichmann. How does one negotiate with ransomware attackers? *International Cybersecurity Law Review*, 5(1):55–65, 2024.

[6] Casey Breen, Cormac Herley, and Elissa M Redmiles. A large-scale measurement of cybercrime against individuals. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pages 1–41, 2022.

[7] Thomas Brzustowski, Alkis Georgiadis-Harris, and Balázs Szentes. Smart contracts and the coase conjecture. *American Economic Review*, 113(5):1334–1359, 2023.

[8] Jack Cable, Ian W Gray, and Damon McCoy. Showing the receipts: understanding the modern ransomware ecosystem. In *2024 APWG Symposium on Electronic Crime Research (eCrime)*, pages 149–161. IEEE, 2024.

[9] Anna Cartwright and Edward Cartwright. Ransomware and reputation. *Games*, 10(2):26, 2019.

[10] Anna Cartwright, Edward Cartwright, James MacColl, Gary Mott, Simon Turner, Julian Sullivan, and Jason RC Nurse. How cyber insurance influences the ransomware payment decision: theory and evidence. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 48(2):300–331, 2023.

[11] Anna Cartwright, Edward Cartwright, Lian Xue, and Julio Hernandez-Castro. An investigation of individual willingness to pay ransomware. *Journal of Financial Crime*, 30(3):728–741, 2023.

[12] Edward Cartwright, Julio Hernandez-Castro, and Anna Cartwright. To pay or not: game theoretic models of ransomware. *Journal of Cybersecurity*, 5(1):tyz009, 2019.

[13] Chainalysis. Ransomware payments exceed $1 billion in 2023, hitting record high after 2022 decline. In *https://www.chainalysis.com/blog/ransomware-2024/*, 2024.

[14] Coalition. 2025 cyber claims report. In *https://web.coalitioninc.com/download-2025-cyber-claims-report.html*, 2025.

[15] Angela Y. Connolly and Huw Borrion. Reducing ransomware crime: analysis of victims' payment decisions. *Computers & Security*, 119:102760, 2022.

[16] Coveware. Coveware's 2018 q4 ransomware marketplace report. In *https://www.coveware.com/blog/2019/1/21/covewares-2018-q4-ransomware-marketplace-report*, 2018.

[17] Coveware. Targeted social engineering is en vogue as ransom payment sizes increase. In *https://www.coveware.com/blog/2025/7/21/targeted-social-engineering-is-en-vogue-as-ransom-payment-sizes-increase*, 2025.

[18] Cybersecurity and Infrastructure Security Agency. #Stopransomware: Cl0p ransomware gang exploits cve-2023-34362 moveit vulnerability. In *https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-158a*, 2023.

[19] Peter R Darke and Darren W Dahl. Fairness and discounts: The subjective value of a bargain. *Journal of Consumer Psychology*, 13(3):328–338, 2003.

[20] Federal Trade Commission. Guides against deceptive pricing. In *https://www.ecfr.gov/current/title-16/chapter-I/subchapter-B/part-233*, 29167.

[21] Burak Filiz, Budi Arief, Orcun Cetin, and Julio Hernandez-Castro. On the effectiveness of ransomware decryption tools. *Computers & Security*, 111:102469, 2021.

[22] Jason Franklin, Adrian Perrig, Vern Paxson, and Stefan Savage. An inquiry into the nature and causes of the wealth of internet miscreants. volume 7, pages 375–388, 2007.

[23] Anjuli Franz, Verena Zimmermann, Gregor Albrecht, Katrin Hartwig, Christian Reuter, Alexander Benlian, and Joachim Vogt. Sok: Still plenty of phish in the sea—a taxonomy of user-oriented phishing interventions and avenues for future research. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 339–358, 2021.

[24] Florian Gassmann, Jonas Beck, Nicolas Gourmelon, and Zinaida Benenson. Valuation of confidentiality and availability in a personal ransomware attack scenario. *(unpublished)*, 2024.

[25] Ian W Gray, Jack Cable, Benjamin Brown, Vlad Cuiujuclu, and Damon McCoy. Money over morals: A business analysis of conti ransomware. In *2022 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–12. IEEE, 2022.

[26] Ece Gumusel, Yue Xiao, Yue Qin, Jiaxin Qin, and Xiaojing Liao. Understanding legal professionals' practices and expectations in data breach incident reporting. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 2711–2725, 2024.

[27] Mischa Hansel and Jantje Silomon. Ransomware as a threat to peace and security: understanding and avoiding political worst-case scenarios. *Journal of Cyber Policy*, 9(2):159–178, 2024.

[28] Danny Yuxing Huang, Maxwell Matthaios Aliapoulios, Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, Alex C Snoeren, and Damon McCoy. Tracking ransomware end-to-end. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 618–631. IEEE, 2018.

[29] Jian Huang, Jun Xu, Xinyu Xing, Peng Liu, and Moinuddin K Qureshi. Flashguard: Leveraging intrinsic flash properties to defend against encryption ransomware. In

Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pages 2231–2244, 2017.

[30] Cyentia Institute. 2025 information risk insights study. 2025. Available: https://www.cyentia.com/wp-content/uploads/2025/06/IRIS-2025.pdf.

[31] Amin Kharaz, Sajjad Arshad, Collin Mulliner, William Robertson, and Engin Kirda. UNVEIL: A large-scale automated approach to detecting ransomware. In *25th USENIX security symposium (USENIX Security 16)*, pages 757–772, 2016.

[32] Liat Klainman and Greg Otto. Backup breakdown: How data recovery impacts the outcome of cyberattacks. In *https://www.at-bay.com/dont-let-backups-break-business/*, 2023.

[33] Zoe Kleinman. Insurers defend covering ransomware payments. In *https://www.bbc.co.uk/news/technology-55811165*, 2021.

[34] Eugene Kolodenker, William Koch, Gianluca Stringhini, and Manuel Egele. Paybreak: Defense against cryptographic ransomware. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 599–611, 2017.

[35] Zhen Li and Qi Liao. Preventive portfolio against data-selling ransomware—a game theory of encryption and deception. *Computers & Security*, 116:102644, 2022.

[36] Kyle D Logue and Adam B Shniderman. The case for banning (and mandating) ransomware insurance. *Conn. Ins. LJ*, 28:247, 2021.

[37] Jonathan Lusthaus. Trust in the world of cybercrime. *Global crime*, 13(2):71–94, 2012.

[38] Jamie MacColl, Pia Hüsch, Gareth Mott, James Sullivan, Jason RC Nurse, Sarah Turner, and Nandita Pattnaik. The scourge of ransomware: Victim insights on harms to individuals, organisations and society. RUSI Occasional Papers, 2024.

[39] Tina Marjanov and Alice Hutchings. Sok: Digging into the digital underworld of stolen data markets. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 1–18. IEEE, 2025.

[40] Sifra R. Matthijsse, Alba Moneva, Marloes S. van 't Hoff-de Goede, and Rutger E. Leukfeldt. Examining ransomware payment decision-making among small- and medium-sized enterprises. *European Journal of Criminology*, 22(4):625–645, 2024. Original work published 2025.

[41] Timothy Mcintosh, Teo Susnjak, Tong Liu, Dan Xu, Paul Watters, Dongwei Liu, Yaqi Hao, Alex Ng, and Malka Halgamuge. Ransomware reloaded: Re-examining its trend, research and mitigation in the era of data exfiltration. *ACM Computing Surveys*, 57(1):1–40, 2024.

[42] Tom Meurs, Edward Cartwright, and Anna Cartwright. Double-sided information asymmetry in double extortion ransomware. In *International Conference on Decision and Game Theory for Security*, pages 311–328. Springer, 2023.

[43] Tom Meurs, Edward Cartwright, Anna Cartwright, Marianne Junger, Robert Hoheisel, Erik Tews, and Abhishta Abhishta. Ransomware economics: A two-step approach to model ransom paid. In *18th Symposium on Electronic Crime Research (eCrime)*, 2023.

[44] Tom Meurs, Robert Hoheisel, Marianne Junger, Abhishta Abhishta, and Damon McCoy. What to do against ransomware? evaluating law enforcement interventions. In *2024 APWG Symposium on Electronic Crime Research (eCrime)*, pages 76–93. IEEE, 2024.

[45] Tom Meurs, Marianne Junger, Maarten Cruyff, and Peter GM Van Der Heijden. Estimating the number of ransomware attacks. *Journal of Quantitative Criminology*, pages 1–17, 2025.

[46] Tom Meurs, Marianne Junger, Erik Tews, and Abhishta Abhishta. 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)*, pages 1–13. IEEE, 2022.

[47] Tyler Moore and Richard Clayton. An empirical analysis of the current state of phishing attack and defence. In *Workshop on the Economics of Information Security*, 2007.

[48] Gareth Mott, Sarah Turner, Jason RC Nurse, Jamie MacColl, James Sullivan, Anna Cartwright, and Edward Cartwright. Between a rock and a hard (ening) place: Cyber insurance in the ransomware era. *Computers & Security*, 128:103162, 2023.

[49] Bernold Nieuwesteeg and Michael Faure. The uneasy case for a ransom tax. *European Journal of Risk Regulation*, 14(2):382–398, 2023.

[50] Office of Fair Trading. Pricing practices guide: Helping you comply with consumer protection law. In *https://assets.publishing.service.gov.uk/media/5a80c18e40 f0b62302695536/10-1312-pricing-practices-guidance-for-traders.pdf*, 2010.

[51] Kris Oosthoek, Jack Cable, and Georgios Smaragdakis. A tale of two markets: Investigating the ransomware payments economy. *Communications of the ACM*, 66(8):74–83, 2023.

[52] Harun Oz, Ahmet Aris, Abbas Acar, Güliz Seray Tuncay, Leonardo Babun, and Selcuk Uluagac. {RØB}: Ransomware over modern web browsers. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 7073–7090, 2023.

[53] Masarah Paquet-Clouston, Bernhard Haslhofer, and Benoit Dupont. Ransomware payments in the bitcoin ecosystem. *Journal of Cybersecurity*, 5(1):tyz003, 2019.

[54] Alex Rebert, Chandler Carruth, Jen Engel, and Andy Qin. Safer with google: Advancing memory safety. In *https://security.googleblog.com/2024/10/safer-with-google-advancing-memory.html*, 2024.

[55] Sasha Romanosky, Lillian Ablon, Andreas Kuehn, and Therese Jones. Content analysis of cyber insurance policies: how do carriers price cyber risk? *Journal of Cybersecurity*, 5(1):tyz002, 2019.

[56] Kimberly Ruth, Raymond Buernor Obu, Ifeoluwa Shode, Gavin Li, Carrie Gates, Grant Ho, and Zakir Durumeric. A first look at governments' enterprise security guidance. In *34th USENIX Security Symposium (USENIX Security 23)*, 2025.

[57] John Sakellariadis. Behind the rise of ransomware. In *Atlantic Council Cyber Statecraft Initiative*, 2022.

[58] Nolen Scaife, Henry Carter, Patrick Traynor, and Kevin RB Butler. Cryptolock (and drop it): stopping ransomware attacks on user data. In *2016 IEEE 36th international conference on distributed computing systems (ICDCS)*, pages 303–312. IEEE, 2016.

[59] Rhythima Shinde, Pieter Van der Veeken, Stijn Van Schooten, and Jan van den Berg. Ransomware: Studying transfer and mitigation. In *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, pages 90–95. IEEE, 2016.

[60] Max Smeets. *Ransom War: How Cyber Crime Became a Threat to National Security*. Oxford University Press, 2025.

[61] Sophos. The sophos annual threat report: Cybercrime on main street 2025. In *https://news.sophos.com/en-us/2025/04/16/the-sophos-annual-threat-report-cybercrime-on-main-street-2025/*, 2025.

[62] Richard Thaler. Mental accounting and consumer choice. *Marketing science*, 4(3):199–214, 1985.

[63] Max van der Horst, Ricky Kho, Olga Gadyatskaya, Michel Mollema, Michel Van Eeten, and Yury Zhauniarovich. High stakes, low certainty: Evaluating the efficacy of high-

level indicators of compromise in ransomware attribution. In *Proceedings of the 34th USENIX Security Symposium (USENIX Sec)*, 2025.

[64] R. Van Wegberg and T. Verburgh. Lost in the dream? measuring the effects of operation bayonet on vendors migrating to dream market. In *Proceedings of the Evolution of the Darknet Workshop*, volume 9, 2018.

[65] Verizon. 2025 data breach investigations report. In *https://www.verizon.com/business/resources/Tea/reports/2025-dbir-data-breach-investigations-report.pdf*, 2025.

[66] Shenao Wang, Feng Dong, Hangfeng Yang, Jingheng Xu, and Haoyu Wang. Cancal: Towards real-time and lightweight ransomware detection and response in industrial environments. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 2326–2340, 2024.

[67] Daniel W Woods. A turning point for cyber insurance. *Communications of the ACM*, 66(3):41–44, 2023.

[68] Daniel W Woods, Rainer Böhme, Josephine Wolff, and Daniel Schwarcz. Lessons lost: Incident response in the age of cyber insurance and breach attorneys. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 2259–2273, 2023.

[69] Daniel W Woods and Lukas Walter. Reviewing estimates of cybercrime victimisation and cyber risk likelihood. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 150–162. IEEE, 2022.

[70] Tongxin Yin, Armin Sarabi, and Mingyan Liu. Deterrence, backup, or insurance: game-theoretic modeling of ransomware. *Games*, 14(2):20, 2023.

[71] Adam Young and Moti Yung. Cryptovirology: Extortion-based security threats and

countermeasures. In *Proceedings 1996 IEEE Symposium on Security and Privacy*, pages 129–140. IEEE, 1996.

[72] Chen Zhang and Feng Luo. Bargaining game theoretical analysis framework for ransomware attacks. *Journal of Information Security and Applications*, 93:104115, 2025.

[73] Huan Zhang, Lixin Zhao, Aimin Yu, Lijun Cai, and Dan Meng. Ranker: Early ransomware detection through kernel-level behavioral analysis. *IEEE Transactions on Information Forensics and Security*, 19:6113–6127, 2024.

[74] Leah Zhang-Kennedy, Hala Assal, Jessica Rocheleau, Reham Mohamed, Khadija Baig, and Sonia Chiasson. The aftermath of a crypto-ransomware attack at a large academic institution. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 1061–1078, 2018.

[75] Lingbo Zhao, Yuhui Zhang, Zhilu Wang, Fengkai Yuan, and Rui Hou. Erw-radar: An adaptive detection system against evasive ransomware by contextual behavior detection and fine-grained content analysis. In *NDSS*, 2025.

Appendix

Table A1 can be found on the next page.

Table A1: Regression Results (estimated β and p-value)

Predictor	M1: Attacker Negotiation	M2: Victim Negotiation	M3: Discount Given	M4: Discount Size	M5: Payment
1a. Revenue (log)	0.04 (0.624)	0.30 (0.048)*	0.39 (0.116)	−0.12 (0.001)**	0.44 (0.085)$^{\cdot}$
1b. Ransomware-as-a-Service (RaaS)	0.31 (0.207)	−0.17 (0.654)	0.75 (0.159)	−0.17 (0.005)**	−0.19 (0.723)
1c. Backups: Unrecoverable (vs. no backups)	0.86 (0.015)*	0.10 (0.831)	0.36 (0.648)	−0.00 (0.986)	0.61 (0.353)
1d. Backups: Partial (vs. no backups)	0.91 (0.010)*	−0.82 (0.127)	0.14 (0.875)	−0.01 (0.960)	−0.84 (0.246)
1e. Backups: Recoverable (vs. no backups)	0.72 (0.029)*	−1.60 (0.007)**	−2.15 (0.033)*	0.20 (0.151)	−2.47 (0.002)**
1f. Data exfiltration (yes/no)	0.66 (0.044)*	0.77 (0.102)	0.89 (0.142)	−0.16 (0.026)*	0.10 (0.878)
2a. Cyber insurance (yes/no)	—	0.38 (0.472)	−0.09 (0.885)	−0.11 (0.039)*	0.14 (0.832)
2b. IR firm involvement (yes/no)	—	3.47 (< .001)***	2.04 (0.004)**	−0.04 (0.696)	0.06 (0.945)
3. Negotiation duration (log)	—	—	1.22 (< .001)***	0.08 (0.061)$^{\cdot}$	−1.63 (0.012)*
4a. Initial ransom demand (log)	—	—	—	0.50 (< .001)***	−0.25 (0.068)$^{\cdot}$
4b. Negotiation (attacker)	—	—	—	−0.07 (0.537)	0.18 (0.775)
4c. Negotiation (victim)	—	—	—	0.08 (0.679)	4.80 (0.001)**
5a. Discount offered (yes/no)	—	—	—	—	1.41 (0.039)*
5b. Discount size (log)	—	—	—	—	−0.30 (0.343)
McFadden's pseudo-R^2	0.2438	0.0549	0.2712	-	0.3369
Adjusted R^2	-	-	-	0.3028	-