Ransomware 2.0: To sell, or not to sell A Game-theoretical Model of Data-selling Ransomware

Zhen Li
zli@albion.edu
Department of Economics and Management
Albion College
Albion, Michigan, USA

Qi Liao*
liao1q@cmich.edu
Department of Computer Science
Central Michigan University
Mt. Pleasant, Michigan, USA

ABSTRACT

Cybercrime such as ransomware denies access to valuable data until a ransom is paid. Recent ransomware attacks on organizations such as hospitals, schools, government agencies and private businesses raise public awareness of the severe impact on the society. In this paper, we propose a hypothetical new revenue model for the ransomware, i.e., selling the stolen data. Through a game-theoretical analysis between attackers and victims, we contribute a novel model to understand the critical decision variables between the traditional ransomware (ransomware 1.0) - demanding ransom only and the new type of ransomware (ransomware 2.0) - selling the data as well as demanding ransom. Both theoretical modeling and simulation studies suggest that in general ransomware 2.0 is more profitable than ransomware 1.0. Common defensive measures that may work to eliminate the financial incentives of ransomware 1.0 may not work on ransomware 2.0, in particular the data backup practice and the never-pay-ransom strategy. Nevertheless, the *uncertainties* created by this new revenue model may affect attackers' reputation and users' willingness-to-pay. In turn, ransomware 2.0 may not *always* increase the profitability of attackers. Another finding of the study suggests that reputation maximization is critical in ransomware 1.0 but not in ransomware 2.0, where attackers should seek imperfect reputation for profit maximization.

CCS CONCEPTS

• **Security and privacy → Network security**; **Malware and its mitigation**; **Economics of security and privacy**.

KEYWORDS

Cyber-security, ransomware 2.0, data selling, game theory, economics

ACM Reference Format:
Zhen Li and Qi Liao. 2020. Ransomware 2.0: To sell, or not to sell A Game-theoretical Model of Data-selling Ransomware. In *The 15th International Conference on Availability, Reliability and Security (ARES 2020), August 25–28, 2020, Virtual Event, Ireland*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3407023.3409196

*Corresponding Author: Dr. Qi Liao, email: liao1q@cmich.edu, Department of Computer Science, Central Michigan University, Mt. Pleasant, MI 48859.

ARES 2020, August 25–28, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-8833-7/20/08…$15.00
https://doi.org/10.1145/3407023.3409196

1 INTRODUCTION

Ransomware as a class of malware has lately appeared as a major cybersecurity threat. The malware affects victims' computers and disables access to system and data files through encryptions, and demands ransom payment for the return of computer functionality and data. Ransomware is believed to be highly lucrative [19]. In 2019, the U.S. was hit by an unprecedented ransomware attacks that impacted at least 113 state and municipal governments and agencies, 764 health care providers, and 89 universities, and 1233 schools. The potential cost of these attacks was estimated at $7.5 billion [16]. In CyberSecurity annual reports, ransomware is listed as one of the top three cyberthreat concerns three years in a row (2017, 2018 and 2019) [8].

There are thousands of different ransomware strains in existence, varying in design and sophistication [4]. The first ransomware attack dates back to 1989 that spread via floppy disks and involved sending money to a post office to pay the ransom [2]. The concept of file-encryption ransomware became known as so called "cryptovirology" in a 1996 IEEE Security & Privacy paper [20]. However, such practice remains relatively uncommon until the mid 2000s [10]. Since then, ransomware has been automated and professionalized. The traditional ransomware relies on encrypting information on the victims' computer to demand ransom payment. Recently, a new version of ransomware was found that is armed with browser and email password-stealing features. While it does encrypt data, it uses a variety of methods to steal credentials in each of the targeted applications [14]. Ransomware attackers have threatened to publicly release stolen data if the victims chose not to respond to their ransom demands[13, 15].

In this paper, we propose a new revenue model for ransomware, i.e., selling the stolen data in addition to demanding ransom. We refer to it as ransomware 2.0 for data-selling ransomware as opposed to ransomware 1.0 for traditional ransomware (demanding ransom only). It is imperative to understand what changes ransomware 2.0 may bring to the ransomware business model. To that end, we conduct game-theoretical modeling of both ransomware 1.0 and 2.0 and study the strategic decision-making by the profit-driven ransom attackers and victims/users. The attacker has both the stolen data and locked files in order to gain profit, either from ransom by victims or from selling data to potential buyers, or both. The best response by the victims is studied with the assumption that decryption is not guaranteed as there have been reports of victims

paying the ransom and not receiving the decryption key [14]. It is even more uncertain whether the attacker will keep the stolen data safe. We derive the profit of the ransomware 2.0 in three cases: the attacker has no reputation, perfect reputation, and imperfect reputation; and compare the profitability of both ransomwares 1.0 and 2.0.

Our model and simulation studies suggest ransomware 2.0 in general is more damaging and can make cybercrimes even more lucrative as selling potentially valuable data generates an additional revenue source to the attackers. The threat of data leakage increases the victims' willingness-to-pay the ransom if the data threat does not negatively affect the value of the locked files to the victims. However, if the market value of the stolen data is limited, and/or if the uncertainty of data leakage reduces the value of the locked files to the victims, the ransomware 2.0 may actually be worse for the attackers. While reputation maximization leads to profit maximization in ransomware 1.0, having a perfect reputation in ransomware 2.0 is not necessarily profit maximizing.

The contribution of this work lies in the novel ransomware 2.0 model. To the best of authors' knowledge, we build the first game-theoretical ransomware model with data selling as an additional revenue source. Contrary to common belief, ransomware 2.0 may not always be more profitable than ransomware 1.0 due to the uncertainties. Not trying to be reputable may bring more profit in ransom business is another counterintuitive findings of our study. This paper explores the effects of numerous important factors on the profit of the new data-selling ransomware. The game-theoretical analysis provides insights in designing defensive measures against ever evolving malware and ransomware business.

2 RELATED WORKS

Ransomware has recently taken center stage as one of the most prevalent cybercrimes. Various reports demonstrate the enormous burden placed on individuals and institutions [19]. Given the significant growth of ransomware attacks, it is important to develop a prevention and protection mechanism. Researchers have conducted a survey on ransomware taxonomy and countermeasures [1]. Like any malware, technical mechanisms to defend against ransomware attacks are on the front line. For example, file system activities may be monitored for I/O requests and Master File Table may be protected to detect zero-day ransomware attacks [11].

In addition to technical approaches, there has been recent research that uses economics and game theory to study ransomware behavior. Economic analysis of ransomware [9] reveals the relationship between the valuation distribution among the population and the optimal ransom demand. The study examines the impact of different price discrimination strategies which can help in estimating an optimal ransom value. Since ransom payments are often in the form of Bitcoins, data collected from Bitcoin transactions at public blockchain suggests that the market for ransomware payments has a minimum worth of USD 12,768,536 (22 967.54 BTC) from 2013 to mid-2017 [17].

Game-theoretical model of the ransomware ecosystem [12] was first developed with emphasis on the decision of companies to invest in backup technologies and which degree backup investments can serve as a deterrent for ongoing attacks. Using a game theory to model the strategic playing by ransomware criminals and victims, researchers can understand potential prevention measures and further investigate similar types of cybercrime [5].

Study of the role of reputation suggests that it is optimal for the criminal to build a good reputation and always return the files [6]. How victims form beliefs influences the victims' intention to pay the ransom. A trust model shows that the trust in the attacker and reasonable ransomware demands positively influence the victims' intention to pay the ransom [21].

While kidnapping and blackmail are typically in a terrorist context [18], ransomware may be modeled as kidnapping. The kidnapping aspect of ransomware was acknowledged at a practical level and the models of hostage were extended to study the role of irrational aggression and crime deterrence [7]. The game theoretic literature on kidnapping and blackmail gives insight on the optimal ransom that criminals should charge and the role of deterrence through preventative measures.

Our work is in line with the economics and game theoretic research on ransomware. This paper is the first study on the new type of ransomware that utilizes the stolen data as either a threat for victims to pay ransom or an asset for attackers to manipulate. We propose an additional revenue for ransomware by selling the valuable data. Our model emphasizes on the profitability of the data-selling ransomware compared to traditional ransomware with varying reputations of the attacker. The findings of this study give insights to help the development of defensive measures against this new ransomware. Notably, common advice of nearly all ransomware literature is a mitigation such as backup technologies [12]. While sufficient data backup has the potential to deter traditional ransomware, it has little effect on the new proposed ransomware model which also sells the stolen data.

3 GAME THEORETIC ANALYSIS OF DATA-SELLING RANSOMWARE

In this section, we first lay out the backgrounds and assumptions to specify the ransomware attacks that will be analyzed. We then develop the game theoretic models in three cases of varying reputation of the attacker. We compare the profit of the data-selling ransomware (2.0) with that of traditional ransomware (1.0) in each case.

3.1 Background and Assumptions

While ransomware may be classified into Scareware, Lock-Screen, and Encrypting, the most common form of ransomware is file encryption ransomware [2]. We consider an potential add-on to this type of ransomware that not only files are encrypted but the whole or a subset of data are also transferred to a cloud storage controlled by the attacker. The victims face dual threats: the threat of losing access to files and the threat of leaking data. Hereinafter, we use the phrase "returning files" to refer to the situation where the attacker delivers decryption keys to remove restrictions to a victim's computing resources and files. We use the phrase "selling data" to refer to the situation where the attacker releases the stolen information to a third party.

The attacker has numerous ways to release the data: to release the data to public for free, to sell the data for revenue, or to keep the

data confidential (do nothing). We assume the attacker is money driven so that the attacker will sell the data if doing so is more profitable. As seen from past ransomware attacks, we assume there is no negotiation or bargaining opportunity. Once hit, the victims face two options: pay the ransom demand or do not pay. If the attacker does not return the files, then all encrypted files are going to be lost for good.

There is a cost of returning files and/or selling them. The cost of returning files may include the cost of delivering the decryption key to the victims and the cost of guiding the victims on how to recover files and dealing with queries about files that fail to recover. The cost of selling data includes the search for potential buyers, delivering channels, and other costs of data-related transactions. In addition, the current underground ransomware practice involving cryptocurrencies via distributed blockchain technologies suggests that the probability of facing punishment for a ransomware attack is very low across legal jurisdictions.

3.2 Timeline and Payoff Matrix

The ransomware game is a sequential, multi-stage game involving the attacker and the victims. The timeline of the game is as follows. Stage 1: The attacker launches a successful ransomware attack on N victims. This is the starting point of the game. The infected machines lose access to files and get confidential data stolen. The attacker demands a ransom payment R, which the victims take as given. Stage 2: After observing R, the victims decide whether to pay the ransom or not to pay it. This stage is the victims' decision-making on the ransom payment. Stage 3: Upon observing the victims' decision on ransom payment, the attacker chooses whether to return files to the victims. Stage 4: The attacker determines what to do with the stolen data, to sell it or do nothing. Both Stages 3 and 4 are the attacker's follow-up decision-making.

Let p be the victim's choice of paying ransom in Stage 2.

$$p = \begin{cases} 0, & \text{Not to pay ransom,} \\ 1, & \text{To pay.} \end{cases} \tag{1}$$

Let r be the attacker's choice of returning files in Stage 3.

$$r = \begin{cases} 0, & \text{Not to return files to the victims,} \\ 1, & \text{To return.} \end{cases} \tag{2}$$

Let s be the attacker's choice of selling data in Stage 4.

$$s = \begin{cases} 0, & \text{Not to sell data,} \\ 1, & \text{To sell.} \end{cases} \tag{3}$$

Consider a representative victim i. The payoff (profit) the attacker expects to receive from victim i is

$$\pi = pR + sA_i - rC_r \tag{4}$$

where $C_r > 0$ is the cost of returning files to the victims. $C_d > 0$ is the data transaction cost. $D_i \geq 0$ is the market value of the data stolen from victim i. We define A_i as the data profit of the attacker where

$$A_i = \begin{cases} D_i - C_d, & \text{if } D_i \geq C_d, \\ 0, & \text{if } D_i < C_d. \end{cases} \tag{5}$$

The payoff (utility) of victim i is

$$u = -pR - (1 - r)V_{r,i} - sL_{d,i} \tag{6}$$

where $V_{r,i} \geq 0$ is the value of the locked files to the victims. $L_{d,i} \geq 0$ is the loss to the victims if the stolen data is sold.

The key difference between the data-selling ransomware and traditional ransomware is the existence of the stolen data. Numerous questions arise. Will ransomware be more profitable with the new feature? Will the victims change their willingness-to-pay the ransom? Will the attacker keep the stolen data confidential? If the victims do not expect the attacker to keep the data safe, why should they pay the ransom? To address these questions, we need to compare the data-selling ransomware to traditional ransomware and show the difference between game outcomes and payoffs.

Since p and r are binary decision variables, the game of traditional ransomware has four possible outcomes. In data-selling ransomware, the strategy variables (p, r and s) suggest that the game of data-selling ransomware has eight possible outcomes. The attacker's and victim i's payoffs to different outcomes are shown in Table 1. The goals of both the attacker and the victims are to maximize their expected payoffs, which depend on the game outcomes.

In the ransomware game, the victims are in a disadvantageous position. As Table 1 shows, the best possible outcome for the victims is to receive a zero payoff. This would be the case if the attacker returned files for free, and would not sell the stolen data. In all the other cases, the victims suffer a negative payoff.

3.3 The Baseline Case: Non-repeated Game with No Trust

As a baseline case, we model a one-shot game with no need for the attacker to build reputation. The attacker's decision-making in Stages 3 and 4 are independent. Let's derive the game outcome using backward deduction from the last stage of the game, i.e., Stage 4.

Proposition 1: In the baseline model, the attacker sets $s = 1$ if $D_i \geq C_d$ and $s = 0$ if $D_i < C_d$.

The attacker sells the stolen data whenever the market price of the data exceeds the transaction cost. The attacker receives a net gain of $A_i = D_i - C_d$ from the victims whose data values more than the transaction cost in the market. The attacker receives a payoff of $A_i = 0$ from the victims whose data values less than the transaction cost.

Proposition 2: In the baseline model, the attacker sets $r = 0$.

Not returning files to the victims is always the dominant strategy for the attacker regardless of ransom payment. When reputation is irrelevant, the attacker has no incentive to return files.

The victims' ransom payment decision in Stage 2 critically depends on the victims' belief that the attacker will honor the ransom payment. In the baseline model, taking the money and run is the dominant strategy of the attacker. Expecting the attacker to default, the victims will choose not to pay the ransom in Stage 2.

Proposition 3: In the baseline model, the victims set $p = 0$.

Combining Propositions 1 to 3, the baseline model between the attacker and one victim has two possible outcomes: $\{p = 0, r = 0, s = 1\}$ if $D_i \geq C_d$, and $\{p = 0, r = 0, s = 0\}$ if $D_i < C_d$.

The total profit of the attacker to receive from all victims in the baseline model is

$$\Pi_b = \sum_{i=1}^{N} A_i \tag{7}$$

Table 1: The payoffs to different outcomes in the data-selling ransomware game

Outcome			Attacker (π)	Victim (u)
$p=0$	$r=0$	$s=0$	0	$-V_{r,i}$
$p=0$	$r=0$	$s=1$	$D_i - C_d$	$-V_{r,i} - L_{d,i}$
$p=0$	$r=1$	$s=0$	$-C_r$	0
$p=0$	$r=1$	$s=1$	$D_i - C_d - C_r$	$-L_{d,i}$
$p=1$	$r=0$	$s=0$	R	$-R - V_{r,i}$
$p=1$	$r=0$	$s=1$	$R + D_i - C_d$	$-R - V_{r,i} - L_{d,i}$
$p=1$	$r=1$	$s=0$	$R - C_r$	$-R$
$p=1$	$r=1$	$s=1$	$R + D_i - C_d - C_r$	$-R - L_{d,i}$

Victim i's utility is $-V_{r,i}$ if $D_i < C_d$ and $-V_{r,i} - L_{d_i}$ if $D_i \geq C_d$.

For traditional ransomware, the game outcome of a one-shot model is $\{p = 0, r = 0\}$. The attacker's profit is 0 and victim i's payoff is $-V_{r,i}$. If the attacker's reputation is irrelevant, the data-selling ransomware is more profitable than traditional ransomware. The two strains of ransomware were only equivalent if none of the stolen data were marketable enough, which is not likely to occur.

Therefore, even if the ransom payment is zero, the attacker may still receive financial benefits as long as the market value of the stolen data exceeds the cost of selling data. This is arguably the biggest advantage of the data-selling ransomware (2.0) compared to traditional ransomware (1.0). Thus some defensive measures that may work to eliminate the financial incentives of ransomware 1.0 may not work on ransomware 2.0, in particular the data backup practice and the never-pay-ransom strategy.

Data backup has been widely considered the most effective strategy to mitigate the loss of ransomware [2, 3]. Having a comprehensive data backup process may effectively protect the victims from the threat of traditional ransomware. The victims could simply ignore the ransom note and have a fresh start with the backed-up files. Data backup, however, will not work as effectively against the data-selling ransomware. The victims are exposed to the risk of data leakage. Even if the files are fully backed up, the attacker may gain from selling the valuable data. Data backup will not eliminate the financial incentives of the data-selling ransomware.

Similarly, the never-pay-ransom strategy may work for traditional ransomware since if no one pays, ransomware will become unprofitable. Therefore, a practical strategy for the victims of traditional ransomware is always to say no to the attacker. However, the never-pay-ransom strategy would not work for the data-selling ransomware because attackers can almost always gain from selling data. The never-pay-ransom strategy does not remove financial incentives of the new ransomware.

In both cases of data backup and never-pay-ransom, the profit of traditional ransomware is zero with no ransom payment, but the profit of the data-selling ransomware can be positive. Nevertheless, it does not imply the data-selling ransomware is always more profitable than traditional ransomware. The equilibrium outcome of the baseline model is not optimal for neither the attacker nor the victims. If there were trust, the victims could benefit from paying the ransom for any $R \leq V_r$. The attacker could benefit from returning files and keeping data confidential for any $R \geq C_r$. Since the value of files to the victims is highly likely to exceed the attacker's cost of returning files, there exists a range of ransom $R \in (C_r, V_r)$ that can be mutually beneficial. The attacker would be better off receiving a ransom higher than the cost of returning files. The victims would be better off to pay a ransom in exchange for the files that value more than the ransom. However, this "win-win" (when compared to the baseline equilibrium outcome) situation requires cooperation of the two parties and the victims to trust the attacker. The attacker cannot ignore reputation if ransomware is to be a sustainable business model.

3.4 Role of Reputation: A Cooperative Game with Perfect Reputation

Reputation matters when the outcome of the game between the attacker and one victim affects the choice of other or future victims. It can be in the attacker's interest to build up a reputation because any short-term gain from taking the money and run may be offset by the unwillingness of other victims to pay any ransom.

Proposition 4: In the perfect reputation model, the attacker sets $r = 1$ and $s = 0$ if ransom is paid; the attacker sets $r = 0$ and $s = 1$ if ransom is not paid and $D_i \geq C_d$.

To illustrate the role of reputation, suppose the attacker had endowed reputation who would honor the agreement with the victims with no need to be self-enforcing. The strategy the attacker shall follow, in response to the victims' choice, would be straightforward: to return files and keep the stolen data confidential if the ransom is paid; or to delete the files and sell the data if the ransom is not paid.

Proposition 5: In the perfect reputation model, victim i sets $p = 0$ if $R > V_{r,i} + L_{d,i}$ and $p = 1$ if $R \leq V_{r,i} + L_{d,i}$.

When the victims trust the attacker, the victims' willingness-to-pay the ransomware is $V_{r,i} + L_{d,i}$. By paying the ransom, the victims avoid the file loss and the data loss.

Suppose there are n victims who set $p = 1$. The profit of the attacker is

$$\Pi_t = n(R - C_r) + \sum_{i=n+1}^{N} A_i \tag{8}$$

For traditional ransomware in the case of perfect reputation, the victims' willingness-to-pay the ransom is capped by $V_{r,i}$ and the profit of the attacker is $n(R - C_r)$. Recall the victims' willingness-to-pay the ransom is 0 in the baseline case with no trust. Building reputation can be rewarding to the attacker by increasing the victims' willingness-to-pay the ransom.

The attacker of the data-selling ransomware receives the same profit from the n victims who choose to pay the ransom. For the victims who choose not to pay the ransom, the attacker's profit increases for the data-selling ransomware, compared to 0 profit of traditional ransomware. Compared to traditional ransomware, the data-selling ransomware is more profitable.

In summary, if the attacker has perfect reputation, the data-selling ransomware is more profitable than traditional ransomware. However, it is difficult for the attacker to build perfect reputation in the underground economy. In reality, although many victims who do not pay the ransom may end up losing their files, victims who do pay may not necessarily retrieve their files. Recent evidence suggests that in 2019 about 60% victims who pay the ransom recovered their files [8]. Next we extend the model to a competitive setting, and examine how the data-selling feature of ransomware may add extra uncertainty to an already risky environment.

3.5 A General Competitive Ransomware Game with Imperfect Reputation

The victims' willingness-to-pay ransom depends on the attacker's reputation. The victims estimate the credibility of the attacker based on the past records of the attacker regarding delivering decryption keys and keeping the stolen data safe, e.g., crawling personal and social networks, forums, search engines, media reports, etc. Suppose past records of the attacker indicate that the attacker has $\beta_r \in [0, 1]$ percentage of the chance to return files with ransom payment and $\beta_d \in [0, 1]$ percentage of the chance to keep the stolen data confidential with ransom payment.

A representative victim's expected utility in the risky environment is

$$u_u = -pR - (1 - p\beta_r)V_r - (1 - p\beta_d)L_d \tag{9}$$

From Equation (9), the victim receives a payoff of $-V_r - L_d$ if not paying ransom ($p = 0, \beta_r = 0, \beta_d = 0$). The victim's expected utility is $-R - (1 - \beta_r)V_r - (1 - \beta_d)L_d$ if paying ($p = 1$). Apparently, the victims will choose to pay the ransom if doing so generates a higher expected payoff, i.e., $p = 1$ if $-R-(1-\beta_r)V_r-(1-\beta_d)L_d \geq -V_r-L_d$. That leads to Proposition 6.

Proposition 6. In the competitive game, the victims will choose to pay ransom if $R \leq \beta_r V_r + \beta_d L_d$.

Proposition 6 specifies the victims' willingness-to-pay in the imperfect reputation case. There are two parts of the victims' willingness-to-pay, the expected value of the locked files and the expected value of the stolen data. The no-reputation case and the perfect-reputation case are two special cases of the general expression: $\beta_r = \beta_d = 0$ for the former and $\beta_r = \beta_d = 1$ for the latter. The reputation of the attacker increases the victims' willingness-to-pay.

The attacker's profit with one victim is

$$\pi_u = pR - p\beta_r C_r + (1 - p\beta_d)A_i \tag{10}$$

From Equation (10), the attacker will receive a profit of A_i from a victim if ransom not paid, and a profit of $R - \beta_r C_r + (1 - \beta_d)A_i$ if ransom paid.

Suppose n victims choose to pay the ransom, the expected profit of the attacker among all victims is

$$\Pi_u = n(R - \beta_r C_r) + \sum_{i=1}^{n}(1 - \beta_d)A_i + \sum_{i=n+1}^{N} A_i \tag{11}$$

Proposition 7. In the competitive game, the attacker sets $\beta_d = 0$ if $\beta_d L_d \leq A_i$ and $\beta_d = 1$ otherwise.

$\beta_d L_d$ is the upper-bound on the potential increase in ransom demand with data threat. If the expected ransom gain is no higher than the profit of selling data, the attacker chooses to sell data. Suppose the condition $\beta_d L_d \leq A_i$ holds true for m out of N victims, the attacker has the likelihood of $\beta_d = 1 - m/N$ to keep the stolen data confidential for a random victim.

In the baseline model, it is optimal for the attacker not to return files. In the cooperative game, the attacker should always return the files with ransom payment. When the game is competitive with imperfect reputation, it may not be optimal to always return files with ransom payment or never to return.

Proposition 8. In the competitive game, the attacker shall return files with ransom payment if $\beta_r V_r \geq C_r$.

Comparing the profit of ransomware in the cooperative game and the competitive game, as in Equations (8) and (11), it is ambiguous which is more profitable. The attacker faces dual tradeoffs. The first is common to ransomware: the tradeoff between building reputation and gaining from defaulting. The second is unique to the data-selling ransomware: the tradeoff between ransom demand and the revenue from selling data.

For example, suppose the number of victims who are willing to pay the ransom is the same in the two games, i.e., the two n's in Equations (8) and (11) take the same value. The ransom demand is R_t in the perfect reputation game and R_u in the competitive game. Then

$$\Pi_u - \Pi_t = n\{(1 - \beta_r)C_r - (R_t - R_u)\} + \sum_{i=1}^{n}(1 - \beta_d)A_i \tag{12}$$

In the competitive game with imperfect reputation, the attacker may gain from the saved cost of returning files ($(1 - \beta_r)C_r$) and selling the stolen data ($(1 - \beta_d)A_i$). The sacrifice is a potential loss in ransom ($R_t - R_u$). The data-selling component of ransomware adds uncertainty to the competitive game. It not only strengthens the existing tradeoff, it also adds a new layer of tradeoff to the game, applicable to both the attacker and the victims.

4 SIMULATION RESULTS

We compare the profit of the data-selling ransomware to traditional ransomware with simulation experiments in three cases discussed in Section 3: the baseline game model with no reputation, the cooperative game model with perfect reputation, and a general competitive game model with imperfect reputation.

Table 2: The comparison of profit between the traditional ransomware and the data-selling ransomware

various cases of reputation	traditional ransomware	data-selling ransomware
$\beta_r = \beta_d = 0$	0	$\sum_{i=1}^{N} A_i$
$\beta_r = \beta_d = 1$	$n(R - C_r)$	$n(R - C_r) + \sum_{i=n+1}^{N} A_i$
$0 < \beta_r < 1, 0 < \beta_d < 1$	$n(R - \beta_r C_r)$	$n(R - \beta_r C_r) + \sum_{i=1}^{n}(1 - \beta_d)A_i + \sum_{i=n+1}^{N} A_i$

4.1 Simulation Setup

The profit formulas of traditional ransomware and the data-selling ransomware in the three cases are in Table 2 where n is the number of victims choosing to pay the ransom. It varies from case to case.

Suppose there are $N = 30$ victims, and the ransom demand be $R = 50$. The victims' valuation of the locked files (V_r) and the stolen data (L_d) are randomly generated in the range from 0 to 100. Without loss of generality, we set the cost of returning files at $C_r = 5$, the cost of selling data at $C_d = 10$, and $D_i = L_{d,i}$.

4.2 The comparison of profit in the no-reputation and perfect-reputation cases

At the specified parameters and randomly generated values of V_r and L_d, the profit of the data-selling ransomware in the case of no reputation ($\beta_r = \beta_d = 0$) is 1,018 (earned from selling data), compared to 0 for traditional ransomware. In the case of perfect reputation ($\beta_r = \beta_d = 1$), a victim chooses to pay the ransom if $R \leq V_r$ for traditional ransomware. The simulation results show that 15 victims choose to pay, generating a profit of 675 at per-victim profit of 45. For the data-selling ransomware, a victim chooses to pay the ransom if $R \leq V_r + L_d$. The simulation results show that 25 victims choose to pay, generating a ransom profit of 1,125. Meanwhile, the attacker receives an additional profit of 30 from selling the data of the 5 victims who do not pay the ransom, bringing the profit of the data-selling ransomware to a total of 1,155.

Therefore, the data-selling ransomware is more profitable than traditional ransomware in both the no-reputation case and the perfect-reputation case. The increase in profit comes from the increased number of victims paying the ransom and the additional revenue from selling the stolen data.

4.3 Profits in the imperfect-reputation case

In the imperfect-reputation case, the victims' willingness-to-pay is capped at $\beta_r V_r + \beta_d L_d$. Given the victims' valuation of the locked files and the stolen data, the attacker's choices of returning files (β_r) and selling data ($1 - \beta_d$) determine the number of victims choosing to pay the ransom. The attacker faces a tradeoff between ransom income and data income when setting β_r and β_d. If the attacker sets higher probabilities of returning files and keeping the data safe, the attacker will gain from increased ransom payments but lose from forgone data income.

4.3.1 How selling data affects ransomware profit. We first study how the probability of selling data affects the ransomware profit at various probability of returning files. The simulation results suggest the tradeoff that the attacker faces when setting β_r and β_d, as in Figure 1. There are five data series in the figure. The two flat lines are the data-selling ransomware profit in the no-reputation and perfect-reputation cases for reference. The other three curves illustrate how the data-selling ransomware profit changes when the probability of selling the stolen data changes, given a certain probability of returning files (β_r).

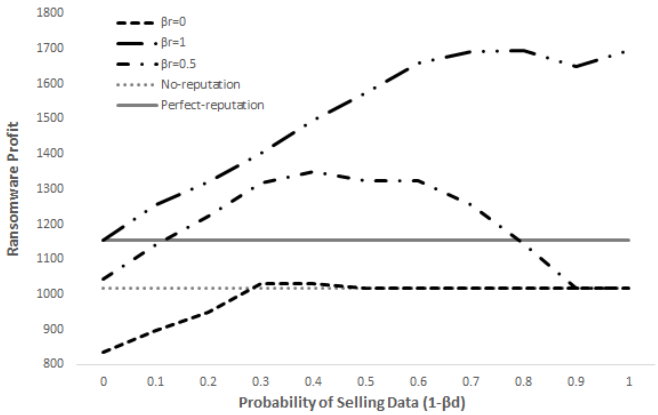

Figure 1: Profitability of the data-selling ransomware changes with the probability of selling data at various probability of returning files. A low probability of returning files and a high probability of selling data decrease the victims' willingness-to-pay. As the probability of selling data increases, fewer victims pay the ransom but the data revenue increases. The net change in ransomware profit depends on the relative changes in ransom profit and data-selling profit.

Because of the tradeoff between ransom revenue and data revenue, none of the three curves is monotonic. Increasing the probability of selling data is not necessarily profit increasing because it decreases the victims' willingness-to-pay the ransom. Since a lower β_r also decreases the victims' willingness-to-pay, the profit-maximizing probability of selling data appears to be at a low or moderate level when β_r is smaller. When β_r is big, a higher probability of selling data tends to be more profitable because a high β_r helps maintain the victims' willingness-to-pay the ransom while the attacker gains additionally from selling data.

4.3.2 How returning files affects ransomware profit. Now we study the effects of the file-returning probability on ransomware profit at various data-selling probabilities, as shown in Figure 2. The two flat lines are the data-selling ransomware profit in the no-reputation and perfect-reputation cases for reference. The other three curves illustrate how the data-selling ransomware profit changes when the

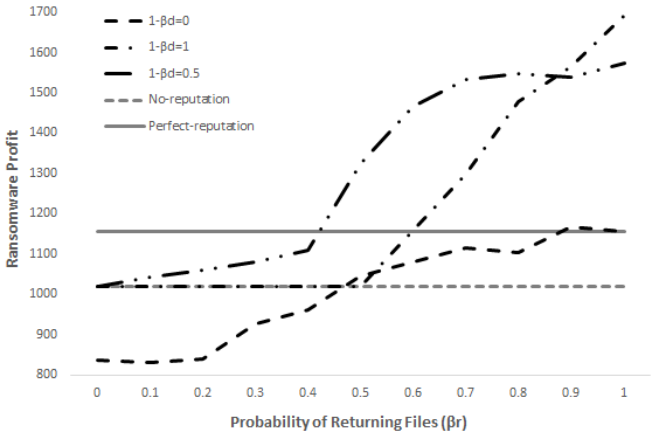

Figure 2: Overall, profitability of the data-selling ransomware increases with higher probability of returning files resulting in more victims paying the ransom. Selling data at (0.5) rate performs better than no selling (0) or always selling (1). Attacker's optimal strategy is a mixed strategy with a combinations of returning files and selling data.

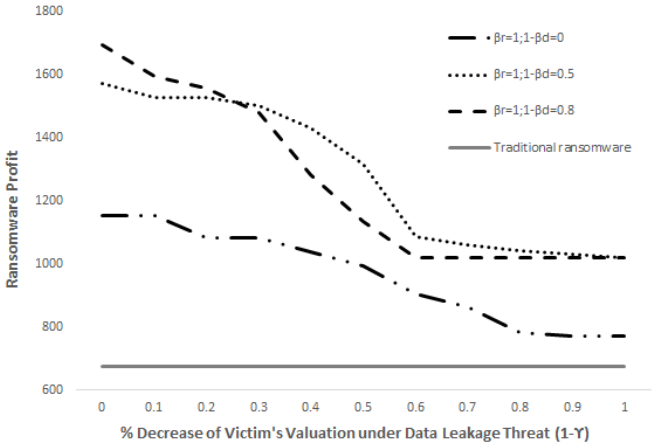

Figure 3: Profitability of the data-selling ransomware decreases as victims' valuation of their locked files decreases under the data leakage threat at various data-selling probabilities. In the case when average market value matches average victims' expected value of their locked data, the data-selling ransomware is always more profitable than traditional ransomware.

probability of returning files changes, given a certain probability of selling data $(1 - \beta_d)$.

The results confirm the tradeoff the attacker faces when setting β_r and β_d. Increasing the probability of returning files increases the victims' willingness-to-pay the ransom, generating more ransom income, potentially causing a loss in data profit. The probability of selling data for the victims who do not pay the ransom is 1, but the probability of selling data for the victims who pay the ransom is $1-\beta_d$. As more victims pay the ransom, the data profit decreases but not by as much. Although there are fluctuations, overall the data-selling ransomware is more profitable when the attacker increases the probability of returning files, at a given probability of selling data.

Based on the above results, we summarize that data-selling ransomware is always more profitable than traditional ransomware in both no-reputation and perfect-reputation models. For traditional ransomware, it is profit maximizing to build perfect reputation by always returning the data files. Building perfect reputation is not necessarily profit maximizing for the data-selling ransomware because the attacker faces a tradeoff between gaining from ransom and gaining from selling data. The relative profit of ransomware in the imperfect-reputation case is nondeterministic, as shown in Figures 1 and 2. It implies that the optimal strategy of the attacker is a mixed strategy with certain combinations of β_r and β_d, in accordance with the victims' valuation of locked files and stolen data.

4.4 Profit under data leakage threat

Under the threat of data leakage as in data-selling ransomware, victims may or may not value the locked files as much as in traditional ransomware. This may inversely affect the victims' willingness-to-pay. For example, a leaked customer database becomes less valuable to the victims since that means mandatory resetting passwords for all customers or closing accounts. The decreasing victims' willingness-to-pay has a potential to negatively affect the relative profit of the data-selling ransomware.

When factoring in the plausible negative effect of data threat on the value of locked files, the leftover value of the files is a fraction of the data-threat-free value of the files, γV_r where $\gamma \in [0, 1]$. A representative victim's expected utility is

$$u_u = -pR - (1 - p\beta_r)\gamma V_r - (1 - p\beta_d)L_d \tag{13}$$

From Equation (13), the victim receives a payoff of $-\gamma V_r - L_d$ if not paying ransom ($p = 0$, $\beta_r = 0$ and $\beta_d = 0$). The victim's expected utility is $-R-(1-\beta_r)\gamma V_r-(1-\beta_d)L_d$ if paying ($p = 1$). The victims will choose to pay if doing so generates a higher expected payoff, i.e., if $\beta_r\gamma V_r + \beta_d L_d \geq R$.

4.4.1 Case 1: average market value matches average victims' expected value. We study how γ affects the profit of the data-selling ransomware with the same randomly generated values of V_r and L_d as above. During simulation, V_r and L_d are drawn from the same range between 0 and 100. Let ransom demand be 50 and $\beta_r = 1$. The profit of the data-selling ransomware remains at 1,018 in the no-reputation case, regardless of γ as the attacker profits only from selling the stolen data. In the perfect-reputation case, the victims' willingness-to-pay is $\gamma V_r + L_d$. We let γ to vary from 0 to 1 to calculate the profit of the data-selling ransomware.

Figure 3 shows the results. The flat line is the profit of traditional ransomware in the perfect-reputation case for reference. The other three curves are the profit of the data-selling ransomware at various probabilities of selling data. The general trend of profitability of the data-selling ransomware is decreasing as the victims' valuation of

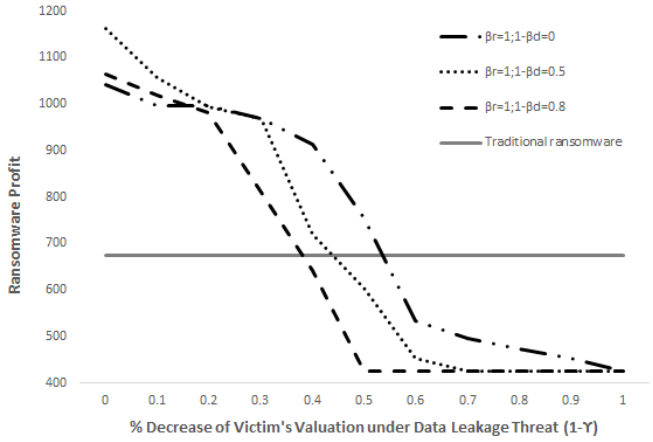

Figure 4: Profitability of the data-selling ransomware decreases as victims' valuation of their locked files decreases under the data leakage threat at various data-selling probabilities. In the case when average market value is less than average victims' expected value of their locked data, the data-selling ransomware may be less profitable than traditional ransomware.

their data decreases. At any given $1 - \gamma$, not selling data is the least profitable because the attacker would not be able to compensate as much the lost ransom income from selling the stolen data. While not selling data performs the worst, selling at a higher rate does not necessarily mean more profitable than selling at a lower rate.

Also shown in Figure 3, the data-selling ransomware stays more profitable than traditional ransomware since even if selling data completely wipes off victims' valuation on the locked files, the attacker can still profit no less from the stolen data.

4.4.2 Case 2: average market value is less than average victims' expected value. In this simulation, we keep $\beta_r = 1$, $R = 50$, and set the average market value at 50% of the average victims' expected value on their locked data. In an analogy of housing market, a house's market value may be \$200,000 but the owner's expected value may be \$400,000 due to affection.

Figure 4 shows the data-selling ransomware profit exhibits a similar trend of decreasing profit as victims' expected valuation decreases as in Figure 3. However, not selling data generally performs better than the other two curves. Another interesting result is that in this case data-selling ransomware is not always more profitable than traditional ransomware (the middle flat line). The above result suggests that using the stolen data as additional threat to force the victims to cooperate may back fire when the potential data-selling profit is limited. If the data is not valuable enough and the data leakage threat reduces the victims' valuation of their locked files, the data-selling ransomware is less profitable than traditional ransomware.

5 CONCLUSION

In this paper we studied a new type of ransomware that gains potential profit by selling stolen data in addition to ransom demand. The game-theoretical models we built analyze the best strategies of both the attacker and the victims in various cases, i.e., baseline game with no reputation, cooperative game with perfect reputation, and the general competitive game with imperfect reputation. The modeling analysis and simulation studies suggest that the data-selling ransomware is more financially rewarding than traditional ransomware in most cases. However, the realization of the potential financial gains largely depends on the marketability of the stolen data and whether and how the threat of data leakage affects the victims' willingness-to-pay ransom. In this sense, the data-selling ransomware is more risky to both the attacker and the victims. Having established reputation is mutually beneficial to both the attacker and the victims, but having perfect reputation is not necessarily profit-maximizing for the attacker of the data-selling ransomware. The finding suggests that the attacker may play strategically with combinations of unlocking and selling data, and manipulate the perception of the victims to gain profit.

REFERENCES

[1] Bander Ali Saleh Al-rimy, Mohd Aizaini Maarof, and Syed Zainudeen Mohd Shaid. 2018. Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. *Computer & Security* 74 (May 2018), 144–166.
[2] Azad Ali. 2017. Ransomware: A Research and a Personal Case Study of Dealing with this Nasty Malware. *Issues in Informing Science and Information Technology* 14 (2017), 87–99.
[3] Mihail Anghel and Andrei Racautanu. 2019. A note on different types of ransomware attacks. *IACR Cryptology ePrint Archive* (2019), 605.
[4] Pranshu Bajpai, Aditya K. Sood, and Richard Enbody. 2018. A key-management-based taxonomy for ransomware. In *Proceedings of APWG Symposium on Electronic Crime Research*. San Diego, CA, 1–12.
[5] Nicholas Caporusso, Singhtararaksme Chea, and Raied Abukhaled. 2018. A Game-Theoretical Model of Ransomware. In *Proceedings of the International Conference on Applied Human Factors and Ergonomics*. Orlando, FL, 69–78.
[6] Anna Cartwright and Edward Cartwright. 2019. Ransomware and Reputation. *Games, MDPI, Open Access Journal* 10(2) (June 2019), 1–14.
[7] Edward J. Cartwright, Julio Hernandez-Castro, and Anna Cartwright. 2019. To pay or not: game theoretic models of ransomware. *Journal of Cybersecurity* 5 (2019), 1–12.
[8] CyberEdge. 2019. 2019 Cyberthreat Defense Report. (2019).
[9] Julio Hernandez-Castro, Edward Cartwright, and Anna Stepanova. 2017. Economic Analysis of Ransomware. *SSRN Electronic Journal* (March 2017). https://doi.org/10.2139/ssrn.2937641
[10] Amin Kharraz, William Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. 2015. Cutting the Gordian Knot: A Look Under the Hood of Ransomware Attacks. In *Proceedings of the 12th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. 3–24.
[11] Amin Kharraz, William Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. 2015. Cutting the Gordian Knot: A Look Under the Hood of Ransomware Attacks. In *Proceedings of the International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA 2015)*. 3–24.
[12] Aron Laszka, Sadegh Farhang, and Jens Grossklags. 2017. On the Economics of Ransomware. In *Proceedings of the 8th Conference on Decision and Game Theory for Security (GameSec 2017)*. 397–417.
[13] Lee Mathews. 2020. Another Ransomware Campaign Threatens To Expose Victims' Data. *Forbes* (January 23 2020).
[14] Cyware Hacker News. 2020. FTCode Ransomware Returns with Credential-Stealing Capabilities. *Cyware* (January 22 2020). https://cyware.com/news/ftcode-ransomware-returns-with-credential-stealing-capabilities-181a6274
[15] Cyware Hacker News. 2020. Ransomware Operators Turn Evil for Late Reposnders and Non-paying Victims. *Cyware* (January 23 2020). https://cyware.com/news/ransomware-operators-turn-evil-for-late-reposnders-and-non-paying-victims-de65c7c1
[16] Cyware Hacker News. 2020. The State of Maryland to Criminalize Ransomware Possession. *Cyware* (January 21 2020). https://cyware.com/news/the-state-of-maryland-to-criminalize-ransomware-possession-4ad3df2f
[17] Masarah Paquet-Clouston, Bernhard Haslhofer, and Benoit Dupont. 2018. Ransomware Payments in the Bitcoin Ecosystem. In *Proceedings of the 17th Annual Workshop on the Economics of Information Security (WEIS)*. Innsbruck, Austria, 10.
[18] Todd Sandler and Daniel G. Arce M. 2003. Terrorism & Game Theory. *Simulation & Gaming* 34(3) (2003), 319–337.

[19] Camelia Simoiu, Christopher Gates, Joseph Bonneau, and Sharad Goel. 2019. "I was told to buy a software or lose my computer. I ignored it": A study of ransomware. In *Proceedings of the Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. Santa Clara, CA, 155–174.

[20] Adam Young and Moti Yung. 1996. Cryptovirology: extortion-based security threats and countermeasures. In *Proceedings of IEEE Symposium on Security and Privacy*. Oakland, CA, 129–140.

[21] Alex Zarifis and Xusen Cheng. 2018. The Impact of Extended Global Ransomware Attacks on Trust: How the Attacker's Competence and Institutional Trust Influence the Decision to Pay. In *Proceedings of the Americas Conference on Information Systems (AMCIS 2018)*. New Orleans, USA.