You Can't Talk Your Way Out of Cyber Extortion: An Empirical Study of Ransomware Negotiations

Tom Meurs
tom.meurs@politie.nl
Dutch Police
Apeldoorn, Netherlands

Anna Cartwright
a.cartwright11@icloud.com
Independent Researcher
Kent, United Kingdom

Edward Cartwright
edward.cartwright@dmu.ac.uk
De Montfort University
Leicester, United Kingdom

Harold Houba
harold.houba@vu.nl
Vrije Universiteit Amsterdam
Amsterdam, Netherlands

Daniel W Woods
daniel.woods@ed.ac.uk
University of Edinburgh
Edinburgh, United Kingdom

Abstract

Ransomware has become one of the most common cyber risks for businesses. Average ransom payments have increased sharply since 2018, despite attackers still exploiting longstanding technical issues. To better understand why attackers can extract increasingly large ransoms from victims, we study various explanations for ransom outcomes: technical (back up availability and data security); economic (insurance and external consultants); and negotiation strategy. We model ransomware negotiation using a six-stage framework synthesized from prior literature, capturing the attacker's approach to negotiation, discounts, and re-extortion, as well as the victim's engagement and payment decisions. We evaluate these stages using two datasets: (1) 461 ransomware incidents reported to the police (2019–2023); and (2) metadata from 237 negotiation transcripts that were published online. Across both datasets, payment outcomes are most strongly associated with victims' early-stage conditions rather than attackers' late-stage pricing strategies. Fully recoverable backups sharply reduce engagement and payment, while higher victim revenue increases the likelihood of negotiation and discounting. Longer negotiations increase the likelihood of discounts, but are associated with lower payment rates. We find no evidence that cyber insurance directly increases payment or that discount size independently affects payment once interaction is accounted for. Our results suggest that improving victims' early-stage conditions, primarily via backup reliability, offers greater leverage for reducing ransomware prevalence and profitability than focusing on negotiation tactics. Our main recommendation is that practitioners and researchers focus on the organizational processes surrounding backups, an overlooked topic within academic computer security.

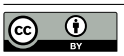
CCS Concepts

• **Security and privacy → Economics of security and privacy**;

Keywords

Ransomware, Cyber Extortion, Security Economics, Empirical Cybercrime Measurement

ACM Reference Format:
Tom Meurs, Anna Cartwright, Edward Cartwright, Harold Houba, and Daniel W Woods. 2026. You Can't Talk Your Way Out of Cyber Extortion: An Empirical Study of Ransomware Negotiations. In *Proceedings of the 2026 ACM SIGSAC Conference on Computer and Communications Security (CCS '26), November 15–19, 2026, The Hague, Netherlands.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3830454.3832611



CCS '26, The Hague, Netherlands

ACM ISBN 979-8-4007-2871-6/2026/11
https://doi.org/10.1145/3830454.3832611

1 Introduction

Ransomware has evolved from automated software that demands a fixed ransom payment into a sophisticated criminal enterprise in which attackers engage in selective targeting, dynamic pricing, and strategic negotiation [7, 54, 58, 81]. The change in business model coincided with a sharp increase in the frequency and impact of ransomware incidents [65]. Ransomware comprised 40% of publicly reported incidents impacting businesses in 2024, up from around 1% in 2016 [21]. In terms of impact, one professional negotiator reports the average ransom payment rose from $6k in Q3 2018 [18] to $1.13m in Q2 2025 [19]. These rapid changes destablized the cyber insurance market [56, 76] and led politicians to declare national emergencies related to specific ransomware incidents [30].

The escalating negative impact of ransomware cannot be explained by ransomware actors using new technical breakthroughs. Industry reports find the majority of incidents begin with stolen credentials [19, 65, 69, 73], obtained via well-established cybercrime techniques like phishing [26, 55] or cybercrime marketplaces [25, 46]. When ransomware groups exploit software vulnerabilities, they typically rely on longstanding issues like memory safety [20, 61]. If technical innovation alone is insufficient to understand the continued rise of ransomware then an alternative lens is needed. We suggest behavioral and economic explanations.

Ransomware has proved highly lucrative for criminal actors, in part, because they can extract high ransom payments from victims through negotiation. Prior work and industry commentary propose several mechanisms that may shape negotiation outcomes. Some argue that paying ransoms may increase the risk of re-extortion or incentivize future attacks [6, 48]. Incident response (IR) firms suggest that inexperienced negotiation on the part of victims can

lead to higher payments. Others emphasize the role of backup reliability, although this is undermined by data exfiltration threats [48]. Finally, the insurance industry has been accused of encouraging ransom payments by absorbing part of the financial risk [42].

Despite their prominence, the many channels through which negotiation can influence ransomware outcomes have received limited empirical evaluation. Ransom victims are a hard-to-reach population because disclosing details of the attack may incur legal and reputation risk. As a result, ransom negotiation is an understudied (extreme) experience for network operators, a user group that "have not received significant attention by prior research" [22].

To help victims achieve better negotiation outcomes, we ask:
RQ1: Which factors influence ransomware negotiations?
RQ2: How does negotiation strategy impact discounts?
RQ3: What problems do victims face post-payment?

To address **RQ1**, we model 461 police-reported ransomware incidents (2019–2023) as a six-stage negotiation process. To address **RQ2**, we analyze metadata from 237 negotiation transcripts across 23 ransomware groups. To address **RQ3**, we conduct a qualitative case study of reinfection and post-payment outcomes among the police-reported ransomware incidents.

Contributions. We find the following:

- **RQ1** Across 461 police-reported incidents, payment rates are higher when attackers start with high ransom demands and negotiate, even though these cases involve higher final prices than low fixed-price demands. Stage-level analysis shows that this reflects a selection effect: only a subset of high-demand cases, typically involving higher-revenue victims without recoverable backups, progress to later stages where payment becomes likely. In contrast to claims from IR companies, these results indicate that payments are not driven by pricing, negotiation tactics or insurance, but by whether victims lack technical recovery options.
- **RQ2** Using metadata from 237 negotiation transcripts across 23 ransomware groups, we find that longer negotiations are associated with receiving discounts but are also associated with lower payment rates. Incident response (IR) firm involvement similarly increases the likelihood of discounts but is not associated with a higher probability of payment. These results challenge the assumption that more intensive or structured negotiation reliably improves outcomes, and suggest that negotiation strategy primarily affects how bargaining unfolds rather than whether payment occurs.
- **RQ3** Across the 461 police-reported incidents, reinfection attempts occur in 3.6% of all cases, while among paying victims, 2.5% never received a decryption key, 6.7% experienced partial or delayed decryption, and post-payment renegotiation occurred in 7.6% of cases. These findings refine claims that post-payment risks are sufficiently severe to consistently deter victims from paying.

Section 2 provides background on ransomware negotiation and synthesizes prior work into a six-stage framework. Section 3 describes the data sources and hypotheses. Section 4 presents the empirical analyses. Section 5 discusses the findings and limitations. Section 6 concludes.

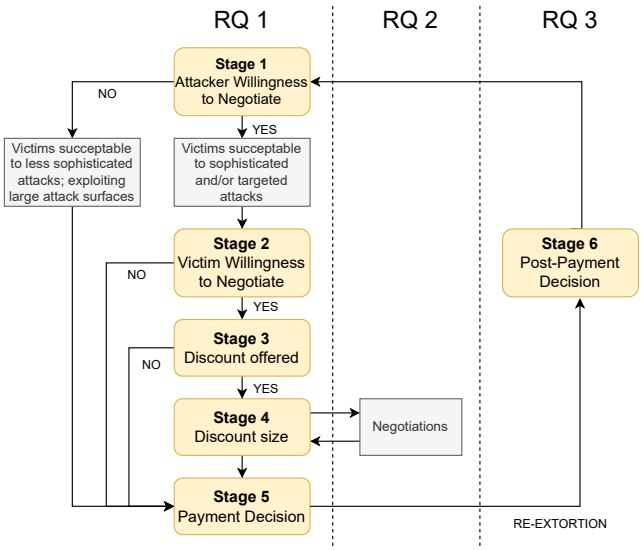


Figure 1: Six-stage framework of ransomware negotiation. The diagram shows the sequence of decision points from attacker willingness to negotiate to post-payment interaction, including possible exits at each stage.

2 Background

Prevalence Ransomware remains rare among individuals, with annualized frequency estimated at 0.1% in a nationally representative survey [8]. Annualized victimization rates among businesses range from around 1% to 60% depending on the study [78]. 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% [53]. In terms of impact, the average ransomware incident cost is estimated to cost over $1 million in a study of public incidents [21]. The mean size mean value of cyber insurance claims resulting from ransomware was $292k in 2024 [16]. These quantitative estimates do not account for impacts like emotional stress and support networks disruptions [44, 84].

The Nature of Ransomware Ransomware differs from most computer security incidents because the attacker actively facilitates detection by displaying a ransom note and communicating with the victim. By contrast, threat actors usually attempt to remain undetected [3, 38]. Indeed, the NIST-800-61 "Computer Security Incident Handling Guide" [15] devotes more subsections to detection than any other stage of incident response.

Ransomware actors want to interact with victims in order to exploit business interruption, data exfiltration, and time pressure. This turns recovery into a negotiated process rather than a purely technical response [12]. The novelty of negotiating with cyber criminals can be seen in NIST-800-61 not mentioning "negotiation" once [15]. As a result, the outcome of a ransomware incident depends not only on the technical details of compromise, but also on a sequence of behavioral decisions made by both attacker and victim.

To model these sequential decisions, Figure 1 presents a six-stage framework based on insights from game theory and industrial economics. In a game theoretic sense the attacker must ultimately

decide: the size of initial ransom demand, whether they want to engage in negotiation with the victim, the maximum discount they are willing to offer (if any) in the case of negotiation, and actions post-payment. The victim must decide: the maximum ransom they are willing to pay and whether they are willing to negotiate with the attackers. The six stages in our framework represent an irreversible ordering of decisions that encapsulate the strategic choices of attacker and victim, transitioning from signaling a willingness to negotiate (stages 1-2), discounts and negotiation (stages 3 and 4) to payment and post-payment decisions (stages 5 and 6).

Crucially, the attacker and victim have incentives to implement a forward looking strategy whereby, for instance, the attacker's decision of whether to show willingness to negotiate in stage 1 will depend on their expectation of outcomes in stages 2-5. Such forward looking expectations create a conditional dependence that can only be studied by modeling all six stages together. In particular, stage 4, and negotiations, are conditional on attacker and victim choosing to engage in those negotiations and not take alternative paths at stages 1-3. Thus, attacker and victim only reach stage 4 if they both believe they have something to gain from the negotiation path.

While some stages have been examined individually, such as payment decisions [9, 12, 13, 47, 50], decryptor reliability [24, 48], and backups [51], no prior work empirically models progression through all stages (see Table 1). Isolated analyses in prior work cannot account for conditional dependencies between stages.

Stage 0. Initial compromise. Before negotiation can occur, ransomware actors must gain access to a system and deploy malware and/or exfiltrate data. Computer security research into ransomware has primarily focused on this stage, including novel attack techniques [59, 80], detection heuristics [35, 74, 83, 85], and defensive measures [32, 39, 66]. These questions of how to attack, detect, and prevent ransomware are familiar to computer scientists, but do not address the defining feature of modern ransomware, namely active negotiation with the victim. Stage 0 is outside the scope of the current study and is included here for completeness.

Stage 1. Attacker willingness to negotiate. Until around 2016, ransomware actors largely relied on fixed ransom demands [64], meaning that interactive negotiation was largely absent. A 2016 study found that only a small fraction of victims paid ransoms, in part due to distrust and limited ability to facilitate cryptocurrency payments [68]. In such cases, victims faced a take-it-or-leave-it demand with no opportunity for interaction. Subsequent work suggests that ransomware actors adapted by allowing negotiation, as attacker–defender communication enabled initial demands to be adjusted based on victim responses [54]. As interactive negotiation became more common after 2016, prior work began to examine individual aspects of the negotiation process.

In Stage 1, attackers must either issue a fixed ransom demand or permit negotiation by providing contact details, such as a chat link or email address. This decision determines whether the incident can progress into an interactive exchange. In many cases, the size of the ransom demand is only revealed after the victim makes contact, which allows attackers to assess responsiveness, operational impact, and apparent ability to pay before committing to a price. Negotiation is costly for attackers in terms of time and resource and so it is optimal to only negotiate where the expected gain in ransom revenue exceeds the cost. Attackers may, thus, be less likely to engage in negotiation when expected returns are low, for example when victims appear unable to pay or to complete cryptocurrency transactions [53, 68].

Stage 2. Victim willingness to negotiate. At Stage 2, victims presented with contact details decide whether to initiate communication with the attacker [17, 51]. Some victims engage immediately, while others remain silent. This decision reflects an assessment of expected losses, recovery prospects, and the potential informational value of dialogue. For example, dialogue is more likely when backups are unavailable or exfiltrated data is sensitive. Prior empirical work has merged this stage into later payment outcomes rather than modeling negotiation initiation itself (see Table 1).

Stage 3. Discount offering. If communication begins, Stage 3 captures whether attackers reduce the initial ransom demand during the exchange. Industry reports and anecdotal evidence indicate that discounts are common in ransomware negotiations, often reducing demands by 50–90% [7]. Attackers may infer victim willingness to pay from message tone, urgency, or perceived payment feasibility. Threats of rapid escalation, such as warnings that failure to pay within 48 hours will double the ransom, are frequently reported, but experienced negotiators know this is rarely enforced [50].

Stage 4. Discount size. This stage captures how large a discount is offered once attackers decide to reduce the ransom demand. It is distinct from the prior stage because the final demand may result from repeated back-and-forth communication between attacker and victim, sometimes spanning dozens of messages. While discounting appears flexible in practice, some ransomware-as-a-service (RaaS) groups have introduced internal rules that restrict discount size in an effort to maintain pricing discipline across affiliates. For example, an industry analysis of LockBit documents formal negotiation rules that prohibit discounts exceeding 50% of the initial ransom demand, motivated by concerns that inconsistent discounting undermines future negotiations [67]. Related academic work proposes theoretical bargaining models to study how negotiation may influence ransom demands [81], but does not empirically validate negotiations across stages.

Stage 5. Payment decision. Stage 5 captures whether a victim ultimately pays the ransom after negotiation. This stage has received the most attention in prior work. One line of research quantifies ransom payments at an aggregate level, primarily through blockchain analysis of cryptocurrency flows [9, 28, 31, 60], with complementary estimates provided by industry reports [14]. Such studies only sample incidents where the victim paid. Other studies model payment as a binary outcome, focusing on whether victims pay or not pay [13, 50].

Non-quantitative studies have emphasized qualitative factors like the role of cyber insurance [77], data breach notification obligations [29], sanctions compliance [71]. Game-theoretic models of ransom payment focus on the structure of incentives without studying real-world outcomes [13, 79]. Another strand of research analyzes payment willingness in hypothetical situations involving individuals and organizations [12, 17, 27, 47, 50]. Across these approaches, payment is typically treated as a standalone decision, without conditioning on earlier negotiation stages.

Table 1: Ransomware studies mapped to our negotiation framework. Ticks indicate explicit analysis of a stage.

				Investigated Stages					
Article	**Year**	**Data Type**	**n**	**1**	**2**	**3**	**4**	**5**	**6**
[40]	2017	Theory	-					✓	
[13]	2019	Theory	-					✓	
[24]	2021	Decryptors	78						✓
[17]	2022	Cases	41		✓			✓	
[41]	2022	Theory	-					✓	
[54]	2022	Cases	453	✓				✓	
[82]	2022	Theory	-					✓	✓
[50]	2023	Cases	382					✓	
[12]	2023	Survey	1,798					✓	
[11]	2023	Theory	-					✓	
[79]	2023	Theory	-					✓	✓
[27]	2024	Survey	800					✓	
[47]	2024	Survey	445					✓	
[81]	2025	Theory	-			✓	✓	✓	
[75]	2025	Survey	51						✓
This work	2026	Cases	698	✓	✓	✓	✓	✓	✓

Stage 6. Post-payment interaction. The final stage captures interactions after payment, including decryption failure and re-extortion. Although frequently discussed in policy and industry reports [48], empirical estimates of post-payment outcomes are rare. One laboratory study found that roughly half of decryptors failed under experimental conditions [24], but it remains unclear how often such failures occur in real incidents. To our knowledge, no prior work estimates the prevalence of re-extortion or other post-payment outcomes in real incidents (**RQ3**).

Summary Prior work focused on isolated stages of ransomware negotiation, but rarely the full process. Table 1 summarizes this literature by mapping studies to the negotiation stages they examine. Most empirical work concentrates on the payment decision, while earlier negotiation stages and post-payment interaction receive limited attention. This gap motivates a stage-based empirical analysis that models negotiation as an end-to-end process.

3 Methodology

Given prior work has identified which variables are of interest, our RQs aim to quantify how those variables influence each stage of ransomware negotiations. This quantitative design requires collecting many observations of ransomware incidents. Data collection is challenging because negotiations are highly sensitive and occur in private channels, in addition to existing challenges recruiting security experts to user studies [33, 34]. These considerations are why the first author was the only researcher authorized to access and code the data, which meant intercoder reliability was not possible. However, the sole coder could at least use six years of experience working ransomware cases at the Dutch Police.

3.1 Data Collection and Labeling

We collect two historical samples to study ransomware negotiation as a staged process within the six-stage model. Incidents reported to the Police and incident response (IR) firms provide insights across all stages (Stages 1–6), enabling analysis of negotiation initiation, progression, discounting, payment, and post-payment outcomes. To gain deeper insights into negotiations into the under-studied Stages 3 and 4 (see Table 1), we also collect negotiation transcripts that capture message volume, discounting, and payment outcomes. Using multiple sources helps mitigate reporting biases associated with each source individually.

3.1.1 Dataset 1: Police and Incident Response Data. Our primary dataset combines 525 ransomware incidents reported to the Dutch police between January 1, 2019 and January 1, 2023 with 116 ransomware incidents handled by a private incident response (IR) firm between February 20, 2020 and January 1, 2023. Both sources were originally collected for operational purposes rather than research. Use of police data for scientific research is governed by Article 22 of the Dutch Police Data Act. Permission was granted under the condition that all results are published in fully anonymized form. This created a constraint that just one member of the research team could access raw case details in order to identify and label ransomware incidents.

Data extraction and labeling. Police cases were identified by querying internal Dutch police information systems for reports containing the keyword ransomware, followed by manual review to confirm. This revealed reports and investigation records for each case, which enabled case-level assessment of post-payment outcomes (**RQ3**), including decryptor failure, renegotiation, and reinfection. To identify delayed reinfection, we cross-matched victim identifiers across the merged police and IR dataset. Incident response (IR) cases were provided directly to the lead author by a private IR firm under a non-disclosure agreement. Due to this agreement, only the lead author had access to identifiable case details, and no other authors could inspect the raw IR reports.

We only included cases if reports explicitly described both file encryption and a corresponding ransom demand. Reports referring solely to attempted intrusions, generic malware, or non-encrypting extortion were excluded during manual validation. Duplicate records or incidents were identified and merged using victim identifiers, timestamps, and contextual descriptions. We extracted variables using explicit decision rules based on observable evidence in the reports. For example, payment was recorded only when a completed transaction was explicitly confirmed; similarly the offer of a discount required a documented reduction between initial and final ransom demands; and post-payment outcomes were coded only when explicitly described in the case record. Ambiguous or incomplete information was conservatively coded as missing rather than inferred. The full codebook, including variable definitions and decision criteria, is provided in Appendix A. The final dataset consists of 461 unique ransomware attacks against organizations, after excluding attempted intrusions, incidents targeting individuals, and cases without encryption or a ransom demand.

Table 2 shows the list of variables. Only one researcher had access to raw case details, which means inter-rater reliability tests were not feasible. To limit subjectivity, we restricted variables to factual, verifiable events rather than interpretive judgments.

Measurement validity and scope. Measurement validity is addressed by focusing on observable actions and decisions recorded

during incident handling, by explicitly accounting for stage-dependent and non-random missingness, and by clearly defining the scope and generalizability of the results. Most structural variables are observed for the majority of incidents, with relatively low missingness for, for example, backup status (10.0%), victim's yearly revenue (2.4%), and payment (7.2%). Ransom demand (49.5% missing) and discount size (53.1% missing) are observed only for incidents that progress to negotiation, resulting in higher missingness by design. Accordingly, stage-wise pricing and discount analyses are estimated on stage-appropriate complete cases (Figure 3), while analyses of negotiation progression and payment outcomes use the full set of 461 incidents (**RQ1** and **RQ3**). No imputation was applied, as missingness primarily reflects the structure of the negotiation process rather than measurement error.

Regarding scope, the sample over-represents victims in the Netherlands. This limits direct generalization to other countries, as negotiation practices may differ across legal, institutional, and economic contexts. To partly address this limitation, we also analyze a second dataset (see the next subsection). A further sampling bias results from victims' reluctance to report cyber incidents to the police [70]. To mitigate this limitation, we combine police data with incident response (IR) firm data, which reduces reliance on police reporting alone. The combined dataset therefore provides a broad view of ransomware negotiation behavior within the Dutch context, while generalization beyond the Netherlands should be made with caution.

3.1.2 Dataset 2: Negotiation Transcripts. To analyze Stages 3–5 of the negotiation model, we collected a corpus of 237 ransomware negotiation transcripts from 23 ransomware groups extracted from *ransomware.live* on July 31, 2025. The dataset consists of publicly accessible attacker-victim chats scraped from leak sites and related infrastructure. Victims represented in the dataset are located worldwide. Because these transcripts only exist once communication has begun, they are used exclusively for analyses conditional on negotiation engagement (Stages 3–5).

Data extraction and coding. Negotiation transcripts were obtained via the public *ransomware.live* API. We used variables provided by the platform including message volume, initial and final ransom demand, and observed payment outcomes. Manual validation of a random 10% sample confirmed that variables are recorded only when explicitly present in the negotiation chats. We estimated discount size by comparing the initial and final demands provided by the API.

Measurement validity and scope. We cannot independently verify each case because transcripts are anonymized. To assess plausibility, we performed several consistency checks. First, all ransomware groups referenced in the transcripts were active during the corresponding time periods. Second, opening messages and negotiation templates closely match those observed in verified police and IR cases. Third, linguistic patterns and negotiation sequences are consistent across transcripts and align with known ransomware group practices. While these checks cannot fully eliminate fabrication risk, they make large-scale fabrication unlikely. Publicly posted transcripts may over-represent unsuccessful negotiations or disputes. Accordingly, this dataset is best suited for analyzing discounting behavior and negotiation dynamics conditional on engagement, rather than estimating overall payment rates.

3.2 Variable Selection

Variables were selected based on three criteria. First, we prioritize factors that recur in prior academic and industry work on ransomware negotiation, including organizational size, backup availability, and cyber insurance. Second, variables are aligned with specific stages of the six-stage negotiation framework, and are included only when they plausibly act as inputs to decisions at that stage. Third, all variables are restricted to information observable in structured police and incident response (IR) reports to ensure consistent measurement across incidents.

Table 2 summarizes variable definitions, measurement, stage alignment, and data sources. Some variables serve as outcomes at one stage and as inputs at later stages; for example, attacker willingness to negotiate (Stage 1 outcome) may proxy unobserved strategic intent that also shapes discounting behavior in subsequent stages. To avoid post-treatment bias, each variable enters the regression framework only from the stage at which it becomes observable to the negotiating parties. (see Table 2).

Organizational characteristics. We include victim revenue (log-transformed) as a proxy for organizational size and financial capacity. Backup availability is categorized as no backups installed or backups unrecoverable, partially recoverable, or fully recoverable. This coding captures either how successful attackers were in destroying backups, or reflect the degree of preparedness by victims.

We also include variables for whether the victim bought cyber insurance and whether they hired an IR firm. These are included to evaluate claims that they influence negotiation outcomes, especially the claim that insurance incentivizes ransomware [37]. These variables enter the model from Stage 2 onward, as they primarily affect victim behavior once communication with the attacker becomes possible.

Attack characteristics. We considered including a variable for each ransomware group, but this would have led to over-fitting given many ransomware groups were observed just once or twice. Instead, we included a binary indicator for ransomware-as-a-service (RaaS) affiliation to capture differences in attacker organization. To capture the details of the specific attacks, we include indicators for whether data was exfiltrated and the initial ransom demand (log-transformed).

Negotiation process and outcomes. Negotiation variables capture whether the attacker and victim engage in negotiation and the duration of interaction. Discount-related variables record whether a discount was offered and the magnitude of that discount. In case of back-and-forth negotiation discount is defined as the difference between initial and final ransom demand. Payment is modeled as the final outcome of the negotiation process.

3.3 Analysis

We estimate regression models aligned with the six-stage negotiation framework and the three research questions. Incident-level police and IR data are used to model progression across negotiation

Table 2: Variables, measurement, stage alignment, and missingness in the six-stage negotiation framework. Stages indicate the first stage at which a variable appears; some variables serve as outcomes at one stage and inputs at subsequent stages.

Role by stage	Variable	Description	Source	Missing (%)
Structural and attack characteristics (inputs from Stage 1 onward)				
Input (Stages 1–5)	**1a. Revenue (log)**	Log_{10} of estimated victim revenue.	Police / IR	2.4
Input (Stages 1–5)	**1b. RaaS affiliation**	Indicator for whether attack was conducted by a known ransomware-as-a-service group.	Police / IR	11.3
Input (Stages 1–5)	**1c–1e. Backup status**	Categorical variable to capture no backups vs un/partially/fully recoverable backups.	Police / IR	10.0
Input (Stages 1–5)	**1f. Data exfiltration**	Whether data exfiltration occurred as part of the ransomware attack (yes/no).	Police / IR	10.3
Stage 1: Attacker allows negotiation				
Outcome (Stage 1); Input (Stages 4–5)	**1g. Negotiation (attacker)**	Whether the attacker allowed negotiation beyond an initial demand (yes/no).	Police / IR	0.0
Stage 2: Victim initiates negotiation				
Input (Stages 2–5)	**2a. Cyber insurance**	Whether the victim held ransomware-relevant cyber insurance during the incident (yes/no).	Police / IR	10.8
Input (Stages 2–5)	**2b. IR firm involvement**	Whether a professional incident response firm assisted the victim (yes/no).	Police / IR	0.0
Outcome (Stage 2); Input (Stages 4–5)	**2c. Negotiation (victim)**	Whether the victim initiated or participated in negotiation (yes/no).	Police / IR	0.0
Stage 3: Discount offering				
Input (Stages 3–5)	**3a. Negotiation duration (log)**	Log_{10} of the total number of days with attacker–victim communication.	Police / IR	15.2
Input (Stages 3–5)	**3b. Initial ransom demand (log)**	Log_{10} of the initial ransom demand (EUR) stated by the attacker.	Police / IR; Transcripts	49.5
Outcome (Stage 3); Input (Stage 5)	**3c. Discount offered**	Whether the attacker offered a reduced ransom during negotiation (yes/no).	Police / IR; Transcripts	49.5
Stage 4: Discount size				
Outcome (Stage 4); Input (Stage 5)	**4. Discount size (log / proportion)**	Magnitude of the price reduction relative to the initial demand.	Police / IR; Transcripts	53.1
Stage 5: Payment				
Outcome (Stage 5)	**5. Payment**	Whether the ransom was paid (yes/no).	Police / IR; Transcripts	7.2

Note: Missingness varies systematically by stage. Structural variables and outcomes are observed for most incidents, while pricing and discount variables are available only for incidents that proceed to negotiation. Missing values therefore reflect mostly stage-dependent observability rather than data loss.

stages (RQ1) and payment and post-payment outcomes (RQ3). Negotiation transcripts are analyzed separately to examine negotiations conditional on engagement (RQ2). All analyses were conducted in R (version 4.3.1), using standard libraries for logit regressions.

Stage-wise models (police and IR data). For the first five stages of the model, we estimate regression models with stage-specific outcomes. Binary outcomes for attacker negotiation (M1), victim negotiation (M2), discount offered (M3), and payment (M5), are modeled using logit regressions. Discount size (M4) is modeled using ordinary least squares (OLS). Predictors are introduced iteratively according to the six-stage negotiation framework (see Table 2). Each model is estimated on the subset of incidents that have reached the corresponding stage. Apparent missing values therefore reflect structural features of the negotiation process rather than data quality issues, and no imputation is performed. This design avoids conditioning on post-treatment information. Goodness-of-fit metrics (McFadden's pseudo-R^2 for logit models and adjusted R^2 for OLS models) are reported in the Appendix. The primary purpose of these models is to identify associations across stages rather than to maximize predictive accuracy.

Transcript-based models (conditional on negotiation). To analyze negotiations once communication has begun (RQ2), we estimate two logistic regression models using the negotiation transcript dataset. The first models whether a discount is offered (Stage 3). The second models payment (Stage 5). Independent variables capture message volume, initial ransom demand, and discount size. These models are estimated conditional on negotiation and are not part of the stage-wise incident-level regressions. McFadden's pseudo-R^2 are reported for both models as Goodness-of-fit metric.

4 Results

4.1 Factors Influencing Negotiations (RQ1)

4.1.1 Descriptive Results. We begin with descriptive patterns observed in the police and incident response (IR) dataset, which provides end-to-end visibility into ransomware incidents, including whether negotiation occurred, whether discounts were offered, and whether payment followed.

Figure 2 shows the same incidents grouped by ransom amount and by their position in the six stage model. It distinguishes incidents that end without negotiation, incidents that involve negotiation but no discount, and incidents that involve both negotiation and discounting. For discounted cases, both the initial and final ransom demands are shown.

Figure 3 shows a staged funnel in which payment rates increase monotonically as incidents progress through the negotiation process. When communication is not allowed, the payment rate is 18.3%. Among incidents where communication is allowed, payment increases to 32.0%. Conditional on victim engagement, the payment

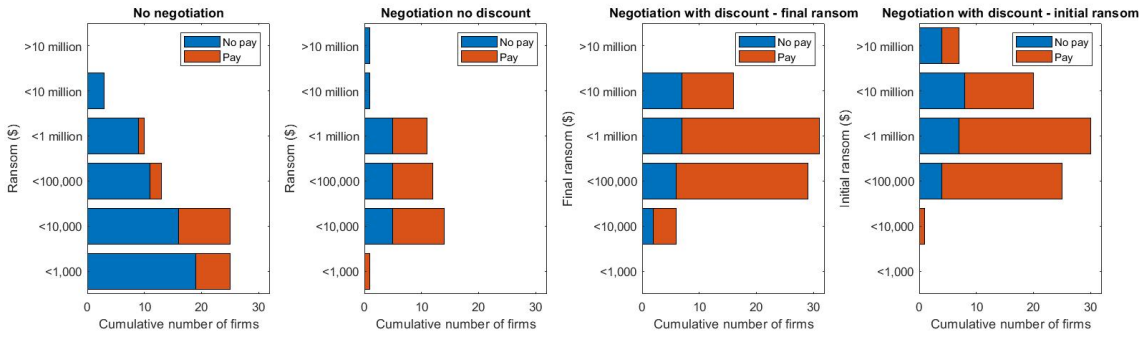


Figure 2: Negotiation outcomes by ransom amount distinguishing cases with no negotiation and negotiation. In the case of negotiation with discount we provide data for both the final ransom demand and initial ransom demand.

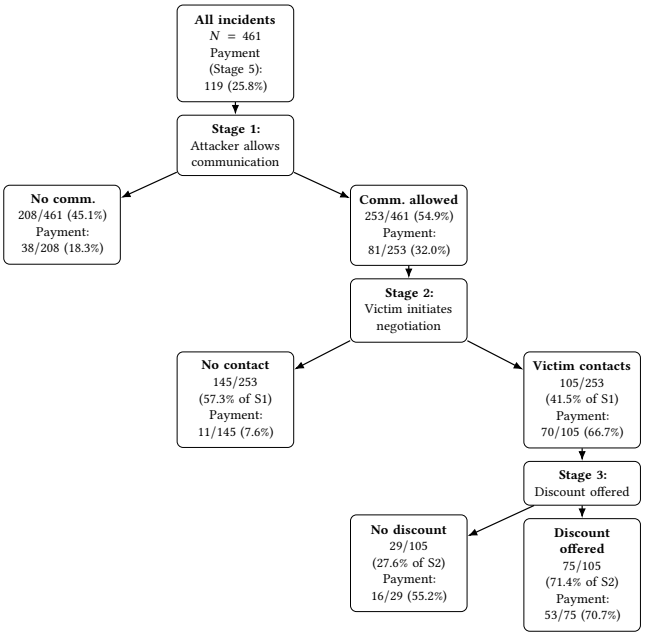


Figure 3: Progression of ransomware incidents through negotiation stages and payment outcomes. Payment rates are reported within each subset.

rate rises to 66.7%. Among negotiated incidents that reach a discount stage, the payment rate increases further to 70.7%. Payment therefore occurs primarily in incidents that proceed beyond the initial stages of the negotiation process.

Across both figures, a consistent pattern emerges. Incidents that end in the early stages of the model are concentrated at lower ransom levels and less frequently result in payment. For fixed demands below €100,000 that do not involve negotiation, payment rates are approximately 15 to 25%, and these incidents make up most observations at the lower end of the ransom distribution.

When incidents proceed to negotiation but do not reach the discount stage, payment rates are substantially higher. For these cases, which typically involve ransom demands between €1,000 and €1 million, approximately 50 to 60% result in payment. The highest payment rates are observed in incidents that proceed further and involve discounting. These incidents are concentrated at higher ransom levels, roughly between €10,000 and €10 million, and show payment rates between 60 and 85%. Although fewer in number, they account for a large share of total ransom revenue.

Taken together, these patterns show three consistent features of the six-stage negotiation process. First, incidents that end before any negotiation begins rarely lead to payment. Second, once negotiation starts, the likelihood of payment increases substantially. Third, although discounts are more common in higher-value cases, larger discounts do not consistently lead to higher payment rates. Payments occur both when discounts are small and when they are large, and the size of the discount does not appear strongly correlated with the probability of payment. These patterns suggest that what matters most is whether an incident progresses through the negotiation stages, rather than the exact price or size of the discount. Incidents that reach later stages represent a selective subset of attacks involving sustained interaction between attacker and victim. These descriptive results motivate the regression analyses that follow, which examine how victim characteristics, attacker behavior, and observable negotiation actions relate to stage progression and payment outcomes.

4.1.2 Regression Results. Table 3 summarizes the direction and magnitude of statistically significant predictors for each stage with a relatively permissive threshold ($\alpha = 0.05$). The full regression results, including estimated β-coefficients and p-values, are presented in the Appendix (see Table 6).

Stage 1: Attacker allows negotiation. Attackers are more likely to enable negotiation when data exfiltration occurred and when some form of backups is present. The backup result appears counterintuitive at first, since backups reduce the victim's need to pay. A plausible explanation is selection: organizations with backups tend to be larger, more complex, or operationally critical, making them attractive targets for extended engagement rather than fixed pricing. We find no evidence that attacker negotiation depends on victim revenue at this stage, possibly because attackers lack visibility into financial capacity before interaction begins.

Stage 2: Victim initiates negotiation. Victim-side behavior changes markedly once communication is possible. Higher-revenue

Table 3: Direction and magnitude of significant predictors across ransomware negotiation (Stages 1-5; RQ1). Cells indicate the sign and relative magnitude of statistically significant effects; dots denote predictors that are not applicable at a given stage.

Predictor	M1: Attacker negotiation	M2: Victim negotiation	M3: Discount offered	M4: Discount size	M5: Payment
Structural and attack characteristics (inputs from Stage 1 onward)					
1a. Revenue (log)	ns	+	ns	−	ns
1b. RaaS affiliation	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	+	ns	ns	−	ns
Stage 1 outcome / later-stage inputs					
1g. Negotiation (attacker)	*Outcome*	.	.	ns	ns
Stage 2: Victim-side preparedness and response					
2a. Cyber insurance	.	ns	ns	−	ns
2b. IR firm involvement	.	++	++	ns	ns
2c. Negotiation (victim)	.	*Outcome*	.	ns	++
Stage 3: Interaction dynamics					
3a. Negotiation duration (log)	.	.	+	ns	−−
3b. Initial ransom demand (log)	.	.	.	++	ns
3c. Discount offered	.	.	*Outcome*	.	++
Stage 4: Price adjustment					
4. Discount size	.	.	.	*Outcome*	ns
Stage 5: Payment					
5. Payment	.	.	.	.	*Outcome*

Legend: ns = not significant; ++ / + = significant positive effect (large / moderate); − / −− = significant negative effect (moderate / large); *Outcome* indicates the dependent variable of that stage; '.' = predictor not applicable at that stage.

organizations are more likely to negotiate, suggesting greater capacity or incentive to engage rather than immediately refuse or restore. Incident response firm involvement is the strongest predictor at this stage, indicating that professional responders actively facilitate negotiation as part of incident handling. In contrast, victims with fully recoverable backups are substantially less likely to negotiate, consistent with having a credible outside option.

Stage 3: Discount offered. Discounts are more likely when negotiations last longer and when an IR firm is involved. This pattern suggests that discounts emerge through sustained interaction rather than as an immediate concession. Longer duration of communication exchanges may give attackers information about victim constraints and willingness to pay, while IR involvement may increase pressure to reach agreement. Structural characteristics such as revenue and backups appear to play a limited role in victim's decision-making once negotiation dynamics take over.

Stage 4: Discount size. The magnitude of the discount is primarily driven by the initial ransom demand. Higher starting prices allow more room for reduction. At the same time, discounts are smaller when attacks are linked to Ransomware-as-a-Service groups or when victims have cyber insurance. This is consistent with more rigid pricing rules in affiliate-based ecosystems and with insurance anchoring expectations about acceptable payments.

Stage 5: Payment. Payment is strongly associated with prior engagement: victims who negotiate and receive a discount are more likely to pay. However, longer negotiations reduce payment likelihood, suggesting that extended bargaining may reflect disagreement, delay, or strategic stalling rather than progress toward settlement. Fully recoverable backups remain a strong deterrent to payment, even at this final stage. Initial ransom size has a weak, negative association after accounting for prior interaction.

In summary, the regression results reinforce the descriptive finding that ransomware outcomes are driven less by the size of ransom demands, and more by whether incidents advance through the negotiation process and the victims' recovery alternatives.

4.2 Negotiation Strategy and Discount (RQ2)

This subsection examines negotiation behavior conditional on communication having started.We use publicly available ransomware negotiation transcripts (see Section 3.1).

4.2.1 Descriptive Results. The negotiation transcript dataset spans 237 negotiations across 23 ransomware groups (see Appendix A, Table 7). Four groups account for 58% of the observed negotiations: Akira ($n = 61$), LockBit 3.0 ($n = 45$), Conti ($n = 32$), and REvil ($n = 20$). Akira, Conti, and REvil show similar payment rates between 31 and 35%, with average discounts in the range of 20 to 40%. LockBit 3.0 differs from this pattern, with a mean discount of 7.7% and a payment rate of 6.7%, indicating more rigid pricing and fewer successful settlements.

Negotiation transcripts illustrate how attackers structure interaction before pricing is discussed. Initial messages typically combine generic pressure with a request that the victim confirm authority to negotiate. Substantive ransom demands are introduced only after the victim signals willingness to engage, consistent with the six-stage model. The following extract illustrates the transition from initating contact (Stage 1–2) to negotiation (Stage 3–4):

> **Attacker:** Hello. You've reached Attacker 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

Table 4: Negotiation outcomes categorized by message volume (Stages 1-5; RQ2). All monetary values are expressed in $ millions.

		Number of messages exchanged				
		0–10	11–25	26–50	51–100	100+
	Number of conversations (*n*)	43	44	65	63	22
Stage 1–2	Initial demand observed (%)	41.9	65.9	80.0	92.1	90.9
Stage 3	Discount offered (%)	0.0	22.7	46.2	73.0	86.4
Stage 4	Discount size (%; mean / median)	0.0 / 0.0	23.3 / 23.3	42.2 / 42.0	49.5 / 58.3	65.1 / 65.8
	Discount size ($M; mean / median)	0.00 / 0.00	0.39 / 0.23	1.24 / 0.29	1.44 / 0.16	2.32 / 0.25
Stage 5	Payment occurred (%)	0.0	4.5	20.0	44.4	45.5

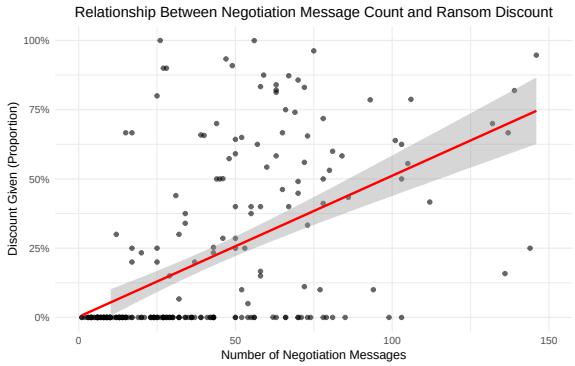


Figure 4: Relationship between message volume and likelihood of discount.

> 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?
>
> **Attacker:** 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?
>
> **Attacker:** 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 interact. Initial ransom demands appear only after that signal (Stage 1 and 2).

Analyzing the data in aggregate, Table 4 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 none of the victims paid. Once talks extended to 11–25 messages, 23% of cases saw a discount (median 23%), with payment 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 4 illustrates the relationship between negotiation length and discount, by plotting discount percentage against raw message count, with a fitted line and confidence interval. While some short exchanges still yield a large discount, the overall trend is that more messages lead to larger discounts.

4.2.2 Regression Results. Two logistic regression models are estimated conditional on negotiation: the first whether a discount is offered (Stage 3), and the second whether payment occurs (Stage 5).

Consistent with the descriptive patterns, 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 ($p = .12$). When modeling payment, message volume remains a strong 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, $p < .001$). Discount size is not statistically significant once message volume is accounted for, while larger initial ransom demands significantly reduce the likelihood of payment.

Taken together, the transcript-based results indicate that negotiation strategy primarily affects how bargaining unfolds rather than whether settlement occurs. Sustained engagement increases the likelihood that attackers concede on price, but does not reliably increase the probability of payment. Within the six-stage framework, outcomes at earlier stages, in particular whether negotiation occurs at all and whether victims remain engaged, appear more closely associated with payment than later-stage discounts. Stages 3 and 4 shape the path of negotiation, but do not independently determine settlement once engagement is established.

4.2.3 Compatibility of Data Sources. We compared descriptive statistics across the police/IR data (N = 461) and the ransomware.live transcript dataset (N = 237). Across both datasets, we observe consistent patterns in negotiations. Discounts are common once negotiation occurs, with discount rates of 37.5% in the police/IR data and 40.9% in the transcript data. Conditional on cases where a discount is offered, median discount sizes are similar across datasets (approximately 50% in the police/IR data and 57.3% in the transcript data). Payment rates are also of comparable magnitude (27.8% in police/IR data and 22.4% in transcript data).

At the same time, the datasets differ in scope and selection. The transcript dataset consists only of cases that progressed to active negotiation, whereas the police/IR dataset includes the full set of

Table 5: Summary of post-payment outcomes (Stage 6; RQ3). Decryptor failures and renegotiations are only possible for the 119 paying victims, whereas reinfection is possible for all 461 incidents. Observations are derived from case study analysis.

Category	Subtype	Observations from Case Studies
1. Decryptor failure	No key delivered	Actor ceased contact after payment; affected one SME and two large organizations; backups ranged from none to partial.
	Partial or delayed or initially unusable	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	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	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	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	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	Later intrusion detected about 18 months after first; encryption prevented in time.

incidents, including those without negotiation. This explains the higher observed discount rates and larger discount sizes in the public transcript data. In addition, ransom demands are systematically higher in the transcript dataset (median €735K) than in the police/IR data (median €40K), reflecting that publicly posted negotiations tend to involve larger, more complex incidents. These patterns are robust when comparing a random subsamples comprising 15% of both datasets. Overall, the similarities suggest that both datasets capture consistent negotiation strategies.

4.3 Post-Payment Outcomes (RQ3)

Across the police and IR ransomware incidents, we examined 461 ransomware incidents, of which 119 victims paid. Stage 6 captures post-payment interactions with threat actors. Three post-payment failure modes emerged (see Table 5):

(1) The decryptor was not delivered or functioned poorly.
(2) The attacker re-opened negotiations for additional payment after an initial payment.
(3) The victim was reinfected by the same or another ransomware strain, either after or without payment.

Precise estimates like 1.5% for delayed reinfection are unlikely to generalize beyond our descriptive case study, due to various limitations discussed in Section 5.3.

Decryptor Failures. Among the 119 victims who paid, 3 cases involved no decryptor being provided, and 8 cases 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 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 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 additional "fees" were invented after payment. We did not observe 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 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.

[1]Throughout we perturb the values for anonymity.

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, about 18 months after the first attack. The intrusion was detected early enough to prevent encryption.

Across all categories, no victims both paid in the initial incident and experienced a subsequent infection within one year. Immediate reinfections were observed only in cases where victims did not pay and restored backups without fully cleaning compromised systems. Overall, reinfection does not appear to be systematically associated with the payment decision itself. Instead, reinfections are more plausibly linked to recovery practices and residual compromise rather than to whether a ransom was paid.

5 Discussion

This section discusses the implications, limitations, and recommendations that follow from our empirical analysis.

5.1 Implications

Technical controls and recovery options. Most security controls aim to prevent or detect ransomware incidents, whereas backups primarily shape outcomes after compromise. Across the staged analyses, backup availability is one of the strongest predictors of whether incidents progress into negotiation and payment. Victims with fully recoverable backups are substantially less likely to engage or pay, while partially recoverable backups provide little benefit over no backups at all. This helps explain why backups are consistently emphasized in national cybersecurity guidance [63] and in policy debates on reducing ransomware impact [6].

At the same time, the prevalence of failed recoveries indicates that many organizations do not realize this protective value in practice [5]. Prior guidance often emphasizes offsite backups [63], yet industry evidence suggests these frequently fail due to operational and organizational factors [36]. Understanding why backup strategies break down under adversarial conditions remains an important direction for future research. This ties into existing work in usable security focused on systems operators [22, 33].

Data Exfiltration had no statistically significant association with the victim's likelihood of negotiating or paying[2], which has potential explanations in harm, trust, and research design. Privacy law may require businesses to notify customers of an unauthorized ransomware actor accessing customer data, incurring legal costs and reputational harm regardless of whether the ransom is paid [29]. By contrast, system encryption is more reversible, providing the decryptor works [24]. Another explanation is that defenders do not trust attackers to delete data after payment [48, 65]. The results may be influenced by a sampling bias given we excluded exfiltration only cases, and also that victims valuing data confidentiality are perhaps less likely to report to the police.

Discounts and negotiation strategy. The results show that discounts are common once negotiation begins and are strongly associated with negotiation duration, consistent with prior observations of prolonged bargaining in ransomware incidents [54, 81]. However, discount size has limited explanatory power for payment once early-stage conditions are accounted for. This challenges the common assumption that securing larger discounts directly increases the likelihood of settlement, an idea frequently implied in practitioner guidance [18, 19]. Instead, discounts appear to function as part of the negotiation process in incidents that have already progressed beyond early stages, rather than as a primary driver of payment decisions. Improved data sharing between victims, law enforcement, incident response firms, and insurers could reduce informational asymmetries that shape negotiation outcomes and perceived bargaining positions [77].

Cyber insurance. Cyber insurance has been criticized for encouraging ransom payments [1, 37]. Within the six-stage framework, we find no evidence that insurance directly increases payment likelihood. Instead, insurance is associated with indirect and offsetting effects. Insured victims are more likely to involve professional incident response firms [77], which increases engagement and the likelihood of receiving a discount. Insurance is also associated with smaller discounts, potentially as ransomware actors discover the policy and aim to extract the full limit.

One explanation for insurance not directly increasing payment likelihood is that insurance frequently covers business interruption losses in addition to extortion [62]. This allows victims to tolerate longer outages without paying. As a result, insurance reduces the victim's out-of-pocket cost of paying and also the cost of not paying. These effects cancel each other out, leaving the overall payment rates largely unchanged. However, cyber insurance is still associated with higher ransom amounts paid [49]. Selection effects likely contribute, as victims with greater dependence on IT systems are both more likely to purchase insurance and to be targeted.

Reputation and post-payment behavior. Post-payment outcomes suggest that re-extortion is relatively rare in our descriptive case study. While decryptor failures, delays, and renegotiation occur, they are infrequent and often linked to technical remediation issues rather than deliberate opportunism. Prior work similarly documents that maintaining decryption reliability is common among established ransomware groups [71]. These findings indicate that ransomware negotiations should be understood as repeated interactions across incidents, in which attackers have incentives to maintain credibility. This aligns with broader evidence that trust and coordination is an important component in cybercrime markets [2, 43]. It also helps explain why infrastructure takedowns and coordinated interventions can have disproportionate effects [52, 72].

Implications for Ransomware Groups. After reading this article, an adversary might focus on encrypting systems instead of exfiltrating data, taking extra effort to destroy backups or locate the insurance policy. This possibility speaks against publication. However, adversaries are unlikely to read English-language scientific articles, let alone use the results to craft criminal playbooks.

[2] Although attackers were more likely to open negotiations and offer a smaller discount when data was exfiltrated

All the aforementioned strategies are discussed in grey literature and also provided by LLMs without any adversarial prompting. Defenders, especially policymakers, are much more likely to use scientific articles to change strategy.

5.2 Recommendations

1. Focus on recoverability and resilience, not just backups. Negotiation outcomes for victims with partial/unrecoverable backups were nearly indistinguishable from victims with no-backups (see Table 3). Further, ensuring recoverable backups appears to positively influence negotiations more than preventing data exfiltration. Defenders should divert resources to ensuring backups actually allow recovery, but granular recommendations are less forthcoming. The 3-2-1 rule recommends keeping three copies, on two different types of media, with at least one stored offsite. However, backups often fail in practice, with offsite backups having the highest failure rate [36].[3] Future work should focus on foundational questions like "who do backups fail" in order to make better recommendations.

2. Hire expert negotiators, and protect the insurance policy. Victims that hire incident response firms are much more likely to receive a discount, but no more likely to pay (see Table 3). This is consistent with IR firms improving negotiations while leaving the moral decision of whether to pay with the client. However, insurance policyholders receive smaller discounts, which is consistent with anecdotal evidence that threat actors search for insurance policies to justify demanding higher ransoms [45]. Hiring experts while keeping the cyber insurance policy secret achieves the best of both worlds, not least because insurance pays for those experts [77].

3. Policy should focus on improving preparedness and negotiation discipline, not outright bans. The six-stage results show that payment outcomes are largely determined at Stages 1 and 2, before pricing and bargaining dynamics unfold. Policies that intervene at later stages, such as payment bans [6], target Stage 5 where leverage is limited and impose high costs on victims lacking effective recovery options. Our findings instead indicate greater promise in interventions that strengthen early-stage preparedness and recovery capabilities, such as by mandating backup maintenance processes as suggested by Recommendation 1.

Policies could be introduced to improve negotiation discipline by capping ransom payments. Rather than banning ransoms, policymakers could establish an upper limit on payments (perhaps proportional to revenue) that improves the credibility of a victim's stated maximum payment. To avoid insured victims facing higher demands (see Table 3), the industry could explore applying a lower sublimit for indemnifying ransom payments (perhaps offset by a higher business interruption sublimit). However, insurers are unlikely to unilaterally introduce these sublimits for fear of losing business, and so a regulator would likely be needed for coordination.

5.3 Limitations

Scope and generalizability. The incident-level analyses rely on ransomware cases reported in the Netherlands between 2019 and 2023. Generalization to victims in other countries may be threatened by factors like reporting requirements, regulatory constraints, and norms around ransom payment. For example, US victims may be more willing to pay for data suppression because the US has more data breach litigation. However, the influence of variables like backups, insurance, and IR firms is less likely to differ by country.

To diversify our sample, we also collected public negotiation transcripts from multiple jurisdictions, although the exact geographic distribution of victims cannot be reliably established. Nevertheless, the findings from these global victims confirmed the equivalent findings from cases reported in the Netherlands. As a result, the findings of our study are most directly applicable to well-resourced organizations operating in Western regulatory environments.

Data access and reproducibility. Access to the police and incident response datasets was restricted to a single researcher due to confidentiality and victim protection requirements. While this prevents data sharing and validation via an independent second coder, it was a necessary pragmatic choice to build a large sample of ransomware victims. For example, Van et al. [71] explain that access constraints are common in empirical ransomware research. To mitigate this limitation, we provide anonymized replication artifacts that preserve the full quantitative analysis workflow and model specifications.

Omitted variables and causal inference. Several factors that likely shape negotiation outcomes are not directly observable. This include the attackers' perceived reputation, victims' expectations about decryption reliability, internal organizational deliberations, and legal or regulatory constraints during incident response [10, 77]. While these omitted variables may impact validity, we at least included proxies for these kinds of factors. For example, our RaaS variable captures aspects related to attacker reputation, and the volume of negotiation provides a proxy for internal deliberation.

A further challenge concerns causality given the analysis is observational. For example, it is unclear whether hiring an IR firm increases the likelihood of negotiation (causal interpretation), or whether victims who want/need to negotiate hire IR firms to help them (non-causal). As such, the analyses identify associations rather than causal effects.

Negotiation content and pricing dynamics. The study used relatively crude variables related to negotiation (duration and volume of messages). We did not analyze the linguistic content of negotiation messages. Consequently, the deployment of psychological, persuasive, or coercive negotiation strategies are only captured indirectly through proxies such as message volume or discounting behavior [23]. We also do not explicitly model potential endogeneity between ransom pricing and payment decisions, as pricing primarily reflects downstream selection into negotiation rather than an independent treatment. Including content-level analysis or structural models of attacker pricing is an important direction for future work.

6 Conclusion

This paper applies a six-stage model of ransomware negotiation to incident-level data and public negotiation transcripts. Modeling ransomware as a sequence of decisions, rather than as a single payment

[3] Storing backups offsite reduces the risk of attackers corrupting them, but makes restoration and testing more time consuming, particularly for tape backups [57].

choice, shows that outcomes are shaped by selective progression through stages of engagement, bargaining, and settlement.

Across both datasets, payment outcomes are largely determined before negotiation begins. Incidents with fully recoverable backups rarely progress to negotiation or payment, whereas destroyed or missing backups are associated with higher rates of negotiation and payment. Discount size itself does not meaningfully influence payment probability. Incident response firm involvement increases the likelihood of negotiation and discounting, but does not affect the probability of payment.

Contradicting common assumptions found in the grey literature, we find that: insurance does not directly increase payment rates; discounts function as bargaining tools rather than simple price reductions; and, post-payment risks such as re-extortion or decryptor failure are not common in our descriptive case study.

Ethical Considerations

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 revictimization. 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 [4], 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 [71]. 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.

Open Science

Artifacts shared. We aim for maximum transparency and reproducibility within the constraints imposed by confidentiality agreements and privacy considerations. To support evaluation of the paper's core contributions, we provide anonymized replication artifacts that reproduce the full analytical workflow used in the study. For the police and incident response (IR) datasets used in Section 4.1, we share anonymized R scripts that replicate data preprocessing, model specification, estimation, and figure/table generation using synthetic placeholder inputs. These scripts preserve the full structure and logic of the analyses while removing victim- and partner-identifying information, enabling reviewers to assess the methodology and re-run the models on substitute data. For the public ransomware negotiation analysis in Section 4.2, we provide an edited R script that performs data cleaning, modeling, and visualization, together with a detailed, reproducible codebook specifying variables, model definitions, and output generation. All shared artifacts are hosted on an anonymous repository at https://anonymous.4open.science/r/ransomware_bargaining-5D01, accessible during double-blind review. For the post-payment case analysis in Section 4.3, no executable script is provided, as the analysis was conducted manually. The coding procedure and decision rules are fully documented in the paper.

Artifacts not shared. We do not release the raw police or IR datasets, as they contain sensitive, non-public information about ransomware incidents and affected organizations. These restrictions reflect legal, ethical, and safety considerations rather than technical limitations.

Generative AI Usage

Generative AI tools were used to assist with language editing and structural refinement of the manuscript. Specifically, a large language model (ChatGPT 5.1) was used to help rewrite and clarify portions of the abstract, introduction, discussion, and conclusion for readability, consistency, and tone. The AI tools were not used to generate research ideas, experimental designs, data, code, statistical analyses, or results. All empirical analyses, figures, tables, and interpretations were produced by the authors. AI-assisted text was reviewed, edited, and validated by the authors through manual verification to ensure accuracy, consistency with the reported results, and correct citation of prior work.

References

[1] Lawrence Abrams. 2021. Conti Ransomware Prioritizes Revenue and Cyberinsurance Data Theft. Retrieved July 18, 2025 from https://www.bleepingcomputer.com/news/security/conti-ransomware-prioritizes-revenue-and-cyberinsurance-data-theft/

[2] Sadia Afroz, Vaibhav Garg, Damon McCoy, and Rachel Greenstadt. 2013. Honor among Thieves: A Common's Analysis of Cybercrime Economies. In 2013 APWG eCrime Researchers Summit. IEEE, Piscataway, NJ, USA, 1–11. https://doi.org/10.1109/eCRS.2013.6805778

[3] Bushra A. Alahmadi, Louise Axon, and Ivan Martinovic. 2022. 99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms. In 31st USENIX Security Symposium (USENIX Security 22). USENIX Association, Berkeley, CA, USA, 2783–2800. Retrieved June 19, 2025 from https://www.usenix.org/conference/usenixsecurity22/presentation/alahmadi

[4] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. 2012. The Menlo Report. IEEE Security & Privacy 10, 2 (2012), 71–75. https://doi.org/10.1109/MSP.2012.52

[5] Pranshu Bajpai and Richard Enbody. 2023. Know Thy Ransomware Response: A Detailed Framework for Devising Effective Ransomware Response Strategies. Digital Threats: Research and Practice 4, 4 (2023), 1–19. https://doi.org/10.1145/3606022

[6] Robert Booth. 2025. UK Government to Ban Public Bodies from Paying Ransoms to Hackers. Retrieved January 14, 2026 from https://www.theguardian.com/technology/2025/jul/22/uk-government-to-ban-public-bodies-from-paying-ransoms-to-hackers

[7] Simona Boticiu and Fabian Teichmann. 2024. How Does One Negotiate with Ransomware Attackers? International Cybersecurity Law Review 5, 1 (2024), 55–65. https://doi.org/10.1365/s43439-023-00106-w

[8] Casey Breen, Cormac Herley, and Elissa M. Redmiles. 2022. A Large-Scale Measurement of Cybercrime Against Individuals. In CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, United States, 1–41. https://doi.org/10.1145/3491102.3517613

[9] Jack Cable, Ian W. Gray, and Damon McCoy. 2024. Showing the Receipts: Understanding the Modern Ransomware Ecosystem. In 2024 APWG Symposium on Electronic Crime Research (eCrime). IEEE, Piscataway, NJ, USA, 149–161. https://doi.org/10.1109/ECRIME66200.2024.00017

[10] Anna Cartwright and Edward Cartwright. 2019. Ransomware and Reputation. Games 10, 2 (2019), 26. https://doi.org/10.3390/g10020026

[11] Anna Cartwright, Edward Cartwright, Jamie MacColl, Gareth Mott, Sarah Turner, James Sullivan, and Jason R. C. Nurse. 2023. How Cyber Insurance Influences the Ransomware Payment Decision: Theory and Evidence. The Geneva Papers on Risk and Insurance – Issues and Practice 48, 2 (2023), 300–331. https://doi.org/10.1057/s41288-023-00288-8

[12] Anna Cartwright, Edward Cartwright, Lian Xue, and Julio Hernandez-Castro. 2023. An Investigation of Individual Willingness to Pay Ransomware. Journal of Financial Crime 30, 3 (2023), 728–741. https://doi.org/10.1108/JFC-02-2022-0055

[13] Edward Cartwright, Julio Hernandez-Castro, and Anna Cartwright. 2019. To Pay or Not: Game Theoretic Models of Ransomware. Journal of Cybersecurity 5, 1 (2019), 1–12. https://doi.org/10.1093/cybsec/tyz009

[14] Chainalysis. 2024. Ransomware Payments Exceed $1 Billion in 2023, Hitting Record High After 2022 Decline. Retrieved March 27, 2025 from https://www.chainalysis.com/blog/ransomware-2024/

[15] Paul Cichonski, Tom Millar, Tim Grance, and Karen Scarfone. 2012. Computer Security Incident Handling Guide. Technical Report NIST Special Publication 800-61 Revision 2. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-61r2

[16] Coalition. 2025. 2025 Cyber Claims Report. Retrieved January 5, 2026 from https://web.coalitioninc.com/download-2025-cyber-claims-report.html

[17] Alena Yuryna Connolly and Hervé Borrion. 2022. Reducing Ransomware Crime: Analysis of Victims' Payment Decisions. Computers & Security 119 (2022), 102760. https://doi.org/10.1016/j.cose.2022.102760

[18] Coveware. 2019. Coveware's 2018 Q4 Ransomware Marketplace Report. Retrieved April 8, 2025 from https://coveware.com/2019/01/covewares-2018-q4-ransomware-marketplace-report/

[19] Coveware. 2025. Targeted Social Engineering Is En Vogue as Ransom Payment Sizes Increase. Retrieved October 4, 2025 from https://www.coveware.com/2025/07/targeted-social-engineering-is-en-vogue-as-ransom-payment-sizes-increase

[20] Cybersecurity and Infrastructure Security Agency. 2023. #StopRansomware: CL0P Ransomware Gang Exploits CVE-2023-34362 MOVEit Vulnerability. Retrieved August 12, 2025 from https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-158a

[21] Cyentia Institute. 2025. 2025 Information Risk Insights Study. Retrieved January 4, 2026 from https://www.cyentia.com/publication/iris2025/

[22] Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. 2018. Investigating System Operators' Perspective on Security Misconfigurations. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. ACM, New York, NY, USA, 1272–1289. https://doi.org/10.1145/3243734.3243794

[23] Juliette Faivre. 2023. Negotiations in Tech: An Analysis of Asymmetric Ransomware Negotiations. https://doi.org/10.2139/ssrn.4530094

[24] Burak Filiz, Budi Arief, Orcun Cetin, and Julio Hernandez-Castro. 2021. On the Effectiveness of Ransomware Decryption Tools. Computers & Security 111 (2021), 102469. https://doi.org/10.1016/j.cose.2021.102469

[25] Jason Franklin, Adrian Perrig, Vern Paxson, and Stefan Savage. 2007. An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants. In Proceedings of the 14th ACM Conference on Computer and Communications Security. ACM, New York, NY, USA, 375–388. https://doi.org/10.1145/1315245.1315292

[26] Anjuli Franz, Verena Zimmermann, et al. 2021. 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). USENIX Association, Berkeley, CA, USA, 339–358. Retrieved May 18, 2025 from https://www.usenix.org/conference/soups2021/presentation/franz

[27] Freya Gassmann, Janina Beck, Nora Gourmelon, and Zinaida Benenson. 2025. Valuation of Confidentiality and Availability in a Personal Ransomware Attack Scenario. Journal of Cybersecurity 11, 1 (2025), tyaf022. https://doi.org/10.1093/cybsec/tyaf022

[28] Ian W. Gray, Jack Cable, Benjamin Brown, Vlad Cuiujuclu, and Damon McCoy. 2022. Money over Morals: A Business Analysis of Conti Ransomware. In 2022 APWG Symposium on Electronic Crime Research (eCrime). IEEE, Piscataway, NJ, USA, 1–12. https://doi.org/10.1109/eCrime57793.2022.10142119

[29] Ece Gumusel, Yue Xiao, Yue Qin, Jiaxin Qin, and Xiaojing Liao. 2024. Understanding Legal Professionals' Practices and Expectations in Data Breach Incident Reporting. In Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security. ACM, New York, NY, USA, 2711–2725. https://doi.org/10.1145/3658644.3690357

[30] Mischa Hansel and Jantje Silomon. 2024. Ransomware as a Threat to Peace and Security: Understanding and Avoiding Political Worst-Case Scenarios. Journal of Cyber Policy 9, 2 (2024), 159–178. https://doi.org/10.1080/23738871.2024.2357092

[31] 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. 2018. Tracking Ransomware End-to-End. In 2018 IEEE Symposium on Security and Privacy (SP). IEEE, Piscataway, NJ, USA, 618–631. https://doi.org/10.1109/SP.2018.00047

[32] Jian Huang, Jun Xu, Xinyu Xing, Peng Liu, and Moinuddin K. Qureshi. 2017. FlashGuard: Leveraging Intrinsic Flash Properties to Defend Against Encryption Ransomware. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, New York, NY, USA, 2231–2244. https://doi.org/10.1145/3133956.3134035

[33] Adam D. G. Jenkins, Linsen Liu, Maria K. Wolters, and Kami Vaniea. 2024. Not as Easy as Just Update: Survey of System Administrators and Patching Behaviours. In Proceedings of the CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, United States, 1–17. https://doi.org/10.1145/3613904.3642456

[34] Harjot Kaur, Sabrina Klivan, Daniel Votipka, Yasemin Acar, and Sascha Fahl. 2022. Where to Recruit for Security Development Studies: Comparing Six Software Developer Samples. In 31st USENIX Security Symposium (USENIX Security 22). USENIX Association, Berkeley, CA, USA, 4041–4058. Retrieved June 19, 2025 from https://www.usenix.org/conference/usenixsecurity22/presentation/kaur

[35] Amin Kharaz, Sajjad Arshad, Collin Mulliner, William Robertson, and Engin Kirda. 2016. UNVEIL: A Large-Scale, Automated Approach to Detecting Ransomware. In 25th USENIX Security Symposium (USENIX Security 16). USENIX Association, Berkeley, CA, USA, 757–772. Retrieved June 18, 2025 from https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/kharaz

[36] Liat Klainman and Greg Otto. 2023. Backup Breakdown: How Data Recovery Impacts the Outcome of Cyberattacks. Retrieved September 3, 2025 from https://www.at-bay.com/wp-content/uploads/2023/10/At-Bay_Backup-Breakdown.pdf

[37] Zoe Kleinman. 2021. Insurers Defend Covering Ransomware Payments. Retrieved August 21, 2025 from https://www.bbc.co.uk/news/technology-55811165

[38] Faris Bugra Kokulu, Ananta Soneji, Tiffany Bao, Yan Shoshitaishvili, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. 2019. Matched and Mismatched SOCs: A Qualitative Study on Security Operations Center Issues. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. ACM, New York, NY, USA, 1955–1970. https://doi.org/10.1145/3319535.3354239

[39] Eugene Kolodenker, William Koch, Gianluca Stringhini, and Manuel Egele. 2017. PayBreak: Defense Against Cryptographic Ransomware. In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. ACM, New York, NY, USA, 599–611. https://doi.org/10.1145/3052973.3053035

[40] Aron Laszka, Sadegh Farhang, and Jens Grossklags. 2017. On the Economics of Ransomware. In International Conference on Decision and Game Theory for Security. Springer Nature Switzerland AG, Cham, Switzerland, 397–417. https://doi.org/10.1007/978-3-319-68711-7_21

[41] Zhen Li and Qi Liao. 2022. Preventive Portfolio Against Data-Selling Ransomware—A Game Theory of Encryption and Deception. Computers & Security 116 (2022), 102644. https://doi.org/10.1016/j.cose.2022.102644

[42] Kyle D. Logue and Adam B. Shniderman. 2021. The Case for Banning (and Mandating) Ransomware Insurance. Connecticut Insurance Law Journal 28 (2021), 247–316. Retrieved September 11, 2025 from https://digitalcommons.lib.uconn.edu/cilj/398

[43] Jonathan Lusthaus. 2012. Trust in the World of Cybercrime. Global Crime 13, 2 (2012), 71–94. https://doi.org/10.1080/17440572.2012.674183

[44] Jamie MacColl, Pia Hüsch, Gareth Mott, James Sullivan, Jason R. C. Nurse, Sarah Turner, and Nandita Pattnaik. 2024. The Scourge of Ransomware: Victim Insights on Harms to Individuals, Organisations and Society. Retrieved September 22, 2025 from https://www.rusi.org/explore-our-research/publications/occasional-papers/ransomware-victim-insights-harms-individuals-organisations-and-society

[45] Jamie MacColl, James Sullivan, Jason R. C. Nurse, Gareth Mott, Sarah Turner, Edward Cartwright, and Anna Cartwright. 2023. Cyber Insurance and the Ransomware Challenge. Retrieved October 1, 2025 from https://www.rusi.org/explore-our-research/publications/occasional-papers/cyber-insurance-and-ransomware-challenge

[46] Tina Marjanov and Alice Hutchings. 2025. SoK: Digging into the Digital Underworld of Stolen Data Markets. In 2025 IEEE Symposium on Security and Privacy (SP). IEEE, Piscataway, NJ, USA, 1–18. https://doi.org/10.1109/SP61157.2025.00037

[47] Sifra R. Matthijsse, Asier Moneva, M. Susanne van 't Hoff-de Goede, and Eric Rutger Leukfeldt. 2025. Examining Ransomware Payment Decision-Making Among Small- and Medium-Sized Enterprises. European Journal of Criminology 22, 4 (2025), 625–645. https://doi.org/10.1177/14773708241285671

[48] Timothy McIntosh, Teo Susnjak, Tong Liu, Dan Xu, Paul Watters, Dongwei Liu, Yaqi Hao, Alex Ng, and Malka Halgamuge. 2024. Ransomware Reloaded: Re-examining Its Trend, Research and Mitigation in the Era of Data Exfiltration.

Comput. Surveys 57, 1 (2024), 1–40. https://doi.org/10.1145/3691340

[49] Tom Meurs, Edward Cartwright, and Anna Cartwright. 2023. Double-Sided Information Asymmetry in Double Extortion Ransomware. In Decision and Game Theory for Security (Lecture Notes in Computer Science, Vol. 14250). Springer, Cham, Switzerland, 311–328. https://doi.org/10.1007/978-3-031-50670-3_16

[50] Tom Meurs, Edward Cartwright, Anna Cartwright, Marianne Junger, Raphael Hoheisel, Erik Tews, and Abhishta Abhishta. 2024. Ransomware Economics: A Two-Step Approach To Model Ransom Paid. In 2023 APWG Symposium on Electronic Crime Research (eCrime). IEEE, Piscataway, NJ, 1–13. https://doi.org/10.1109/eCrime61234.2023.10485506

[51] Tom Meurs, Raphael Hoheisel, Marianne Junger, and Abhishta Abhishta. 2025. NAS Ransomware: Ransomware Targeting NAS Devices. Global Crime 1, 1 (2025), 1–24. https://doi.org/10.1080/17440572.2025.2591628

[52] Tom Meurs, Raphael Hoheisel, Marianne Junger, Abhishta Abhishta, and Damon McCoy. 2025. What To Do Against Ransomware? Evaluating Law Enforcement Interventions. In 2024 APWG Symposium on Electronic Crime Research (eCrime). IEEE, Piscataway, NJ, USA, 76–93. https://doi.org/10.1109/eCrime66200.2024.00012

[53] Tom Meurs, Marianne Junger, Maarten Cruyff, and Peter G. M. van der Heijden. 2025. Estimating The Number Of Ransomware Attacks. Journal of Quantitative Criminology 42 (2025), 227–243. https://doi.org/10.1007/s10940-025-09625-7

[54] Tom Meurs, Marianne Junger, Erik Tews, and Abhishta Abhishta. 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). IEEE, Piscataway, NJ, USA, 1–13. https://doi.org/10.1109/eCrime57793.2022.10142138

[55] Tyler Moore and Richard Clayton. 2007. An Empirical Analysis of the Current State of Phishing Attack and Defence. Retrieved October 14, 2025 from https://tylermoore.utulsa.edu/weis07.pdf

[56] Gareth Mott, Sarah Turner, Jason R. C. Nurse, Jamie MacColl, James Sullivan, Anna Cartwright, and Edward Cartwright. 2023. Between a Rock and a Hard(ening) Place: Cyber Insurance in the Ransomware Era. Computers & Security 128 (2023), 103162. https://doi.org/10.1016/j.cose.2023.103162

[57] Sharon D. Nelson and John W. Simek. 2018. The Basics of Backup. Law Practice 44 (2018), 26. Retrieved October 15, 2025 from https://senseient.com/wp-content/uploads/The-Basics-of-Backup.pdf

[58] Kris Oosthoek, Jack Cable, and Georgios Smaragdakis. 2023. A Tale of Two Markets: Investigating the Ransomware Payments Economy. Commun. ACM 66, 8 (2023), 74–83. https://doi.org/10.1145/3582489

[59] Harun Oz, Ahmet Aris, Abbas Acar, Güliz Seray Tuncay, Leonardo Babun, and Selcuk Uluagac. 2023. RØB: Ransomware over Modern Web Browsers. In 32nd USENIX Security Symposium (USENIX Security 23). USENIX Association, Berkeley, CA, USA, 7073–7090. Retrieved May 15, 2025 from https://www.usenix.org/conference/usenixsecurity23/presentation/oz

[60] Masarah Paquet-Clouston, Bernhard Haslhofer, and Benoit Dupont. 2019. Ransomware Payments in the Bitcoin Ecosystem. Journal of Cybersecurity 5, 1 (2019), tyz003. https://doi.org/10.1093/cybsec/tyz003

[61] Alex Rebert, Chandler Carruth, Jen Engel, and Andy Qin. 2024. Safer with Google: Advancing Memory Safety. Retrieved May 2, 2025 from https://security.googleblog.com/2024/10/safer-with-google-advancing-memory.html

[62] Sasha Romanosky, Lillian Ablon, Andreas Kuehn, and Therese Jones. 2019. Content Analysis of Cyber Insurance Policies: How Do Carriers Price Cyber Risk? Journal of Cybersecurity 5, 1 (2019), tyz002. https://doi.org/10.1093/cybsec/tyz002

[63] Kimberly Ruth, Raymond Buernor Obu, Ifeoluwa Shode, Gavin Li, Carrie Gates, Grant Ho, and Zakir Durumeric. 2025. A First Look at Governments' Enterprise Security Guidance. In 34th USENIX Security Symposium (USENIX Security 25). USENIX Association, Berkeley, CA, USA, 119–138. Retrieved January 4, 2026 from https://www.usenix.org/conference/usenixsecurity25/presentation/ruth

[64] John Sakellariadis. 2022. Behind the Rise of Ransomware. Retrieved June 12, 2025 from https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/behind-the-rise-of-ransomware/

[65] Armin Sarabi, Ziyuan Huang, Chenlan Wang, Tai Karir, and Mingyan Liu. 2025. The Ransomware Decade: The Creation of a Fine-Grained Dataset and a Longitudinal Study. In 34th USENIX Security Symposium (USENIX Security 25). USENIX Association, Berkeley, CA, USA, 4799–4818. Retrieved January 4, 2026 from https://www.usenix.org/conference/usenixsecurity25/presentation/sarabi

[66] Nolen Scaife, Henry Carter, Patrick Traynor, and Kevin R. B. Butler. 2016. CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data. In 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS). IEEE, Piscataway, NJ, USA, 303–312. https://doi.org/10.1109/ICDCS.2016.46

[67] Anastasia Sentsova. 2023. Negotiating with LockBit: Uncovering the Evolution of Operations and Newly Established Rules. Retrieved November 6, 2025 from https://analyst1.com/blog-negotiating-with-lockbit-uncovering-the-evolution-of-operations-and-newly-established-rules/ Contributor: Jon DiMaggio.

[68] Rhythima Shinde, Pieter Van der Veeken, Stijn Van Schooten, and Jan van den Berg. 2016. Ransomware: Studying Transfer and Mitigation. In 2016 International Conference on Computing, Analytics and Security Trends (CAST). IEEE, New York, NY, USA, 90–95. https://doi.org/10.1109/CAST.2016.7914946

[69] Sophos. 2025. The Sophos Annual Threat Report: Cybercrime on Main Street 2025. Retrieved May 29, 2025 from https://www.sophos.com/en-us/blog/the-sophos-annual-threat-report-cybercrime-on-main-street-2025

[70] Steve G. A. van de Weijer, Eric Rutger Leukfeldt, and Sophie van der Zee. 2021. Cybercrime Reporting Behaviors Among Small- and Medium-Sized Enterprises in the Netherlands. In Cybercrime in Context: The Human Factor in Victimization, Offending, and Policing. Springer, Chalm, Switzerland, 303–325. https://doi.org/10.1007/978-3-030-60527-8_17

[71] Max van der Horst, Ricky Kho, Olga Gadyatskaya, Michel Mollema, Michel van Eeten, and Yury Zhauniarovich. 2025. High Stakes, Low Certainty: Evaluating the Efficacy of High-Level Indicators of Compromise in Ransomware Attribution. In 34th USENIX Security Symposium (USENIX Security 25). USENIX Association, Berkeley, CA, USA, 4819–4838. Retrieved January 4, 2026 from https://www.usenix.org/conference/usenixsecurity25/presentation/van-der-horst

[72] Rolf van Wegberg and Thijmen Verburgh. 2018. Lost in the Dream? Measuring the Effects of Operation Bayonet on Vendors Migrating to Dream Market. In Evolution of the Darknet Workshop at the Web Science Conference (WebSci 18). ACM, New York, NY, USA, 1–5. Retrieved September 9, 2025 from https://research.tudelft.nl/en/publications/lost-in-the-dream-measuring-the-effects-of-operation-bayonet-on-v

[73] Verizon. 2025. 2025 Data Breach Investigations Report. Retrieved October 7, 2025 from https://www.verizon.com/business/resources/reports/2025-dbir-data-breach-investigations-report.pdf

[74] Shenao Wang, Feng Dong, Hangfeng Yang, Jingheng Xu, and Haoyu Wang. 2024. CanCal: Towards Real-time and Lightweight Ransomware Detection and Response in Industrial Environments. In Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security. ACM, New York, NY, USA, 2326–2340. https://doi.org/10.1145/3658644.3690269

[75] Noelle Warkentin, Yuxuan Zhang, and Richard Frank. 2025. The Business of Ransomware and Its Effects on Business. Journal of Cyber Security Technology 10, 1 (2025), 1–31. https://doi.org/10.1080/23742917.2025.2454620

[76] Daniel W. Woods. 2023. A Turning Point for Cyber Insurance. Commun. ACM 66, 3 (2023), 41–44. https://doi.org/10.1145/3545795

[77] Daniel W. Woods, Rainer Böhme, Josephine Wolff, and Daniel Schwarcz. 2023. Lessons Lost: Incident Response in the Age of Cyber Insurance and Breach Attorneys. In 32nd USENIX Security Symposium (USENIX Security 23). USENIX Association, Berkeley, CA, USA, 2259–2273. Retrieved June 19, 2025 from https://www.usenix.org/conference/usenixsecurity23/presentation/woods

[78] Daniel W. Woods and Lukas Walter. 2022. Reviewing Estimates of Cybercrime Victimisation and Cyber Risk Likelihood. In 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE, Piscataway, NJ, USA, 150–162. https://doi.org/10.1109/EuroSPW55150.2022.00021

[79] Tongxin Yin, Armin Sarabi, and Mingyan Liu. 2023. Deterrence, Backup, or Insurance: Game-Theoretic Modeling of Ransomware. Games 14, 2 (2023), 20. https://doi.org/10.3390/g14020020

[80] Adam Young and Moti Yung. 1996. Cryptovirology: Extortion-Based Security Threats and Countermeasures. In Proceedings 1996 IEEE Symposium on Security and Privacy. IEEE, Piscataway, NJ, USA, 129–140. https://doi.org/10.1109/SECPRI.1996.502676

[81] Chen Zhang and Feng Luo. 2025. Bargaining Game Theoretical Analysis Framework for Ransomware Attacks. Journal of Information Security and Applications 93 (2025), 104115. https://doi.org/10.1016/j.jisa.2025.104115

[82] Chunming Zhang, Fengji Luo, and Gianluca Ranzi. 2022. Multistage Game Theoretical Approach for Ransomware Attack and Defense. IEEE Transactions on Services Computing 16, 4 (2022), 2800–2811. https://doi.org/10.1109/TSC.2022.3220736

[83] Huan Zhang, Lixin Zhao, Aimin Yu, Lijun Cai, and Dan Meng. 2024. Ranker: Early Ransomware Detection Through Kernel-Level Behavioral Analysis. IEEE Transactions on Information Forensics and Security 19 (2024), 6113–6127. https://doi.org/10.1109/TIFS.2024.3410511

[84] Leah Zhang-Kennedy, Hala Assal, Jessica Rocheleau, Reham Mohamed, Khadija Baig, and Sonia Chiasson. 2018. The Aftermath of a Crypto-Ransomware Attack at a Large Academic Institution. In 27th USENIX Security Symposium (USENIX Security 18). USENIX Association, Berkeley, CA, USA, 1061–1078. Retrieved August 7, 2025 from https://www.usenix.org/conference/usenixsecurity18/presentation/zhang-kennedy

[85] Lingbo Zhao, Yuhui Zhang, Zhilu Wang, Fengkai Yuan, and Rui Hou. 2025. ERW-Radar: An Adaptive Detection System against Evasive Ransomware by Contextual Behavior Detection and Fine-grained Content Analysis. In Network and Distributed System Security Symposium (NDSS) 2025. The Internet Society, Reston, VA, USA, 1–12. Retrieved January 4, 2026 from https://www.ndss-symposium.org/ndss-paper/erw-radar-an-adaptive-detection-system-against-evasive-ransomware-by-contextual-behavior-detection-and-fine-grained-content-analysis/

A Codebook for Police and IR Dataset

This appendix provides the coding protocol used to extract variables from unstructured police and incident response (IR) reports. Variables are defined based on observable evidence in the reports and aligned with the six-stage negotiation framework. Coding prioritizes precision over recall: variables are recorded only when explicitly supported by the source data. Ambiguous or incomplete cases are coded as missing.

A.1 General Coding Principles

- **Unit of analysis:** Individual ransomware incident affecting an organization.
- **Inclusion criteria:** Cases must explicitly describe both (i) file encryption and (ii) a corresponding ransom demand note or contact details to receive this information from the attackers.
- **Exclusion criteria:** Attempted intrusions without encryption, incidents targeting individuals, and cases without a ransom demand or contact details.
- **Deduplication:** Duplicate records referring to the same incident are identified using victim identifiers, timestamps, and contextual descriptions, and merged into a single case.
- **Missing data:** Variables are coded as missing (*NA*) when no explicit information is available.

A.2 Variable Definitions and Coding Rules

1a. Revenue (log).

- **Definition:** Estimated yearly revenue of the victim organization.
- **Coding:** Recorded as numerical value where available; $\log_{10}$ transformation applied in analysis.
- **Criteria:** Based on victim business information from the Dutch Chamber of Commerce, as available in the Dutch police information system. For multinational organizations, total estimated annual revenue is used.

1b. RaaS affiliation.

- **Definition:** Whether the ransomware group is identified as operating under a ransomware-as-a-service (RaaS) model.
- **Coding:** 1 = known RaaS group; 0 = non-RaaS or unknown.
- **Criteria:** A group is coded as RaaS only when explicit evidence of affiliate recruitment is observed, such as public advertisements on darknet forums, leak sites, or ransomware group blogs. Secondary sources are used only when they provide verifiable evidence (e.g., screenshots) of such advertisements, in line with [52].

1c–1e. Backup status.

- **Definition:** Availability and usability of backups at the time of the incident.
- **Coding:**
  - 0 = no backups
  - 1 = backups destroyed/unusable
  - 2 = partially recoverable backups
  - 3 = fully recoverable backups
- **Criteria:** Based on explicit statements regarding backup functionality.

1f. Data exfiltration.

- **Definition:** Whether data was exfiltrated during the attack.
- **Coding:** 1 = yes; 0 = no.
- **Criteria:** Data exfiltration is coded when explicitly evidenced by (i) monitoring logs indicating data transfer, or (ii) exposure of victim data in negotiation chats or on the attacker's public leak site.

1g. Negotiation (attacker).

- **Definition:** Whether the attacker allowed negotiation beyond an initial ransom demand.
- **Coding:** 1 = negotiation possible (e.g., contact details provided); 0 = fixed demand only (as stated in the ransom note).
- **Criteria:** Negotiation is coded as present when the attacker provides a communication channel (e.g., chat link, email address) or explicitly invites interaction. It is coded as absent when only a fixed ransom demand is presented without any means of communication.

2a. Cyber insurance.

- **Definition:** Whether the victim had cyber insurance relevant to ransomware.
- **Coding:** 1 = yes; 0 = no.
- **Criteria:** Based on explicit statements regarding insurance in the Police or IR report.

2b. IR firm involvement.

- **Definition:** Whether a professional incident response firm was involved.
- **Coding:** 1 = yes; 0 = no.
- **Criteria:** Based on either (i) included in the IR dataset, or (ii) explicitly mentioned in Police report.

2c. Negotiation (victim).

- **Definition:** Whether the victim initiated or engaged in communication with the attacker.
- **Coding:** 1 = yes; 0 = no.
- **Criteria:** Based on explicit statements regarding negotiation in the Police or IR report.

3a. Negotiation duration (log).

- **Definition:** Duration of attacker-victim communication.
- **Coding:** Number of days between first and last recorded communication; $\log_{10}$ transformation used in analysis.
- **Criteria:** Based on explicit statements regarding negotiations in the Police or IR report, or negotiations themselves if provided with time stamps.

3b. Initial ransom demand (log).

- **Definition:** Initial ransom demand communicated by the attacker.
- **Coding:** Recorded in EUR where available or converted from USD, Bitcoin, or other currencies using the exchange rate on the date of the incident; $\log_{10}$ transformation applied.
- **Criteria:** Must be explicitly stated in report or provided negotiation chats.

3c. Discount offered.

- **Definition:** Whether the attacker reduced the ransom demand during negotiation.
- **Coding:** 1 = reduction between initial and final demand observed; 0 = no reduction observed.
- **Criteria:** Requires explicit documentation of both initial and final demand.

4. Discount size.

- **Definition:** Size of reduction between initial and final ransom demand.
- **Coding:** Calculated as difference between initial and final demand.
- **Criteria:** Only coded when both demands are available, otherwise coded as missing (*NA*).

5. Payment.

- **Definition:** Whether the victim paid the ransom.
- **Coding:** 1 = payment confirmed; 0 = explicitly not paid.
- **Criteria:** Payment is coded when (i) the victim explicitly reports having paid the ransom, (ii) a (bitcoin) transaction is confirmed by the police or IR firm and attributed to the incident, or (iii) payment is explicitly indicated in the negotiation chat.

6. Post-payment outcomes.

- **Definition:** Outcomes following ransom payment.
- **Coding:**
  - **Decryptor failure:** No key delivered or key non-functional.
  - **Renegotiation:** Additional payment demanded after initial settlement.
  - **Reinfection:** Subsequent ransomware incident affecting same victim.
- **Coding:** Each outcome coded as 1 = observed; 0 = not observed.
- **Criteria:** Must be explicitly described in reports.

B Factors Influencing Negotiations (RQ1): Regression analysis results

Model fit. For the regression analyses of the six-stage negotiation framework, we report the following fit statistics. For M1 (attacker willingness to negotiate, logit) the McFadden's pseudo-R^2 is 0.2438. For M2 (victim willingness to negotiate, logit) the McFadden's pseudo-R^2 is 0.0549. For M3 (discount offered, logit) the McFadden's pseudo-R^2 is 0.2712. For M4 (discount size, OLS) the adjusted R^2 is 0.3028 (with a multiple R^2 of 0.3415). Finally, for M5 (payment decision, logit) the McFadden's pseudo-R^2 is 0.3369.

For logistic models, McFadden's pseudo-R^2 values between approximately 0.20 and 0.40 are generally considered indicative of good model fit in applied settings, particularly for behavioral and decision-making processes. The reported values therefore suggest substantial explanatory power for models capturing attacker behavior, discounting, and payment decisions, while lower fit for victim negotiation initiation reflects greater heterogeneity at earlier stages of the process.

Table 6: Stage-wise regression coefficients (β) and p-values across the first five stages of ransomware negotiation and payment. Cells report estimated coefficients with p-values in parentheses; "." indicates the predictor is not applicable at that stage.

Predictor	M1: Attacker negotiation	M2: Victim negotiation	M3: Discount offered	M4: Discount size	M5: Payment
Structural and attack characteristics (inputs from Stage 1 onward)					
1a. Revenue (log)	0.04 (0.624)	0.30 (0.048)*	0.39 (0.116)	−0.12 (0.001)**	0.44 (0.085)$^{\circ}$
1b. RaaS affiliation	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	0.66 (0.044)*	0.77 (0.102)	0.89 (0.142)	−0.16 (0.026)*	0.10 (0.878)
Stage 1 outcome / later-stage inputs					
1g. Negotiation (attacker)	*Outcome*	.	.	−0.07 (0.537)	0.18 (0.775)
Stage 2: Victim-side preparedness and response					
2a. Cyber insurance	.	0.38 (0.472)	−0.09 (0.885)	−0.11 (0.039)*	0.14 (0.832)
2b. IR firm involvement	.	3.47 (< .001)***	2.04 (0.004)**	−0.04 (0.696)	0.06 (0.945)
2c. Negotiation (victim)	.	*Outcome*	.	0.08 (0.679)	4.80 (0.001)**
Stage 3: Interaction dynamics					
3a. Negotiation duration (log)	.	.	1.22 (< .001)***	0.08 (0.061)$^{\circ}$	−1.63 (0.012)*
3b. Initial ransom demand (log)	.	.	.	0.50 (< .001)***	−0.25 (0.068)$^{\circ}$
3c. Discount offered	.	.	*Outcome*	.	1.41 (0.039)*
Stage 4: Price adjustment					
4. Discount size (log)	.	.	.	*Outcome*	−0.30 (0.343)
Stage 5: Payment					
5. Payment	.	.	.	.	*Outcome*

Notes: Logit models are used for M1–M3 and M5, OLS for M4. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.10$.

C Negotiation Strategy and Discount (RQ2): Ransomware Groups

Negotiation outcomes per ransomware group. Table 7 reports descriptive negotiation outcomes by ransomware group in the transcript dataset (N=237). The table highlights substantial heterogeneity in initial demands, negotiation rates, discounting, and observed payments across groups, but several groups have small sample sizes so group-level percentages should be interpreted cautiously. These descriptives motivate the transcript-based regressions in Section 4.2, where we model discount offering and payment conditional on negotiation length.

Table 7: 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)