On the Economics of Ransomware

Aron Laszka[1], Sadegh Farhang[2], and Jens Grossklags[3]

[1] Vanderbilt University
[2] Pennsylvania State University
[3] Technical University of Munich

Abstract. While recognized as a theoretical and practical concept for over 20 years, only now ransomware has taken centerstage as one of the most prevalent cybercrimes. Various reports demonstrate the enormous burden placed on companies, which have to grapple with the ongoing attack waves. At the same time, our strategic understanding of the threat and the adversarial interaction between organizations and cybercriminals perpetrating ransomware attacks is lacking.
In this paper, we develop, to the best of our knowledge, the first game-theoretic model of the ransomware ecosystem. Our model captures a multi-stage scenario involving organizations from different industry sectors facing a sophisticated ransomware attacker. We place particular emphasis on the decision of companies to invest in backup technologies as part of a contingency plan, and the economic incentives to pay a ransom if impacted by an attack. We further study to which degree comprehensive industry-wide backup investments can serve as a deterrent for ongoing attacks.

Keywords: Ransomware, Backups, Security Economics, Game Theory

1 Introduction

Already in 1996, Young and Yung coined the term *cryptovirological attacks* and provided a proof-of-concept implementation of what could now be considered a major building block of ransomware malware [34]. Due to the perceived seriousness of this attack approach, they also suggested that "access to cryptographic tools should be well controlled."

Malware featuring ransomware behavior was at first deployed at modest scale (e.g., variants of PGPCoder/GPCode between approximately 2005-2010), and often suffered from technical weaknesses, which even led a researcher in the field to proclaim that "ransomware as a mass extortion mean is certainly doomed to failure" [11]. However, later versions of GPCode already used 1024-bit RSA key encryption; a serious threat even for well-funded organizations.

Ransomware came to widespread prominence with the CryptoLocker attack in 2013, which utilized Bitcoin as a payment vehicle [20]. Since then, the rise of ransomware has been dramatic, culminating (so far) with the 2017 attack waves of the many variants of the WannaCrypt/WannaCry and the Petya ransomwares. Targets include all economic sectors and devices ranging from desktop computers, entire business networks, industrial facilities, and also mobile

devices. Security industry as well as law enforcement estimates for the amount of money successfully extorted and the (very likely much larger) overall damage caused by ransomware attacks differ widely. However, the figures are significant (see Section 5). Observing these developments, in a very recent retrospective article, Young and Yung bemoan the lack of adequate response focused on ransomware attacks by all stakeholders even though the threat was known for over 20 years [35].

As with any security management decision, there is a choice between doing nothing to address a threat, or selecting an appropriate investment level. In the case of responding to a sophisticated ransomware attack, this primarily concerns decisions on how to invest in backup and recovery technologies, and whether to pay a ransom in case of a successful attack. These decisions are interrelated.

The empirical evidence is mixed (and scarce). It is probably fair to say that backup technologies have always been somewhat of a stepchild in the overall portfolio of security technologies. In 2001, a survey showed that only 41% of the respondents did data backups and 69% had not recently facilitated a backup; at the same time, 25% reported to have lost data [7]. In 2009, another survey found backup usage of less than 50%; and 66% reported to have lost files (42% within the last 12 months) [15]. In a backup awareness survey that has been repeated annually since 2008, the figures for individuals who *never* created backups have been slowly improving. Starting at 38% in 2008, in the most recent survey in June 2017 only 21% reported to have never made a backup. Still, only 37% now report to create at least monthly backups [3], despite the heightened media attention given to ransomware.

Regarding ransom payment behavior, IBM surveyed 600 business leaders in the U.S about ransomware, and their data management practices and perceptions. Within their sample, almost 50% of the business representatives reported ransomware attacks in their organizations. Interestingly, 70% of these executives reported that ransom payments were made in order to attempt a resolution of the incident. About 50% paid over $10,000 and around 20% reported money transfers of over $40,000 [14]. In contrast, a different survey of ransomware-response practices found that 96% of those affected (over the last 12 months) did *not* pay a ransom [18]. However, the characteristics of the latter sample are not described [18]. Finally, recent cybercrime measurement studies have tracked the approximate earnings for particular ransomware campaigns (for example, by tracking related Bitcoin wallets). These studies typically do not succeed in pinpointing the percentage of affected individuals or organizations paying the ransom (e.g., [17,20]).

Our work targets two key aspects of a principled response to sophisticated ransomware attacks. First, we develop an economic model to further our strategic understanding of the adversarial interaction between organizations attacked by ransomware and ransomware attackers. As far as we know, our work is the first such game-theoretic model.

Second, we study the aforementioned response approaches to diminish the economic impact of the ransomware threat on organizations. As such our model

focuses on organizations' decision-making regarding backup investments (as part of an overall contingency plan), which is an understudied subject area. We further determine how backup security investments interact with an organization's willingness to pay a ransom in case of a ransomware attack.

Further, we numerically show how (coordinated) backup investments by organizations can have a deterrent effect on ransomware attackers. Since backup investments are a private good and are not subject to technical interdependencies, this observation is novel to the security economics literature and relatively specific to ransomware. Note, for example, that in the context of cyberespionage and data breaches to exfiltrate data, such a deterrence effect of backup investments is unobservable.

We proceed as follows. In Section 2, we develop our game-theoretic model. We conduct a thorough analysis of the model in Section 3. In Section 4, we complement our analytic results with a numerical analysis. We discuss additional related work on ransomware as well as security economics in Section 5, and offer concluding remarks in Section 6.

2   Model

We model ransomware attacks as a multi-stage, multi-defender security game. Table 1 shows a list of the symbols used in our model.

Table 1. List of Symbols

Symbol	Description
G_j	set of organizations belonging to group j
W_j	initial wealth of organizations in group j
β	discounting factor for uncertain future losses
F_j	cost of data loss due to random failures in group j
L_j	cost of permanent data loss due to ransomware attacks in group j
T_j	loss from business interruptions due to ransomware in group j
D	base difficulty of perpetrating ransomware attacks
C_B	unit cost of backup effort
C_A	unit cost of attack effort
C_D	fixed cost of developing ransomware
b_i	backup effort of organization i
p_i	decision of organization i about ransom payment
a_j	attacker's effort against group j
r	ransom demanded by the attacker
$V_j(a_1, a_2)$	probability of an organization $i \in G_j$ becoming compromised

2.1   Players

On the defenders' side, players model organizations that are susceptible to ransomware attacks. Based on their characteristics, we divide these organizations

into two groups (e.g., hospitals and universities). We will refer to these two groups as group 1 and group 2, and we let set G_1 and set G_2 denote their members, respectively. On the attacker's side, there is a single player, who models cybercriminals that may develop and deploy ransomware. Note that we model attackers as a single entity since our goal is to understand and improve the behavior of defenders; hence, competition between attackers is not our current focus. Our model—and many of our results—could be extended to multiple attackers in a straightforward manner.

2.2 Strategy Spaces

With our work, we focus on the mitigation of ransomware attacks through backups (as a part of contingency plans), and we will not consider the organizations' decisions on preventative effort (e.g., firewall security policies). The tradeoff between mitigation and preventative efforts has been subject of related work [12]. We let $b_i \in \mathbb{R}_+$ denote the backup effort of organization i, which captures the frequency and coverage of backups as well as contingency plans and preparations. Compromised organizations also have to decide whether they pay the ransom or sustain permanent data loss. We let $p_i = 1$ if organization i pays, and $p_i = 0$ if it does not pay.

The attacker first decides whether it wishes to engage in cybercrime using ransomware. If the attacker chooses to engage, then it has to select the amount of effort spent on perpetrating the attacks. We let $a_1 \in \mathbb{R}_{\geq 0}$ and $a_2 \in \mathbb{R}_{\geq 0}$ denote the attacker's effort spent on attacking group 1 and group 2, respectively. If the attacker chooses not to attack group j (or not to engage in cybercrime at all), then $a_j = 0$. We assume that each organization within a group falls victim to the attack with the same probability $V_j(a_1, a_2)$, which depends on the attacker's effort, independently of the other organizations. Since the marginal utility of attack effort is typically decreasing, we assume that the infection probability $V_j(a_1, a_2)$ is

$$V_j(a_1, a_2) = \frac{a_j}{D + (a_1 + a_2)}, \tag{1}$$

where D is the base difficulty of attacks. In the formula above, the numerator expresses that as the attacker increases its effort on group j, more and more organizations fall victim. Meanwhile, the denominator captures the decreasing marginal utility: as the attacker increases its attack effort, compromising additional targets becomes more and more difficult. In practice, this corresponds to the increasing difficulty of finding new targets as organizations are becoming aware of a widespread ransomware attack and are taking precautions, etc.

The attacker also has to choose the amount of ransom r to demand from compromised organizations in exchange for restoring their data and systems.

Stages The game consists of two stages:

  - Stage I: Organizations choose their backup efforts $\boldsymbol{b}$, while the attacker chooses its attack effort a_1 and a_2, as well as its ransom demand r.

– Stage II: Each organization $i \in G_j$ becomes compromised with probability $V_j(a_1, a_2)$. Then, organizations that have fallen victim to the attack choose whether to pay the ransom or not, which is represented by $\boldsymbol{p}$.

2.3 Payoffs

Defender's Payoff If an organization i, which belongs to group $j \in \{1, 2\}$, has not fallen victim to a ransomware attack, then its payoff is

$$\mathcal{U}_{O_i}\big|_{\text{not compromised}} = W_j - C_B \cdot b_i - \beta \frac{F_j}{b_i}, \tag{2}$$

where W_j is the initial wealth of organizations in group j, F_j is their loss resulting from corrupted data due to random failures[4], and C_B is the unit cost of backup effort. The parameter β is a behavioral discount factor, which captures the robust empirical observation that individuals underappreciate the future consequences of their current actions [23]. The magnitude of β is assumed to be related to underinvestment in security and privacy technologies [1,13]; in our case, procrastination of backup investments [4].[5]

Otherwise, we have two cases. If organization i decides to pay the ransom r, then its payoff is

$$W_j - C_B \cdot b_i - \beta \left(\frac{F_j}{b_i} + T_j + r \right),$$

where T_j is the loss resulting from temporary business interruption due to the attack. On the other hand, if organization i does not pay the ransom, then its payoff is

$$W_j - C_B \cdot b_i - \beta \left(\frac{F_j + L_j}{b_i} + T_j \right),$$

where L_j is the loss resulting from permanent data loss due to the ransomware attack. Using p_i, we can express a compromised organization's payoff as

$$\mathcal{U}_{O_i}\big|_{\text{compromised}} = W_j - C_B \cdot b_i - \beta \left(\frac{F_j + (1 - p_i) \cdot L_j}{b_i} + T_j + p_i \cdot r \right). \tag{3}$$

---

[4] Since we interpret effort b_i primarily as the frequency of backups, the fraction $\frac{1}{b_i}$ is proportional to the expected time since the last backup. Consequently, we assume that data losses are inversely proportional to b_i. Note that alternative interpretations, such as assuming b_i to be the level of sophistication of backups (e.g., air-gapping), which determines the probability that the backups remain uncompromised, also imply a similar relationship.

[5] We are unaware of any behavioral study that specifically investigates the impact of the *present bias* behavioral discount factor on backup decisions, but industry experts argue strongly for its relevance. For example, in the context of the 2017 WannaCry ransomware attacks a commentary about backups stated: "This may be stating the obvious, but it's still amazing to know the sheer number of companies that keep procrastinating over this important task [31]."

By combining Equations (2) and (3) with V_j, we can express the expected utility of an organization $i \in G_j$ as

$$
\begin{aligned}
\mathrm{E}\left[\mathcal{U}_{O_i}\right] = {} & (1 - V_j(a_1, a_2)) \left[W_j - C_B \cdot b_i - \beta \frac{F_j}{b_i} \right] \\
& + V_j(a_1, a_2) \left[W_j - C_B \cdot b_i - \beta \left(\frac{F_j + (1 - p_i) \cdot L_j}{b_i} + T_j + p_i \cdot r \right) \right].
\end{aligned}
\tag{4}
$$

Attacker's Payoff For the attacker's payoff, we also have two cases. If the attacker decides not to participate (i.e., if $a_1 = 0$ and $a_2 = 0$), then its payoff is simply zero. Otherwise, its payoff depends on the number of organizations that have fallen victim and decided to pay. We can calculate the expected number of victims who pay the ransom as

$$
\mathrm{E}[\text{number of victims who pay the ransom}] = \sum_j \sum_{i \in G_j} V_j(a_1, a_2) \cdot p_i \tag{5}
$$

since each organization $i \in G_j$ is compromised with probability V_j, and $p_i = 1$ if organization i chooses to pay (and $p_i = 0$ if it does not pay).

Then, we can express the attacker's expected payoff simply as

$$
\mathrm{E}\left[\mathcal{U}_A\right] = \left[\sum_j \sum_{i \in G_j} V_j(a_1, a_2) \cdot p_i \right] \cdot r - C_A \cdot (a_1 + a_2) - C_D, \tag{6}
$$

where C_A is the unit cost of attack effort, and C_D is the fixed cost of developing a ransomware, which the attacker must pay if it decides to engage (i.e., if $a_1 > 0$ or $a_2 > 0$).

2.4 Solution Concepts

We assume that every player is interested in maximizing its expected payoff, and we use subgame perfect Nash equilibrium as our solution concept. We also assume that organizations always break ties (i.e., when both paying and not paying are best responses) by choosing to pay. Note that the latter assumption has no practical implications, it only serves to avoid pathological mathematical cases.

Further, in our numerical analysis in Section 4, we will use the *social optimum* concept for comparison with the Nash equilibrium results. In the social optimum, a social planner can coordinate the decisions of organizations, such that it yields the maximum aggregate outcome for the organizations, subject to an optimal response by the attacker (who is not guided by the social planner).

3 Analysis

In this section, we analyze our proposed game-theoretic model of the ransomware ecosystem. Our solution concept, as mentioned in Section 2.4, is the subgame

perfect Nash equilibrium. Hence, in our analysis, we first calculate each organization's decision in Stage II in Section 3.1. In other words, we derive under what conditions a victim organization will pay the requested ransom from the attacker. Then, we calculate the best-response backup strategy for each organization in Stage I of the game in Section 3.2. Third, we calculate the attacker's best-response, i.e., demanded ransom and the attacker's effort, in Section 3.3. By calculating the attacker's and the organizations' best-responses, we can then derive the Nash equilibrium in Section 3.4.

3.1 Compromised Organizations' Ransom Payment Decisions

We begin our analysis by studying the compromised organizations' best-response payment strategies in the second stage of the game.

Lemma 1. *For organization $i \in G_j$, paying the ransom (i.e., $p_i = 1$) is a best response if and only if*

$$r \leq \frac{L_j}{b_i}. \tag{7}$$

Proof of Lemma 1 is provided in Appendix A.1.

Lemma 1 means that an organization will pay the demanded ransom if the demanded value is not higher than the average permanent data loss due to ransomware attack.

3.2 Organizations' Backup Decisions

We next study the organizations' best-response backup strategies in the first stage. We assume that compromised organizations will play their best responses in the second stage (see Lemma 1), but we do not make any assumptions about the attacker's effort or ransom strategies. We first characterize the organizations' best-response backup strategies when they do not face any attacks (Lemma 2) and then in the case when they are threatened by ransomware (Lemma 3).

Lemma 2. *If the attacker chooses not to attack group j (i.e., $a_j = 0$), then the unique best-response backup strategy for organization $i \in G_j$ is*

$$b_i^* = \sqrt{\beta \frac{F_j}{C_B}}. \tag{8}$$

Proof of Lemma 2 is provided in Appendix A.2.

Note that in Lemma 2, an organization chooses its backup strategy by considering data loss due to random failures rather than data loss due to ransomware attack since that organization is not chosen to be attacked by the attacker.

Lemma 3 calculates an organization's best-response backup strategy. Note that an organization chooses its backup strategy at Stage I. In this stage, an organization does not know whether it is the target of a ransomware attack and if that organization is the target of a ransomware attack, whether the attack is successful.

Lemma 3. *If the attacker chooses to attack group j (i.e., $a_j > 0$), then the best-response backup strategy b_i^* for organization $i \in G_j$ is*

- *if $b_j^{low} > \frac{L_j}{r}$, then $b_i^* = b_j^{high}$;*
- *if $b_j^{high} < \frac{L_j}{r}$, then $b_i^* = b_j^{low}$;*
- *otherwise, $b_i^* \in \left\{ b_j^{low}, b_j^{high} \right\}$ (the one that gives the higher payoff or both if the resulting payoffs are equal),*

where $b_j^{low} = \sqrt{\beta \frac{F_j}{C_B}}$ and $b_j^{high} = \sqrt{\beta \frac{(F_j + V_j(a_1, a_2)L_j)}{C_B}}$.

Proof of Lemma 3 is provided in Appendix A.3.

Lemma 3 shows the best-response backup strategy when an organization is under attack. If the demanded ransom value is high, i.e., $r > \frac{L_j}{b_j^{low}}$, an organization takes into account the data loss due to ransomware attack as well as the data loss due to random failure when choosing the backup strategy level. On the other hand, if the demanded ransom is low, i.e., $r < \frac{L_j}{b_j^{high}}$, an organization does not care about the data loss due to ransomware attack even when that organization is under attack. In other words, that organization behaves like an organization that is not under ransomware attack, i.e., similar to Lemma 2.

3.3 Attacker's Best Response

Building on the characterization of the organizations' best responses, we now characterize the attacker's best-response strategies. Notice that the lemmas presented in the previous section show that an organization's best response does not depend on the identity of the organization, only on its group. Since we are primarily interested in studying equilibria, in which everyone plays a best response, we can make the following assumptions:

- All organizations within a group j play the same backup strategy, which is denoted by $\hat{b}_j$.
- $\frac{L_1}{\hat{b}_1} \leq \frac{L_2}{\hat{b}_2}$.

The second assumption is without loss of generality since we could easily renumber the groups.

In Lemma 4, we calculate the attacker's best-response demanded ransom given the attacker's effort and the organizations' backup strategies.

Lemma 4. *If the attacker's effort (a_1, a_2) is fixed, then its best-response ransom demand r^* is*

- $\frac{L_1}{\hat{b}_1}$ *if $|G_1| V_1(a_1, a_2)\frac{L_1}{\hat{b}_1} > |G_2| V_2(a_1, a_2) \left(\frac{L_2}{\hat{b}_2} - \frac{L_1}{\hat{b}_1} \right)$*
- $\frac{L_2}{\hat{b}_2}$ *if $|G_1| V_1(a_1, a_2)\frac{L_1}{\hat{b}_1} < |G_2| V_2(a_1, a_2) \left(\frac{L_2}{\hat{b}_2} - \frac{L_1}{\hat{b}_1} \right)$*
- *both $\frac{L_1}{\hat{b}_1}$ and $\frac{L_2}{\hat{b}_2}$ otherwise.*

Proof of Lemma 4 is provided in Appendix A.4.

Lemma 5 shows how the attacker divides its best-response attack effort between the two groups of organizations. Here, we assume that $a_1 + a_2 = a_{sum}$, where a_{sum} is a constant. Note that it is possible that the attacker decides not to attack either of the groups of organizations. The reason is that the benefit for the attacker from a ransomware attack may be lower than the cost of the attack. Hence, a rational attacker will abstain from attacking either of the groups.

Lemma 5. *The attacker's best-response attack effort (a_1^*, a_2^*) is as follows:*

- $a_1^* = 0$ *and* $a_2^* = a_{sum}$ *if* $|G_1| \cdot 1_{\left\{r \le \frac{L_1}{b_1}\right\}} < |G_2| \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}}$ *and* $\frac{a_{sum}}{D + a_{sum}} |G_2| \cdot r \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}} > C_A \cdot a_{sum} + C_D$,

- $a_1^* = a_{sum}$ *and* $a_2^* = 0$ *if* $|G_1| \cdot 1_{\left\{r \le \frac{L_1}{b_1}\right\}} > |G_2| \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}}$ *and* $\frac{a_{sum}}{D + a_{sum}} |G_1| \cdot r \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}} > C_A \cdot a_{sum} + C_D$,

- *any* a_1^* *between* 0 *and* a_{sum} *and* $a_2^* = a_{sum} - a_1^*$ *if* $|G_1| \cdot 1_{\left\{r \le \frac{L_1}{b_1}\right\}} = |G_2| \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}}$ *and* $\frac{a_{sum}}{D + a_{sum}} |G_2| \cdot r \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}} > C_A \cdot a_{sum} + C_D$.

- $a_1^* = a_2^* = 0$ *otherwise.*

Proof of Lemma 5 is provided in Appendix A.5.

3.4 Equilibria

Proposition 1 provides the necessary and sufficient conditions for the attacker's strategy to abstain from attack, i.e., $a_1^* = a_2^* = 0$, and $\hat{b}_1^* = \sqrt{\beta \frac{F_1}{C_B}}$ and $\hat{b}_2^* = \sqrt{\beta \frac{F_2}{C_B}}$ is Nash equilibrium.

Proposition 1. *The attacker choosing not to attack and the organizations choosing backup efforts $\sqrt{\beta \frac{F_1}{C_B}}$ and $\sqrt{\beta \frac{F_2}{C_B}}$ is an equilibrium if and only if each of the following conditions are satisfied:*

- $\frac{L_2 \cdot a_{sum} \cdot |G_2| \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}}}{L_2 (D + a_{sum})(C_A \cdot a_{sum} + C_D)} < \sqrt{\beta \frac{F_2}{C_B}}$ *and* $|G_1| \cdot 1_{\left\{r \le \frac{L_1}{b_1}\right\}} \le |G_2| \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}}$

- $\frac{L_2 \cdot a_{sum} \cdot |G_1| \cdot 1_{\left\{r \le \frac{L_1}{b_1}\right\}}}{L_2 (D + a_{sum})(C_A \cdot a_{sum} + C_D)} < \sqrt{\beta \frac{F_2}{C_B}}$ *and* $|G_1| \cdot 1_{\left\{r \le \frac{L_1}{b_1}\right\}} > |G_2| \cdot 1_{\left\{r \le \frac{L_2}{b_2}\right\}}$

Proof of Proposition 1 is provided in Appendix A.6.

4 Numerical Illustrations

In this section, we present numerical results on our model. We first compare equilibria to social optima, and we study the effect of changing the values of key parameters (Section 4.1). We then investigate interdependence between multiple groups of organizations, which is caused by the strategic nature of attacks, and we again study the effect of changing key parameters (Section 4.2).

For any combination of parameter values, our game has at most one equilibrium, which we will plot in the figures below. However, for some combinations, the game does not have an equilibrium. In these cases, we used iterative best responses:

1. starting from an initial strategy profile,
2. we changed the attacker's strategy to a best response,
3. we changed the organization's strategy to a best response,
4. and then we repeated from Step 2.

We found that regardless of the initial strategy profile, the iterative best-response dynamics end up oscillating between two strategy profiles. Since these strategy profiles were very close, we plotted their averages in place of the equilibria in the figures below.

4.1 Equilibria and Social Optima

For clarity of presentation, we consider a single organization type in this subsection. The parameter values used in this study are as follows: $|G| = 100$, $W = 100$, $\beta = 0.9$, $F = 5$, $L = 5$, $T = 10$, $C_B = 1$, $D = 10$, $C_A = 10$, and $C_D = 10$ (unless stated otherwise).
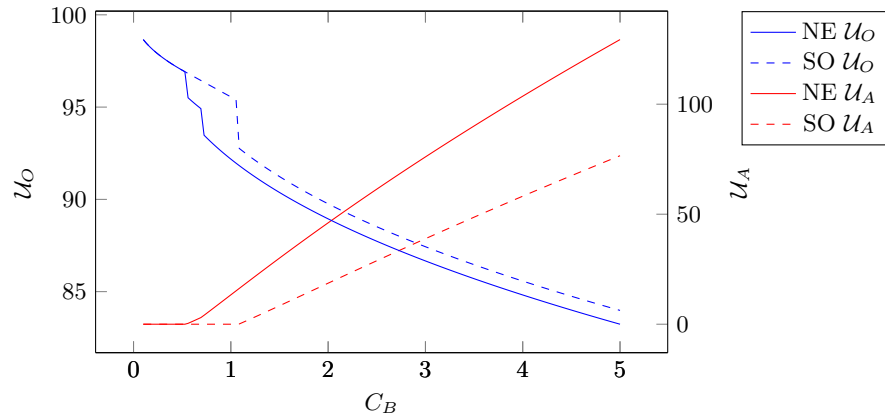

Fig. 1. The expected payoff of the attacker (red) and an individual organization (blue) in equilibrium (solid line —) and in social optimum (dashed line - - -) for various backup cost values C_B.

Figure 1 shows the expected payoffs of an individual organization and the attacker for various values of the unit cost C_B of backup effort. In practice, the unit cost of backup effort may change, for example, due to technological improvements (decreasing the cost) or growth in the amount of data to be backed up (increasing the cost). When this cost is very low ($C_B < 0.5$), organizations can perform frequent and sophisticated backups, which means that the amount of data that may be compromised—and hence, the ransom that they are willing to pay—is very low. As a result, the attacker is deterred from deploying ransomware ($\mathcal{U}_A = 0$) since its income from ransoms would not cover its expenses. For higher costs ($0.5 \leq C_B < 1$), the organizations' equilibrium payoff is much lower since they choose to save on backups, which incentivizes the attacker to deploy ransomware and extort payments from them. In this case, the social optimum for the organizations is to maintain backup efforts and, hence, deter the attacker. For even higher costs ($C_B \geq 1$), deterrence is not socially optimal. However, the equilibrium payoffs are still lower since organizations shirk on backup efforts, which leads to more intense attacks and higher ransom demands.
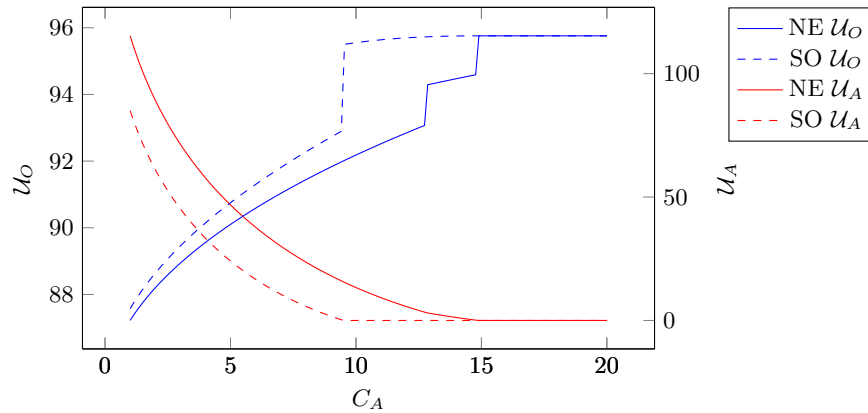

Fig. 2. The expected payoff of the attacker (red) and an individual organization (blue) in equilibrium (solid line —) and in social optimum (dashed line - - -) for various attack cost values C_A.

Figure 2 shows the expected payoff of an individual organization and the attacker for various values of the unit cost C_A of attack effort. In practice, the unit cost of attack effort can change, e.g., due to the development of novel attacks and exploits (lowering the cost) or the deployment of more effective defenses (increasing the cost). Figure 2 shows phenomena that are similar to the ones exhibited in Figure 1. When the attacker is at a technological advantage (i.e., when C_A is low), deterrence is not a realistic option for organizations. However, they can improve their payoffs—compared to the equilibrium—by coordinating and investing more in backups, thereby achieving social optimum. For higher

attack costs ($10 < C_A \leq 15$), this coordination can result in significantly higher payoffs since deterrence becomes a viable option. For very high attack costs ($C_A > 15$), compromising an organization costs more than what the attacker could hope to collect with ransoms; hence, coordination is no longer necessary to deter the attacker.
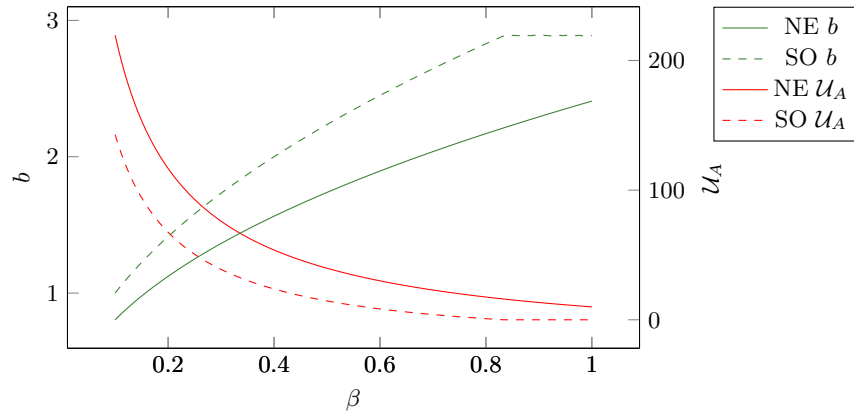

Fig. 3. The attacker's expected payoff (red) and the organizations' backup strategy (green) in equilibrium (solid line —) and in social optimum (dashed line - - -) for various discounting factor values β.

Figure 3 shows how the organizations' backup efforts b and the attacker's payoff are effected by the behavioral discount factor β. With low values of β, organizations underappreciate future consequences; hence, they shirk on backup efforts (as evidenced by low values of b). With high values of β, organizations care more about future losses, so they invest more in backup efforts (resulting in high values of b). We see that in all cases, there is a significant difference between the equilibrium and the social optimum. This implies that regardless of the organizations' appreciation of future consequences, coordination is necessary. In other words, low backup efforts cannot be attributed only to behavioral factors.

4.2 Interdependence

Now, we study the interdependence between two groups of organizations. We instantiate the parameters of both organizations (and the attacker) with the same values as in the previous subsection. Note that more numerical illustrations are available in Appendix B.

Figure 4 shows the payoffs of individual organizations from the two groups as well as the attacker, for various values of the costs L_1 and L_2 of permanent data loss. As expected, we see from the attacker's payoff (Figure 4(c)) that as loss costs increase, organizations become more willing to pay higher ransoms, so the
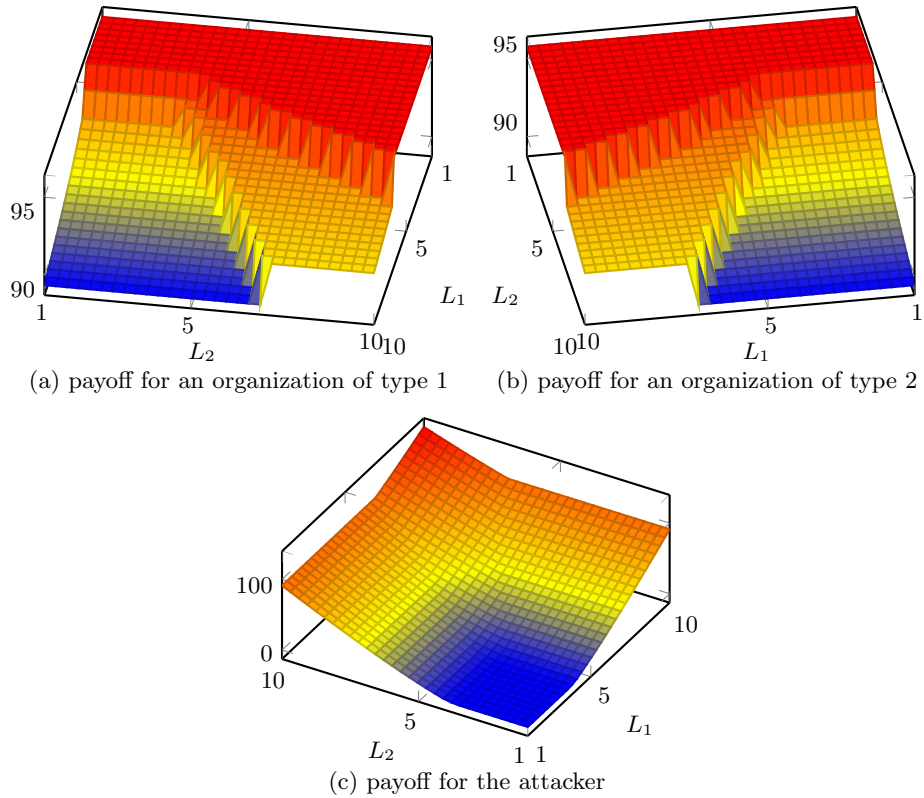
(a) payoff for an organization of type 1   (b) payoff for an organization of type 2

(c) payoff for the attacker

Fig. 4. The expected payoff of individual organizations of (a) type 1 and (b) type 2 as well as (c) the attacker in equilibrium for various data loss costs L_1 and L_2.

attacker's payoff increases. On the other hand, we observe a more interesting phenomenon in the organizations' payoffs. As the loss cost (e.g., L_1) of one group (e.g., group 1) increases, the payoff of organizations in that group (e.g., Figure 4(a)) decreases. However, we also see an *increase* in the payoff (e.g., Figure 4(b)) of organizations in the *other* group (e.g., group 2). The reason for this increase is in the strategic nature of attacks: as organizations in one group become more attractive targets, attackers are more inclined to focus their efforts on this group, which results in lower intensity attacks against the other group. This substitution effect, which can be viewed as a negative externality between groups of organizations, is strong when the attacker's efforts are focused (e.g., when ransomware is deployed using spear-phishing campaigns).

5 Related Work

Ransomware: Early work by Luo and Liao, in 2007 and 2009, respectively, represented first exploratory analyses of the ransomware phenomenon [21,22]. They

focus on increased awareness (in particular, by employees) as a major means to diminish the effectiveness of ransomware attacks, which is a key recommendation regarding ransomware mirrored in the 2017 Verizon DBIR report ten years later: "stress the importance of software updates to anyone who'll listen [32]."

In 2010, Gazet investigated the quality of code, functionalities and cryptographic primitives of 15 samples of ransomware [11]. The studied sample of ransomware malware was quite basic from the perspective of programming quality and sophistication, and did not demonstrate a high level of thoroughness regarding the application of cryptographic primitives. However, the analysis also showed the ability to mass propagate as a key feature.

Highlighting ransomware's increasing relevance, Proofpoint reported that 70% of all malware encountered in the emails of its customer base during a 10-month interval in 2016 was ransomware. At the same time, the same company reported that the number of malicious email attachments grew by about 600% in comparison to 2015 [25]. In addition, many modern forms of ransomware have worm capabilities as demonstrated in a disconcerting fashion by the 2017 WannaCrypt/WannaCry attack, which affected 100,000s of systems of individual users and organizations leading even to the breakdown of industrial facilities.

Other studies also focus on providing practical examples of and empirical data on ransomware including work by O'Gorman and McDonald [24], who provide an in-depth perspective of specific ransomware campaigns and their financial impact. In a very comprehensive fashion, Kharraz et al. analyze ransomware code observed in the field between 2006 and 2014 [17]. Their results mirror Gazet's observation on a much broader pool of 1,359 ransomware samples, i.e., currently encountered ransomware lacks complexity and sophistication.

Nevertheless, it causes major harm. To cite just a few figures, the total cost of ransomware attacks (including paid ransoms) increased to $209 Million for the first three months in 2016 according to FBI data. In comparison, for 2015 the FBI only reported damages of about $24 Million [9].

Drawing on their earlier research, Kharraz et al. developed practices to stem the impact of ransomware. Their key insight is that ransomware needs to temper with user or business data, so that increased monitoring of data repositories can stop ransomware attacks while they unfold, and detect novel attacks that bypassed even sophisticated preventative measures [16]. Scaife et al. also present an early-warning detection approach regarding suspicious file activity [26].

Extending this line of work to a new context, Andronio et al. study ransomware in the mobile context and develop an automated approach for detection while paying attention to multiple typical behaviors of ransomware including the display of ransom notes on the device's screen [2]. Likewise, Yang et al. focus on mobile malware detection, and specifically ransomware identification [33].

Ransomware attacks appear to be predominantly motivated by financial motives, which supports the usage of an economic framework for their analysis. However, other related types of attacks such as malicious device bricking (see, for example, the BrickerBot attack focusing on IoT devices [28]) may be based on a purely destructive agenda, with less clearly identifiable motives.

While knowledge about the technical details and financial impact of ransomware is growing, we are unaware of any research which focuses on the strategic economic aspects of the interactions between cybercriminals that distribute ransomware and businesses or consumers who are affected by these actions.

Economics of Security: Game-theoretic models to better understand security scenarios have gained increased relevance due to the heightened professionalism of cybercriminals. Of central interest are models that capture interdependencies or externalities arising from actions by defenders or attackers [19].

A limited number of research studies focus on the modeling of the attack side. For example, Schechter and Smith capture different attacker behaviors [27]. In particular, they consider the cases of serial attacks where attackers aim to compromise one victim after another, and the case of a parallel attack, where attackers can automate their attacks to focus on multiple defenders at one point in time. We follow the latter approach, which has high relevance for self-propagating ransomware such as WannaCrypt/WannaCry.

Another relevant aspect of our work are incentives to invest in backup technologies, which have found only very limited consideration in the literature. Grossklags et al. investigate how a group of defenders individually decide on how to split security investments between preventative technologies and recovery technologies (called self-insurance) [12]. In their model, preventative investments are subject to interdependencies drawing on canonical models from the literature on public goods [30], while recovery investments are effective independent from others' choices. Fultz and Grossklags [10] introduce strategically acting attackers in this framework, who respond to preventative investments by all defenders. In our model, backup investments are also (partially) effective irrespective of others' investment choices. However, in the context of ransomware, pervasive investments in backup technologies can have a deterrence and/or displacement effect on attackers [5], which we capture with our work.

While we draw on these established research directions, to the best of our knowledge, our work is the first game-theoretic approach focused on ransomware.

6   Concluding Remarks

In this paper, we have developed a game-theoretic model, which is focused on key aspects of the adversarial interaction between organizations and ransomware attackers. In particular, we place significant emphasis on the modeling of security investment decisions for mitigation, i.e., level of backup effort, as well as the strategic decision to pay a ransom or not.

These factors are interrelated and also influence attacker behavior. For example, in the context of kidnappings by terrorists it has been verified based on incident data that negotiating with kidnappers and making concessions encourages substantially more kidnappings in the future [6]. We would expect a similar effect in the context of ransomware, where independently acting organizations who are standing with the "back against the wall" have to make decisions about

ransom payments to get operations going again, or to swallow the bitter pill of rebuilding from scratch and not giving in to cybercriminals. Indeed, our analysis shows that there is a sizable gap between the decentralized decision-making at equilibrium and the socially optimal outcome. This raises the question whether organizations paying ransoms should be penalized? However, this (in turn) poses a moral dilemma, for example, when patient welfare at hospitals or critical infrastructure such at power plants are affected *now*.

An alternative pathway is to (finally) pay significantly more attention to backup efforts as a key dimension of overall security investments. The relative absence of economic research focused on optimal mitigation and recovery strategies is one key example of this omission. A laudable step forward is the recently released factsheet document by the U.S. Department of Health & Human Service on ransomware and the Health Insurance Portability and Accountability Act (HIPAA) [29]. It not only states that the encryption of health data by ransomware should be considered a security breach under HIPAA (even though no data is exfiltrated[6]), but also that having a data backup plan is a required security effort for all HIPAA covered organizations.

An interesting question for future research is the role of cyberinsurance in the context of ransomware, i.e., specifically policies including cyber-extortion. How would these policies have to be designed to achieve desirable outcomes? As discussed above, in the case of kidnappings one would worry about incentivizing future kidnappings by making concessions via kidnapping insurance [8]; however, the design space in the context of ransomware is significantly more complex, but also offers more constructive directions.

References

1. Acquisti, A., Grossklags, J.: What can behavioral economics teach us about privacy? In: Digital Privacy: Theory, Technologies, and Practices, pp. 363–379. Auerbach Publications (2007)
2. Andronio, N., Zanero, S., Maggi, F.: HelDroid: Dissecting and detecting mobile ransomware. In: Proceedings of the International Workshop on Recent Advances in Intrusion Detection. pp. 382–404. Springer (2015)
3. Backblaze: Backup awareness survey, our 10th year, industry report. Available at: `https://www.backblaze.com/blog/backup-awareness-survey/`.

---

[6] The reasoning is as follows: "When electronic protected health information (ePHI) is encrypted as the result of a ransomware attack, a breach has occurred because the ePHI encrypted by the ransomware was acquired (i.e., unauthorized individuals have taken possession or control of the information), and thus is a "disclosure" not permitted under the HIPAA Privacy Rule. [29]"

4. Baddeley, M.: Information security: Lessons from behavioural economics. In: Workshop on the Economics of Information Security (WEIS) (2011)
5. Becker, G.: Crime and punishment: An economic approach. Journal of Political Economy 76(2), 169–217 (March–April 1968)
6. Brandt, P., George, J., Sandler, T.: Why concessions should not be made to terrorist kidnappers. European Journal of Political Economy 44, 41 – 52 (2016)
7. Bruskin Research: Nearly one in four computer users have lost content to blackouts, viruses and hackers according to new national survey (2001), survey conducted for Iomega Corporation.
8. Fink, A., Pingle, M.: Kidnap insurance and its impact on kidnapping outcomes. Public Choice 160(3), 481–499 (Sep 2014)
9. Finkle, J.: Ransomware: Extortionist hackers borrow customer-service tactics (2016), available at: http://www.reuters.com/article/us-usa-cyber-ransomware-idUSKCN0X917X.
10. Fultz, N., Grossklags, J.: Blue versus red: Towards a model of distributed security attacks. In: Dingledine, R., Golle, P. (eds.) Financial Cryptography and Data Security, Lecture Notes in Computer Science, vol. 5628, pp. 167–183. Springer (2009)
11. Gazet, A.: Comparative analysis of various ransomware virii. Journal in Computer Virology 6(1), 77–90 (2010)
12. Grossklags, J., Christin, N., Chuang, J.: Secure or insure?: A game-theoretic analysis of information security games. In: Proceedings of the 17th International World Wide Web Conference. pp. 209–218 (2008)
13. Grossklags, J., Barradale, N.: Social status and the demand for security and privacy. In: Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium. pp. 83–101. Springer (2014)
14. IBM: IBM study: Businesses more likely to pay ransomware than consumers (2016), industry report. Available at: http://www-03.ibm.com/press/us/en/pressrelease/51230.wss.
15. Kabooza: Global backup survey: About backup habits, risk factors, worries and data loss of home PCs (Jan 2009), available at: http://www.kabooza.com/globalsurvey.html.
16. Kharraz, A., Arshad, S., Mulliner, C., Robertson, W., Kirda, E.: UNVEIL: A large-scale, automated approach to detecting ransomware. In: Proceedings of the 25th USENIX Security Symposium (USENIX Security). pp. 757–772 (2016)
17. Kharraz, A., Robertson, W., Balzarotti, D., Bilge, L., Kirda, E.: 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 (DIMVA). pp. 3–24. Springer (2015)
18. KnowBe4: The 2017 endpoint protection ransomware effectiveness report (2017), industry report. Available at: https://www.knowbe4.com/hubfs/Endpoint%20Protection%20Ransomware%20Effectiveness%20Report.pdf.
19. Laszka, A., Felegyhazi, M., Buttyan, L.: A survey of interdependent information security games. ACM Computing Surveys 47(2), 23:1–23:38 (Aug 2014)
20. Liao, K., Zhao, Z., Doupé, A., Ahn, G.J.: Behind closed doors: Measurement and analysis of CryptoLocker ransoms in Bitcoin. In: Proceedings of the 2016 APWG Symposium on Electronic Crime Research (eCrime) (2016)
21. Luo, X., Liao, Q.: Awareness education as the key to ransomware prevention. Information Systems Security 16(4), 195–202 (2007)
22. Luo, X., Liao, Q.: Ransomware: A new cyber hijacking threat to enterprises. In: Gupta, J., Sharma, S. (eds.) Handbook of Research on Information Security and Assurance, pp. 1–6. IGI Global (2009)

23. O'Donoghue, T., Rabin, M.: Doing it now or later. American Economic Review 89(1), 103–124 (1999)
24. O'Gorman, G., McDonald, G.: Ransomware: A growing menace. Symantec Security Response (2012)
25. Proofpoint: Threat summary: Q4 2016 & year in review, industry report. Available at: `https://www.proofpoint.com/sites/default/files/proofpoint_q4_threat_report-final-cm.pdf`.
26. Scaife, N., Carter, H., Traynor, P., Butler, K.: Cryptolock (and drop it): Stopping ransomware attacks on user data. In: Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS). pp. 303–312 (2016)
27. Schechter, S., Smith, M.: How much security is enough to stop a thief? In: Proceedings of the Seventh International Financial Cryptography Conference (FC 2003). pp. 122–137 (2003)
28. Simon, R.: Mirai, BrickerBot, Hajime attack a common IoT weakness (2017), available at: `https://securingtomorrow.mcafee.com/mcafee-labs/mirai-brickerbot-hajime-attack-common-iot-weakness/`.
29. U.S. Department of Health & Human Service: Fact sheet: Ransomware and HIPAA (2016), available at: `https://www.hhs.gov/sites/default/files/RansomwareFactSheet.pdf`.
30. Varian, H.: System reliability and free riding. In: Camp, L., Lewis, S. (eds.) Economics of Information Security (Advances in Information Security, Volume 12), pp. 1–15. Kluwer Academic Publishers, Dordrecht, The Netherlands (2004)
31. Venkat, S.: Lessons for telcos from the WannaCry ransomware attack (2017), cerillion blog. Available at: `http://www.cerillion.com/Blog/May-2017/Lessons-for-Telcos-from-the-WannaCry-attack`.
32. Verizon: 2017 Data breach investigations report: Executive summary, industry report.
33. Yang, T., Yang, Y., Qian, K., Lo, D.C.T., Qian, Y., Tao, L.: Automated detection and analysis for Android ransomware. In: Proceedings of the 1st IEEE International Conference on Big Data Security on Cloud (DataSec). pp. 1338–1343. IEEE (2015)
34. Young, A., Yung, M.: Cryptovirology: Extortion-based security threats and countermeasures. In: Proceedings of the 1996 IEEE Symposium on Security and Privacy. pp. 129–140 (1996)
35. Young, A., Yung, M.: Cryptovirology: The birth, neglect, and explosion of ransomware. Communications of the ACM 60(7), 24–26 (Jul 2017)

A Proofs

A.1 Proof of Lemma 1

From Equation (3), we have that the best-response strategy p_i^* of organization i is

$$p_i^* \in \underset{p \in \{0,1\}}{\mathrm{argmax}} \left[W_j - C_B \cdot b_i - \beta \left(\frac{F_j + (1-p) \cdot L_j}{b_i} + T_j + p \cdot r \right) \right] \qquad (9)$$

$$= \underset{p \in \{0,1\}}{\mathrm{argmax}} \, p \cdot \left(\frac{L_j}{b_i} - r \right). \qquad (10)$$

Clearly, $p_i^* = 1$ is a best response if and only if $\frac{L_j}{b_i} - r \geq 0$, and $p_i^* = 0$ is a best response if and only if $\frac{L_j}{b_i} - r \leq 0$. □

A.2 Proof of Lemma 2

From Equation (2), we have that the best-response strategy b_i^* of organization i is

$$b_i^* \in \underset{b_i \in \mathbb{R}_+}{\operatorname{argmax}} \left[W_j - C_B \cdot b_i - \beta \frac{F_j}{b_i} \right]. \tag{11}$$

To find the maximizing b_i^*, we take the first derivative of the payoff, and set it equal to 0:

$$-C_B + \beta \frac{F_j}{b_i^{*2}} = 0 \tag{12}$$

$$b_i^* = \pm \sqrt{\beta \frac{F_j}{C_B}}, \tag{13}$$

Since $b_i \in \mathbb{R}_+$, the only local optima is $b_i^* = \sqrt{\beta \frac{F_j}{C_B}}$. Further, the payoff is a concave function of b_i as the second derivative is negative, which means that this b_i^* is the global optimum and, hence, a unique best response. $\square$

A.3 Proof of Lemma 3

We can express the expected payoff of organization i as

$$\begin{aligned} \mathrm{E}\left[\mathcal{U}_O\right] =& (1 - V_j(a_1, a_2)) \left[W_j - C_B \cdot b_i - \beta \frac{F_j}{b_i} \right] \\ &+ V_j(a_1, a_2) \left[W_j - C_B \cdot b_i - \beta \left(\frac{F_j + (1 - p_i^*(b_i)) \cdot L_j}{b_i} + T_j + p_i^*(b_i) \cdot r \right) \right] \\ =& W_j - C_B \cdot b_i - \beta \frac{F_j}{b_i} - \beta V_j(a_1, a_2) \left[\frac{(1 - p_i^*(b_i)) \cdot L_j}{b_i} + T_j + p_i^*(b_i) \cdot r \right]. \end{aligned}$$

Notice that we incorporated the second-stage best-response payment strategy $p_i^*(b_i)$, which is given by Lemma 1 and depends on the backup strategy b_i.

We find the best response b_i^* by dividing the search space into two regions, $b_i \leq \frac{L_j}{r}$ and $b_i \geq \frac{L_j}{r}$, and find the best strategies in both regions.

Case 1: First, consider $b_i \leq \frac{L_j}{r}$, which implies that $p_i^*(b_i) = 1$. In this case, the best backup strategy $b_i^* \leq \frac{L_j}{r}$ is

$$b_i^* \in \underset{b_i \leq \frac{L_j}{r}}{\operatorname{argmax}} \left[W_j - C_B \cdot b_i - \beta \frac{F_j}{b_i} - \beta V_j(a_1, a_2)(T_j + r) \right] \tag{14}$$

$$= \underset{b_i \leq \frac{L_j}{r}}{\operatorname{argmax}} \left[-C_B \cdot b_i - \beta \frac{F_j}{b_i} \right]. \tag{15}$$

Notice that this is equivalent to the maximization problem considered in Lemma 2, whose solution was $\sqrt{\beta \frac{F_j}{C_B}}$. From the concavity of the objective, it follows that the best strategy $b_i^* \leq \frac{L_j}{r}$ is $\min \left\{ \sqrt{\beta \frac{F_j}{C_B}}, \frac{L_j}{r} \right\}$.

Case 2: Second, consider $b_i \geq \frac{L_j}{r}$. In this case, the best backup strategy $b_i^* \geq \frac{L_j}{r}$ can be expressed as

$$b_i^* \in \operatorname*{argmax}_{b_i \geq \frac{L_j}{r}} \left[W_j - C_B \cdot b_i - \beta \frac{F_j}{b_i} - \beta V_j(a_1, a_2) \left(\frac{L_j}{b_i} + T_j \right) \right] \quad (16)$$

$$= \operatorname*{argmax}_{b_i \geq \frac{L_j}{r}} \left[-C_B \cdot b_i - \frac{\beta}{b_i} \left(F_j + V_j(a_1, a_2) L_j \right) \right]. \quad (17)$$

Note that the equation is correct even when $b_i = \frac{L_j}{r}$ (which implies $p_i^* = 1$), since we can substitute $\frac{L_j}{b_i}$ for r.

To find the best strategy $b_i^* \geq \frac{L_j}{r}$, we take the first derivative of the objective and set it equal to 0:

$$0 = -C_B + \frac{\beta}{b_i^2} \left(F_j + V_j(a_1, a_2) L_j \right) \quad (18)$$

$$b_i = \pm \sqrt{\beta \frac{(F_j + V_j(a_1, a_2) L_j)}{C_B}}. \quad (19)$$

Further, it is clear that the objective concave as the first derivative is strictly decreasing. Combining with the constraint $b_i \geq \frac{L_j}{r}$, this implies that the best strategy $b_i^* \geq \frac{L_j}{r}$ is $\max \left\{ \sqrt{\beta \frac{(F_j + V_j(a_1, a_2) L_j)}{C_B}}, \frac{L_j}{r} \right\}$.

Let's define $b_j^{\text{low}} = \sqrt{\beta \frac{F_j}{C_B}}$ and $b_j^{\text{high}} = \sqrt{\beta \frac{(F_j + V_j(a_1, a_2) L_j)}{C_B}}$. Note that it is obvious that $b_j^{\text{low}} \leq b_j^{\text{high}}$. If we have $b_j^{\text{low}} > \frac{L_j}{r}$, the organization's payoff is increasing in Case 1 and for Case 2, it is increasing in $[\frac{L_j}{r}, b_j^{\text{high}}]$ and decreasing in $[b_j^{\text{high}}, \infty]$. Hence, the organization's payoff is maximized in b_j^{high}. In a similar way, the organization's payoff is maximized at b_j^{low} when $b_j^{\text{high}} < \frac{L_j}{r}$. Otherwise, the organization's payoff is increasing in $[0, b_j^{\text{low}}]$, decreasing in $[b_j^{\text{low}}, \frac{L_j}{r}]$, increasing in $[\frac{L_j}{r}, b_j^{\text{high}}]$, and decreasing in $[b_j^{\text{high}}, \infty]$. Therefore, $b_i^* \in \left\{ b_j^{\text{low}}, b_j^{\text{high}} \right\}$ (the one that gives the higher payoff or both if the resulting payoffs are equal). $\square$

A.4   Proof of Lemma 4

The best-response ransom demand r^* is

$$r^* \in \operatorname*{argmax}_{r \in \mathbb{R}_+} \left[\sum_j \sum_{i \in G_j} V_j(a_1, a_2) \cdot r \cdot p_i^*(r) \right] - C_A \cdot (a_1 + a_2) - C_D \quad (20)$$

$$= \operatorname*{argmax}_{r \in \mathbb{R}_+} \sum_j \sum_{i \in G_j} V_j(a_1, a_2) \cdot r \cdot 1_{\left\{ r \leq \frac{L_j}{\bar{b}_j} \right\}} \quad (21)$$

$$= \operatorname*{argmax}_{r \in \mathbb{R}_+} \sum_j |G_j| \cdot V_j(a_1, a_2) \cdot r \cdot 1_{\left\{ r \leq \frac{L_j}{\bar{b}_j} \right\}}. \quad (22)$$

Clearly, the optimum is attained at either $\frac{L_1}{\hat{b}_1}$ or $\frac{L_2}{\hat{b}_2}$. Since we assumed that $\frac{L_1}{\hat{b}_1} \leq \frac{L_2}{\hat{b}_2}$, we have that $r = \frac{L_1}{\hat{b}_1}$ is a best response if and only if

$$(|G_1|\, V_1(a_1, a_2) + |G_2|\, V_2(a_1, a_2)) \frac{L_1}{\hat{b}_1} \geq |G_2|\, V_2(a_1, a_2) \frac{L_2}{\hat{b}_2} \tag{23}$$

$$|G_1|\, V_1(a_1, a_2) \frac{L_1}{\hat{b}_1} \geq |G_2|\, V_2(a_1, a_2) \left(\frac{L_2}{\hat{b}_2} - \frac{L_1}{\hat{b}_1} \right). \tag{24}$$

Further, an analogous condition holds for $r = \frac{L_2}{\hat{b}_2}$ being a best response, which concludes our proof. $\square$

A.5 Proof of Lemma 5

Recall that the attacker's expected payoff is

$$\mathrm{E}\left[\mathcal{U}_A\right] = \left(\sum_{j} \sum_{i \in G_j} V_j(a_1, a_2) \cdot p_i \cdot r \right) - C_A \cdot (a_1 + a_2) - C_D \cdot 1_{\{a_1 > 0 \text{ or } a_2 > 0\}}. \tag{25}$$

Consider that $a_1 + a_2 = a_{\mathrm{sum}}$ and r are given, and $a_{\mathrm{sum}} > 0$. Under these conditions, the attacker's best strategy is

$$a_1^* \in \operatorname*{argmax}_{a_1 \geq 0} \left(\sum_{j} \sum_{i \in G_j} V_j(a_1, a_2) \cdot p_i^*(r) \cdot r \right) - C_A \cdot (a_1 + a_2) - C_D \tag{26}$$

$$= \operatorname*{argmax}_{a_1 \geq 0} \frac{a_1}{D + a_{\mathrm{sum}}} |G_1| \cdot 1_{\left\{ r \leq \frac{L_1}{\hat{b}_1} \right\}} + \frac{a_{\mathrm{sum}} - a_1}{D + a_{\mathrm{sum}}} |G_2| \cdot 1_{\left\{ r \leq \frac{L_2}{\hat{b}_2} \right\}}, \tag{27}$$

giving the non-negative payoff. The best strategy can be calculated readily. $\square$

A.6 Proof of Proposition 1

Lemma 5 shows the attacker's best-response attack effort for fixed effort level, i.e., a_{sum}. In this Lemma, for example, $a_1^* = 0$ and $a_2^* = a_{sum}$ is the attacker's best-response effort if $|G_1| \cdot 1_{\left\{ r \leq \frac{L_1}{\hat{b}_1} \right\}} < |G_2| \cdot 1_{\left\{ r \leq \frac{L_2}{\hat{b}_2} \right\}}$ and the resulting attacker's payoff is non-negative. According to Lemma 4, the attacker's best-response ransom demand is either $\frac{L_1}{\hat{b}_1}$ or $\frac{L_2}{\hat{b}_2}$ and without loss of generality, we have assumed that $\frac{L_1}{\hat{b}_1} \leq \frac{L_2}{\hat{b}_2}$.

For this case, the attacker's payoff is equal to:

$$\mathrm{E}\left[\mathcal{U}_A\right] = \frac{a_{sum}}{D + a_{sum}} |G_2| \cdot r \cdot 1_{\left\{ r \leq \frac{L_2}{\hat{b}_2} \right\}} - C_A \cdot a_{sum} - C_D. \tag{28}$$

If the above equation is negative, i.e.,

$$r < \frac{(D + a_{sum})(C_A \cdot a_{sum} + C_D)}{a_{sum} \cdot |G_2| \cdot 1_{\left\{ r \leq \frac{L_2}{\hat{b}_2} \right\}}},$$

the attacker's best-response effort is $a_1^* = a_2^* = 0$. To satisfy the above condition, we replace r with $\frac{L_2}{\hat{b}_2}$, which gives

$$\frac{L_2 \cdot a_{sum} \cdot |G_2| \cdot 1_{\left\{r \leq \frac{L_2}{\hat{b}_2}\right\}}}{L_2\left(D + a_{sum}\right)\left(C_A \cdot a_{sum} + C_D\right)} < \hat{b}_2^*.$$

Further, the defender's best-response backup strategy when there is no attack, i.e., $a_1^* = a_2^* = 0$ is calculated based on Lemma 2. By inserting the value of $\hat{b}_2^*$ from Lemma 2, we can readily have the following:

$$\frac{L_2 \cdot a_{sum} \cdot |G_2| \cdot 1_{\left\{r \leq \frac{L_2}{\hat{b}_2}\right\}}}{L_2\left(D + a_{sum}\right)\left(C_A \cdot a_{sum} + C_D\right)} < \sqrt{\beta \frac{F_2}{C_B}}.$$

Another condition can be calculated similarly. □

B Further Numerical Illustrations

Figure 5 shows the payoffs of individual organizations from two groups as well as the attacker, for various values of the costs F_1 and F_2 of random-failure data losses. Figure 5(c) shows that the attacker's payoff *decreases* as these costs increase, which is in contrast with the cost-payoff relationship shown by Figure 4(c). The reason for this decrease is that organizations spend more on backups when the cost of random-failure data losses is higher, which in turn reduces the potential loss from ransomware attacks and, hence, the ransom that an attacker may demand.

We see more interesting phenomena in the organizations' payoffs. For most combinations of cost values, as expected, the payoff of an organization decreases gradually as the random-failure data loss cost in its group increases. However, for certain combinations of cost values (e.g., $F_1 = 4$ and $F_2 = 1$), increasing the cost of one group (e.g., group 1) *increases* the payoff of organizations in that group (see, e.g., Figure 5(a)) but *reduces* the payoff in the other group. This surprising phenomenon is due to the incentives that increased costs of random-failure data losses provide for organizations to invest more in backups. At certain tipping points, such as $F_1 = 4$ and $F_2 = 1$, even a small change in the costs can lead to a significantly different equilibrium, in which the focus of the attacker is shifted from one group to another. This substitution effect is very similar to the one exhibited in Figure 4.
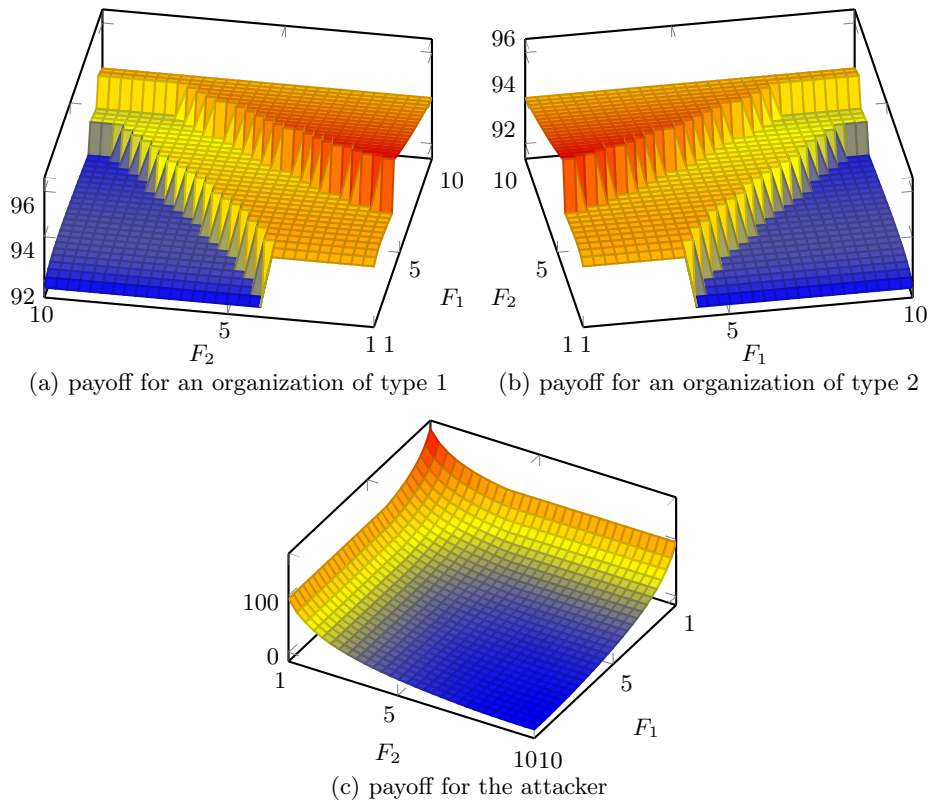
(a) payoff for an organization of type 1

(b) payoff for an organization of type 2

(c) payoff for the attacker

Fig. 5. The expected payoff of individual organizations of (a) type 1 and (b) type 2 as well as (c) the attacker in equilibrium for various costs of random-failure data losses F_1 and F_2.