DISCUSSION PAPER SERIES

DP17403

Cyber security and ransomware in financial markets

Toni Ahnert, Michael Brolley, David Cimon and Ryan Riordan

FINANCIAL ECONOMICS

CEPR

ISSN 0265-8003

Cyber security and ransomware in financial markets

Toni Ahnert, Michael Brolley, David Cimon and Ryan Riordan

Discussion Paper DP17403
Published 23 June 2022
Submitted 22 June 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Cyber security and ransomware in financial markets

Abstract

Financial markets are under constant threat of cyber attacks. We develop a principal-agent model of cyber-attacking with fee-paying clients who delegate security decisions to financial platforms. We derive testable implications about cyber attack vulnerability and fees charged. We also characterize the form of cyber attack chosen by attackers. Successful ransomware attacks are more likely than traditional attacks. When security is unobservable, platforms underinvest in security. Welfare can improve by targeting security investment through regulation (e.g. minimum security standards), or by improving transparency (e.g. security ratings). Our results support regulatory efforts to increase transparency around cyber security and cyber attacks.

JEL Classification: N/A

Keywords: N/A

Toni Ahnert - toni.ahnert@gmail.com
European Central Bank and CEPR

Michael Brolley - mbrolley@wlu.ca
Wilfrid Laurier University

David Cimon - dcimon@bankofcanada.ca
Bank of Canada

Ryan Riordan - ryan.riordan@queensu.ca
Queen's University

Acknowledgements
We thank Jason Allen, Kartik Anand, Shota Ichihashi, Marco Macchiavelli, and Sophie Moinas for thoughtful comments. We thank conference participants at the Northern Finance Association and seminar participants at Wilfrid Laurier University, the University of Toronto, the University of Hawaii, the Bank of Canada and the CFTC, as well as the staff from the Bank of Canada's Resolution and Crisis Preparedness team for their helpful feedback. We thank Michael Beckenhauer and William Wootton for excellent research assistance. David Cimon and Michael Brolley acknowledge financial support from the Social Sciences and Humanities Research Council, Grant No. 430-2019-00814. The views of the authors do not necessarily represent the views of the Bank of Canada or the European Central Bank.

Cyber security and ransomware in financial markets

Toni Ahnert Michael Brolley David Cimon Ryan Riordan

June 22, 2022

Abstract

Financial markets are under constant threat of cyber attacks. We develop a principal-agent model of cyber-attacking with fee-paying clients who delegate security decisions to financial platforms. We derive testable implications about cyber attack vulnerability and fees charged. We also characterize the form of cyber attack chosen by attackers. Successful ransomware attacks are more likely than traditional attacks. When security is unobservable, platforms underinvest in security. Welfare can improve by targeting security investment through regulation (e.g. minimum security standards), or by improving transparency (e.g. security ratings). Our results support regulatory efforts to increase transparency around cyber security and cyber attacks.

In modern markets, financial institutions often act as online platforms to facilitate transactions, store data, and provide advice; as their digital connectivity has grown, so to has their vulnerability to cyber attacks. For example, the FinCyber Project details over 150 notable incidents of cyber attacks and threats to global financial institutions since 2019.[1] The impact of cyber attacks is significant: in December 2021, attackers stole over $150 million from users at cryptocurrency exchange Bitmart. In March 2021, U.S. insurance firm CNA suffered an attack that disrupted client services and locked employees out of the network; two weeks later, CNA paid $40 million in ransom to regain control.[2] These instances exemplify two common categories of cyber attacks: (i) conventional attacks that steal client assets or data directly from the platform, and (ii) ransomware attacks that seize control of a platform's information technology system and hold it hostage until ransom payments are made.

Financial firms are especially susceptible to cyber risk because they not only sell services, but also provide intermediation. Clients of financial firms seek to avoid direct loss of assets, information, or access, while the platform wishes to maintain their business relationship with the client to earn fees. This relationship raises questions about the dangers that cyber attacks pose to financial institutions. Does the platform-client relationship impact cybersecurity investment, and ultimately the vulnerability of platforms to breaches? Does the introduction of ransomware technology improve or worsen the problem? Which platform types are most affected by ransomware? How might a regulator improve welfare, and does its solution depends on the type of attack (i.e., theft or ransomware)?

In this paper we construct a parsimonious model of the financial system with clients, financial infrastructure providers, and cyber attackers. In our model, clients delegate security decisions to these infrastructure providers, which we refer to as platforms. In turn, these

[1] Carnegie Endowment for International Peace: https://carnegieendowment.org/specialprojects/protectingfinancialstability/timeline

[2] Bloomberg: https://www.bloomberg.com/news/articles/2021-05-20/cna-financial-paid-40-million-in-ransom-after-march-cyberattack

platforms charge fees to their clients to arrange transactions and invest in security to protect client assets. Clients wish to maximize successful transaction volume net of these fees by allocating their transactions across several platforms. These platforms can be thought of as securities or cryptocurrency exchanges, banks, or payment providers and networks.

In our model, platforms are aware of the threat from cyber attackers, and may choose to invest in security to protect clients' assets. Cyber attackers attack platforms to either to steal client assets (e.g., the 2021 Bitmart theft) or disrupt client transactions until a ransom is paid (e.g., the 2021 CNA breach). The likelihood of attacker success increases with effort but decreases in platform security investment. Platforms, however, face diminishing returns to investment (Goyal and Vigier 2014). The model resembles a traditional "attacker-defender" game. However, unique to the financial setting, the defender (platform) is not the *owner* of the assets requiring protection. Cyber attackers target client transactions and these losses are borne primarily by clients (principal) and only partially by the platforms (agents) that assume the defender role. Herein lies a key friction in our model, in the form of a classic principal-agent problem: as cyber attackers target client transactions, these losses are borne by clients and not the platforms that assume the defender role.

In equilibrium, platform vulnerability depends on the relative cost of security and the value of the targeted asset to the attacker. The combination of a low cost of security and a low value of the targeted asset can lead to "full security", a state with zero expected breaches. However, if the marginal cost of security is sufficiently high relative to the attractiveness of the asset, the diminishing returns of security investment to the platform yields positive expected breaches. Moreover, the likelihood of breaches increases in both the security cost and target asset value dimensions. Our model predicts that, in the cross section, one would observe a higher (expected) number of security breaches in industries where i) security is more costly, or ii) the targeted assets have higher value to attackers. Interestingly, the size of the market and the degree of competition have no impact on platform vulnerability, as an

increase in the magnitude or concentration of assets at any one platform attracts attackers but also increases the benefits to platforms from enhanced security.

Beyond the realm of conventional attacks that target client assets, our model allows us to address the recent phenomena of *ransomware* attacks. In practice, ransom attacks can take many forms, including the encryption of files or the take-over of critical systems. By demanding a ransom payment from the platform, often using some form of cryptocurrency, the attackers change participants' incentives within the model: instead of targeting clients directly by attacking platforms with the intent to steal client assets, attackers block transactions, thereby depriving the platform operator of transaction based-revenues.

Because clients rationally believe that platforms will pay ransoms to protect their assets—as CNA did[3]—they ignore the degree to which a platform is vulnerable to ransomware attacks. Consequently, in the ransomware case of our model, clients main concern is platform fees. In equilibrium, we find that ransom attacks are more likely to be successful, driven by a lower level of security investment (relative to the conventional attack equilibrium) that, in turn, results in a lower fee to attract clients. We also show that measures like insurance against ransomware attacks may increase the likelihood of these attacks by documenting the willingness and ability of platforms to pay ransoms. Indeed, the FBI has cautioned against paying ransoms for this very reason, as hackers target firms with cyber insurance policies in anticipation that they are more likely to pay.[4] We show that if platforms can commit to not paying ransoms the incidence of successful cyber attacks is reduced.[5]

In our model, regulating a minimum security spending does not always reduce breaches or improve welfare. When security levels are publicly observable, competition amongst platforms leads to security investment that is constrained efficient; that is, identical to the

[3]Gavin Souter, *Business Insurance,* November 1, 2021: "CNA's insurance won't cover all of ransomware loss"

[4]Bloomberg: https://www.bloomberg.com/news/articles/2021-05-20/cna-financial-paid-40-million-in-ransom-after-march-cyberattack

[5]Alternatively, a regulator who is able to influence the underlying parameters of the model may improve welfare by lowering the value of ransoms, such as by banning ransom payments by insurers (or ransom payments itself), or increasing surveillance of cryptocurrency transactions. However, vulnerability will not decline, as security spending also declines.

social planner's outcome in which they cannot dictate the attack intensity of the cyber attacker. We take this constrained efficient outcome as the welfare benchmark throughout the paper. Perhaps more realistically, when clients cannot observe platform security investment, the resulting equilibrium does not achieve constrained efficiency. In this environment, a regulator may achieve constrained efficiency through two channels. First, by regulating security investment through appropriate minimum security standards, a regulator could ensure platforms invest to the constrained efficient level. Second, a regulator may improve transparency by requiring platforms to obtain a signal of security investment, similar to security ratings offered by firms like BitSight and UpGuard.[6] Such a policy appears palatable, as the Securities and Exchange Commission in the U.S. is currently considering a proposal to improve cybersecurity transparency through mandating cybersecurity disclosures by public companies. A key goal, SEC Chair Gary Gensler argues, is to "strengthen investors' ability to evaluate public companies' cybersecurity practices and incident reporting."[7]

One additional highlight of our paper, in relation to the welfare implications of unobservable security, is the importance of the interaction between the attacker and defender (the platform) game, and the principal-agent problem. In our model, we show that security costs and asset value (to the attacker) impact which attack mode–conventional or ransomware–that an attacker will use in equilibrium. This choice is, in turn, affected further by the transparency of security investment, but interestingly, only in the case of conventional attacks. That is, security investment against ransomware attacks is independent of the *transparency* of security investment, whereas conventional attacks lead to under investment in security when that investment decision is unobservable. As a result of this under investment, it is optimal for attackers to use conventional attacks for a larger region of the parameter space

[6] BitSight and UpGuard are two U.S.-based firms that provide third-party cyber risk management services to firms and institutions. Both companies offer an assessment of a firm's vulnerability to cyber attacks, which is summarized by a numerical "security score". These systems are similar in design to a credit ratings score.

[7] See SEC Statement on Proposal for Mandatory Cybersecurity Disclosures, March 9, 2022: https://www.sec.gov/news/statement/gensler-cybersecurity-20220309

when security investment is unobservable. Because conventional attacks under unobservable security do not obtain constrained efficiency, welfare is thus lower.

The model is robust along several dimensions. First, similar results arise in a monopolist setting: the monopoly platform chooses the same welfare-maximizing level of security investment as the aggregate security investment in the competitive equilibrium. However, the monopolistic platform charges higher fees to clients. Second, when attackers are subject to a budget constraint, or platforms face increasing marginal costs to providing security, the probability of successful attacks may decrease as competition increases. These extensions of our model suggest that, similar to other common reasons for competition policy, one regulatory response may be to increase financial market competition by breaking up monopolies, or by supporting firm entry.

Related Literature. Our paper contributes to the literature on cyber crime. On the surface, the incentives to commit cyber crimes are not different to the conventional crimes studied in Becker (1968). Cyber crime is unique, however, in that securing information systems is expensive relative to the cost of attacking—attackers often do not have to leave the comfort of their own homes. In the context of Automated Telling Machines, a literature in computer science (see e.g., Anderson (2001); Anderson and Moore (2006); Anderson et al. (2013)) highlights the economic incentives of cyber security. They show that when banks are liable for losses, security is higher. Hal Varian highlighted a similar incentive problem with anti-virus software:[8] customers are willing to spend $100 to protect their own systems, but unwilling to spend even $1 to protect the systems of others. Other areas of computer science and information technology have discussed the security investment problem in different contexts: as a user's responsibility (August and Tunca 2006); as a profit maximization problem (Dynes et al. 2007); as a function of the importance of a vulnerability (Gordon and Loeb 2002); in the presence of state actors with almost infinite resources (Anderson 2001).[9]

[8] www.nytimes.com/library/financial/columns/060100econ-scene.html

[9] See Manshaei et al. (2013) for a survey on the earlier economics of security literature.

We contribute to this literature by modelling the principal-agent problem inherent in the provision of security by platforms that manage client assets, transactions, and data.

Investigating the true costs of cyber crime has proven challenging (e.g., Anderson et al. (2013), Biancotti (2017) and Paquet-Clouston et al. (2018)). Anderson et al. (2013) argue that direct losses are equal to spending on security, but indirect losses (reputation) may be larger. Moreover, Eisenbach et al. (2021) sheds light on the spillover effect of cyber attacks on the U.S. financial system, where an attack on one of the major banks in the country would negatively impact almost a third of the wholesale payment network between U.S. financial institutions. Recent papers have sought to quantify some of these indirect and spillover costs. Florackis et al. (2022) describe cyber risk at a firm level and show that increased cyber risk is associated with higher equity returns. This suggests that firms, or industries, that are more susceptible to cyber attacks will have higher costs of capital. Relatedly, Kamiya et al. (2021) find evidence that following a successful cyber attack, the decrease in shareholder wealth exceeds that of the out-of-pocket costs of the attack; Crosignani et al. (2022) document the contagion effects of a large-scale attack, showing that the costs of such an attack reach far beyond the targeted firm. Kotidis and Schreft (2022) highlight the importance of bank contingency planning towards mitigating spillover effects when attacks continue over a period of days. In our paper, we analyze the impact of cyber crime on welfare by accounting for several key sources of welfare loss, including direct theft from market participants, the need to invest in security, and the loss of business.

Recently, cyber attacks have shifted to use ransom-type threats. Foley et al. (2019) estimate that 46% of all Bitcoin transactions are related to illicit activity, including the payment of ransoms in ransomware attacks, and Paquet-Clouston et al. (2018) describe the use of the Bitcoin network as a means to pay ransom in cyber attacks. Most recently, August et al. (2022) examine the welfare implications of a ransom option to victims versus conventional attacks, the outcome of which depends non-monotonically on the level of risk of an attack. Laszka et al. (2017) model the response of an industry of firms to the threat

of ransomware attacks. We contribute to the literature of ransom-based cyber attacks by providing what is, to the best of our knowledge, the first model to analyze their impact on firm security investment, platform vulnerability, and the subsequent welfare effects. Moreover, our work i) contrasts the efficacy of ransomware attacks against traditional cyber attacks, showing that, in equilibrium, ransomware attacks are more likely to succeed, and ii) evaluates under what conditions attackers may prefer ransomware to conventional attacks.

From a mitigation perspective, cyber-insurance as a solution has been widely studied; see, for example, Grossklags et al. (2008), Böhme and Schwartz (2010), Johnson et al. (2011), Moore (2010) and Massacci et al. (2017) discuss solutions such as cyber insurance. Our model suggests that cyber insurance is either ineffective; or worse it increases the overall probability of successful attacks.

Finally, our model belongs to the broad class of attack and defender games in the economics literature. This literature, however, focuses primarily on the structure of networks. For example, Bier et al. (2007), Dziubiński and Goyal (2013), Goyal and Vigier (2014), Acemoglu et al. (2016), Hoyer and de Jaegher (2016) and Kovenock and Roberson (2018) analyze the incentives for attackers and defenders who expend resources to secure nodes and the entire network. Duffie and Younger (2019) explores the repercussions of a cyber-run on 12 of the largest U.S. financial institutions. In our paper, we depart from the network-structure setup and assume a single point-of-failure or "weak point". We do not model the contagion that hacking in one institution could have on other institutions. The simplification allows us to tractably study the principal-agent problem inherent in many financial market applications, and one that is core to our particular attacker-defender game. In our model, it is not only the service providers who fear an attack, but also clients that rely on the service.

Our analysis in the paper proceeds as follows. In Section 1, we lay out the basic model. In Section 2 we outline the equilibrium. We extend the model in Section 3 to include the case of ransomware. We examine unobservable security investment in Section 4. We explore robustness checks in Section 5 and conclude in Section 6.

1 Model

There are three dates $t = 0, 1, 2$, a single good, and three types of risk-neutral agents: clients, platforms, and an attacker. At date 0 platforms invest in cyber security and set a fee for transacting on the platform. At date 1 clients allocate their transaction needs across platforms. At date 2 the attacker chooses the intensity with which to attack each platform.

Platforms. At $t = 0$, $N \geq 2$ platforms indexed by $i = 1, ..., N$ each simultaneously invest $s_i \geq 0$ in the security of their platform (e.g., hiring an information systems analyst, implementing biometric identification and/or multi-factor authentication) and choose a fee $f_i \geq 0$ per unit of transaction on the platform.[10] Each platform maximizes its expected profits

$$\pi_i = (1 - \delta_i) f_i Q_i - c s_i, \tag{1}$$

where δ_i is the probability of a successful attack on platform i, Q_i is the transaction volume on platform i, and c is the cost of security investment. If the platform is successfully attacked, we refer to it as having been *breached*. We assume that the platform receives the fee only if the transaction is completed successfully, which occurs with probability $1 - \delta_i$.[11] The platform's payoff therefore corresponds to the expected probability that it is not breached, multiplied by the total fees it charges, net the cost of its security investment.

Following Goyal and Vigier (2014), an attack on platform i is successful with probability

$$\delta_i = \delta(a_i, s_i) = \frac{a_i}{a_i + s_i} \tag{2}$$

if $s_i + a_i > 0$ and 0 otherwise, where a_i is the attack intensity. A higher attack intensity increases the probability of a breach, $\frac{d\delta_i}{da_i} > 0$, while higher security lowers it, $\frac{d\delta_i}{ds_i} < 0$.

[10] We consider the case of a monopoly platform in Section 5.

[11] This assumption can be viewed as either a refund requested by the client when the platform does not adequately perform its function or a loss of future business by the client in reduced form.

Clients. A mass M of identical clients indexed by m have exogenous transaction needs Q_m at $t = 1$.[12] They simultaneously allocate their transactions across platforms, where $q_{im} \geq 0$ is the transaction volume of client m on platform i and $Q_m = \sum_{i=1}^{N} q_{im}$. We normalize $Q_m \equiv 1$ without loss of generality. Thus, the total market size for all platforms is M and the transaction volume of platform i is $Q_i \equiv \int_0^M q_{im}\,\mathrm{d}m$. Each client maximizes her expected utility

$$U_m = \sum_{i=1}^{N}(1-\delta_i)(1-f_i)q_{im}, \tag{3}$$

where the client enjoys the value of successful transactions net of fees.

Attacker. We start the analysis with conventional attacks and turn to ransomware attacks in Section 3. The attacker wishes to disrupt the clients' transactions and steal a portion of their value. The attacker chooses the attack intensity $a_i \geq 0$ at $t = 2$ to maximize his payoff

$$\pi_A = r\sum_{i=1}^{N}\left(\delta_i Q_i - a_i\right), \tag{4}$$

where $r \in (0, 1]$ is the portion of the transaction the attacker is able to steal (his reward).[13]

We offer two interpretations of r. First, we can interpret r as the inherent ease of stealing the asset or a recovery rate. For example, records of physical asset ownership may have an ease parameter $r = 0$ as even if the records are stolen or corrupted, back-up copies exist which prevent the transfer. Digital assets (e.g., crypto wallet addresses and banking information) on centralized systems may have a higher r, by contrast, as digital records of asset ownership may be accessed and transactions authorized and cleared before the hacker can be intercepted. The 2021 attack on Bitmart, and other similar incidents at cryptocurrency exchanges, are prime examples of realized attacks at venues where r may be high and digital assets can be taken.

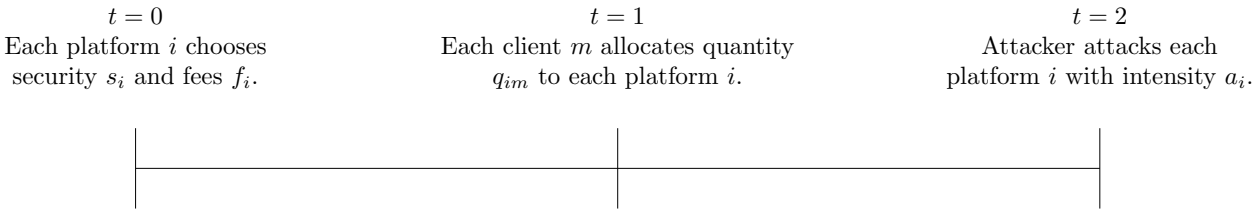
[12] Our approach effectively assumes that the demand for transactions is insensitive to cyber risk and studies its effect on the supply side instead. It is similar in nature to the assumptions in Parlour (1998), where investors transact in fixed size.

[13] While the reward r is exogenous, both the transaction volume Q_i and the probability of a successful attack δ_i are endogenous in our model.

Second, we can interpret r as the relative value of data or assets that can be stolen or the ease with which they can be monetized by the hacker. This interpretation reflects the disparity between the transaction that the client wishes to complete, and the value of the asset truly at risk. For example, personal data may offer a hacker the *possibility* of stealing all of a client's assets, but in practice the hacker may not be able to realize the full value of the data before the platform recognizes the breach and denies access.[14]

Timeline. The game proceeds as shown in Figure 1.

Figure 1: Timeline of events.



Definition 1 (Equilibrium.) *An equilibrium of this game is given by a_i^*, s_i^*, f_i^*, and q_{im}^* for all $i = 1, ..., N$ and $m \in [0, M]$ and is found via backward induction:*

1. *At $t = 2$, the attacking strategy on platform i is $a(Q_i, s_i) = \arg\max_{a_i} \pi_A$, for any Q_i and s_i.*

2. *At $t = 1$, the transaction allocation strategy is $q_{im}(\mathbf{s}, \mathbf{f}) = \arg\max_{q_{im}} U_m$ subject to $\sum_{i=1}^{N} q_{im} = Q_m$ and the attacking strategies $a(Q_i, s_i)$, for any $(\mathbf{s}, \boldsymbol{f}) \equiv \{f_i, s_i\}_{i=1}^{N}$. (Note that $Q_i(\mathbf{s}, \boldsymbol{f}) = \int_0^M q_{im'}(\mathbf{s}, \boldsymbol{f})\, dm'$ is independent of m because each client has zero mass.)*

3. *At $t = 0$, $(\mathbf{s}^*, \mathbf{f}^*)$ is a Nash equilibrium among the platforms. That is, $(s_i^*, f_i^*) = \arg\max_{s_i, f_i} \pi_i$, given the choices of the other platforms (s_{-i}, f_{-i}), the allocation strategies of clients $q_{im}(\mathbf{s}, \mathbf{f})$, and the attacking strategies $a(Q_i, s_i)$.*

[14] In an extended version of the model, an attacker may receive non-pecuniary benefits (e.g., from the disruption of transactions), one could interpret r as a sum of financial gains and non-pecuniary benefits (i.e., a high value of r represents a combination of a high degree of financial and non-financial motivations). This interpretation may be particularly relevant for some state-sponsored cyber attacks.

4. *Thus,* $q_{im}^* = q_{im}(\mathbf{s}^*, \mathbf{f}^*)$ *and* $a_i^* = a(Q_i^*, s_i^*)$.

2 Equilibrium

Attack intensity. We first solve the problem in (4) to obtain the attacking strategy, which is an attacking intensity on platform i for any security investment s_i and transaction volume Q_i:

$$a(Q_i, s_i) \equiv \begin{cases} \sqrt{s_i r Q_i} - s_i & \text{if } s_i \leq rQ_i \equiv \overline{s}_i \\ 0 & \text{if } s_i > \overline{s}_i. \end{cases} \tag{5}$$

Intuitively, the attacker chooses to attack with positive intensity when the share of the platform's transaction volume that can be stolen, rQ_i, is high enough relative to the security level s_i. Moreover, the attack intensity increases in the platform's transaction volume Q_i and the share that can be stolen, r.

Using the attacking strategy $a(Q_i, s_i)$, the probability of a successful attack on platform i reduces to a function of platform security investment and its transaction volume:

$$\delta(Q_i, s_i) \equiv \delta(a(Q_i, s_i), s_i) = \begin{cases} 1 - \sqrt{\frac{s_i}{rQ_i}} & \text{if } s_i \leq \overline{s}_i \\ 0 & \text{if } s_i > \overline{s}_i. \end{cases} \tag{6}$$

Equation (6) highlights that a sufficiently high level of security investment, $s_i \geq \overline{s}_i$, deters any attacks. The investment threshold required to deter all attacks, $\overline{s}_i$, increases in r and Q_i, the components of the reward to the attacker from a successful attack (Equation 5).

Allocation of transactions across platforms. To maximize her utility, each client allocates her transaction needs among the platforms offering the highest value

$$(1 - \delta(Q_i, s_i))\,(1 - f_i),$$

which is the highest cyber risk-adjusted return net of fees. The client allocates zero to all other platforms. We consider a symmetric equilibrium in which clients allocate their transaction needs equally across platforms with the best deal for clients. When $n \geq 1$ platforms within P offer the highest cyber risk-adjusted net return, then $q^*_{im}(\mathbf{s}, \mathbf{f}) = \frac{1}{n(\mathbf{s},\mathbf{f})} \mathbf{1}_{\{i \in P(\mathbf{s},\mathbf{f})\}}$.

Security investment and fees. Given the the allocation strategy of clients, $q_{im}(\mathbf{s}, \mathbf{f})$, and the attacker's intensity strategy, $a(Q_i, s_i)$, each platform chooses s_i and f_i to maximize the expected profits in Equation (1) subject to attracting a positive market share

$$(1 - \delta(Q_i, s_i))(1 - f_i) \geq \zeta, \tag{7}$$

where $\zeta = \max_{j \neq i} (1 - \delta(Q_j, s_j))(1 - f_j)$ is the best offer made by another platform. Competition for positive market share by platforms leads to a Bertrand-style 'race-to-the-bottom' competition in fees such that each platforms earns zero expected profit in equilibrium.[15] At equilibrium, lowering fees would make the platform more attractive to clients and would earn the platform the entire market share; however, this platform would earn negative profits. Similarly, increasing fees would result in the platform earning zero market share and therefore zero profits.

We consider a symmetric equilibrium in which i) all platforms invest identically in security, $s^*_i = s^*$, and offer identical fees, $f^*_i = f^*$; ii) clients allocate $\int_0^M q^*_{im} = \frac{M}{N}$ to each platform, and iii) the attacker chooses the same intensity on all platforms, $a^*_i = a^*$. We are ready to state our first main result, which is proven in Appendix A.1.

Proposition 1 (Competitive Equilibrium for conventional attacks.) *A symmetric equilibrium exists and is unique. It is characterized by Equations (8)-(11).*

[15] The dual problem in which each platform chooses s_i and f_i to maximize $(1 - \delta(Q_i, s_i))(1 - f_i)$ subject to earning at least zero expected profits yields the same allocation.

Solving the platforms' problem, the equilibrium security investment is

$$s^* = \begin{cases} \frac{rM}{N} & \text{if } 2rc \leq 1 \\ \frac{M}{4c^2rN} & \text{if } 2rc > 1, \end{cases} \tag{8}$$

and the equilibrium fee is

$$f^* = \begin{cases} rc & \text{if } 2rc \leq 1 \\ \frac{1}{2} & \text{if } 2rc > 1. \end{cases} \tag{9}$$

As shown in Figure 2 (shown in the Appendix), the equilibrium security investment s^* and fee f^* exhibit a kink at $2rc = 1$, which depends on the relative value of attacking r and the cost of security c. Its importance is highlighted by the the same kink in attack intensity a^*, where the attacker chooses not to attack when the value of $2rc$ is sufficiently low. The corresponding attack intensity on a platform demonstrates a similar structure:

$$a^* = \begin{cases} 0 & \text{if } 2rc \leq 1 \\ \frac{M}{2cN}\left(1 - \frac{1}{2rc}\right) & \text{if } 2rc > 1. \end{cases} \tag{10}$$

Then (a^*, s^*) yields the equilibrium vulnerability as the probability of a successful attack is:

$$\delta^* = \begin{cases} 0 & \text{if } 2rc \leq 1 \\ 1 - \frac{1}{2rc} & \text{if } 2rc > 1. \end{cases} \tag{11}$$

Equation (11) states that once the combination of r and c are low enough, $2rc \leq 1$, platforms choose to secure client transactions fully (see also Figure 3a). In this case, the combination of the cost of security c and the value of the asset to the attacker r are sufficiently low such that the platforms choose a level of security investment that prevents attacking completely. In this region, the cost of security does not factor directly into s^* (Equation 8): for any marginal difference in c, investment is already such that $\delta^* = 0$, so no change

in investment occurs. However, for a marginal increase in r, the increase in the value of attacking requires greater security investment by the platform to maintain $\delta^* = 0$. Moreover, fees increase only with the costs of security, c, and the reward, r.

When $2rc > 1$, however, the combination of relative asset value to the attacker and the cost of security are higher, so it is not optimal for platforms to invest in a level of security that would eliminate all attacks. Instead, platforms maintain a constant optimal fee level $f^* = 1/2$, and opt for a positive value of δ^*.

2.1 Comparative statics

Our model produces several results for how parameters affect the equilibrium. In particular, we emphasize the i) cross-section of market tightness (clients-per-platform), ii) cost of security (relative to the unit marginal cost of attacking), and iii) the degree to which an attacker can monetize the transactions they interrupt (e.g., by stealing valuable assets). In what follows, we focus on the impact of these parameters on platform vulnerability (δ^*) and fees (f^*) because these variables have direct empirical counterparts. Proposition 2 provides a summary of the comparative statics in $(M/N, r, c)$, for all equilibrium values $(\delta^*, a^*, s^*, f^*)$.

Proposition 2 (Comparative statics for conventional attacks.) *The impact of marginal increases in parameters* $(\frac{M}{N}, r, c)$ *on the equilibrium outcomes* $(\delta^*, a^*, s^*, f^*)$ *are given in the table below, where arrows indicate increasing or decreasing.*

	M/N		r		c		
	$2rc \leq 1$	$2rc > 1$	$2rc \leq 1$	$2rc > 1$	$2rc \leq 1$	$rc \in (0.5,1)$	$rc \geq 1$
Vulnerability (δ^*)	0		0	↑	0	↑	
Attack intensity (a^*)	0	↑	0	↑	0	↑	↓
Security Investment (s^*)	↑		↑	↓	0	↓	
Fees (f^*)	0		↑	0	↑	0	

We start by investigating as the ratio, M/N. Higher market tightness reflects an increase in the mass of client transactions at a (representative) platform. In equilibrium, market tightness has no impact on the probability of a breach δ^*, nor on fees f^*. The former is driven

by counteracting investments in attack intensity and cyber security: the attacker earns more per successful attack as the mass of transactions at each platform increases, leading to an increase in attack intensity, while platforms respond with an increase in security to protect the additional fees earned from a higher transactions volume.

Focusing on per-platform total investment, Equation (8) highlights that security s^* increases linearly in the mass of transactions (M) and decreases in competition (N), but this implies that total market security per transaction, $\frac{N}{M}\, s^*$, is invariant to changes in N and M: platforms increase security linearly as M grows to protect those additional transactions, but the aggregate security investment per transaction remains unchanged. Moreover, additional venues induce each platform to reduce their security investment, as their share of total transactions declines so that their security investment per transaction remains constant.

A key factor in the attacker's choice is the value of transactions to the attacker, r. It plays an important role in platform vulnerability, separating security investment into two regions: investment under low r and high r. When r is low, the platform increases security investment for any increase in r, funded via higher fees. Maintaining this level of security keeps attacking from being profitable, ensuring a fully-secured platform. For platforms that manage transactions with relatively higher value to the attacker r, however, platforms favour competition in fees as r increases versus fully securing against attackers. Moreover, this competition in fees does not merely halt security investment as r increases, but even *reduces it* (see s^* in Figure 2a) to to keep fees constant. The main driver of the reduction in security investment is the reduction in its effectiveness: as the attacker increases its attack intensity, any additional investment in security has decreasing returns on vulnerability, yet the marginal cost of security is constant. Hence, when the attacker derives a relatively large payoff from successful attacks (high r), equilibrium security levels are lower, all else equal. Taken together, higher r coincides with greater platform vulnerability δ^* (Figure 2a).

Following the discussion of the value of the transactions to the attacker, the marginal cost of security also plays a key role in the platform's investment decision (Figure 2b). Generally,

when the cost of security is sufficiently low ($2rc \leq 1$), security investment is invariant to changes in c. At this level, a platform fully secures all transactions, and funds any marginal increase in the cost of security through an increase in fees. Clients earn a higher utility from security versus their disutility from fees at these cost levels. On the other side of the threshold ($2rc > 1$), the convex nature of returns to security lead clients to prefer the risk of successful attacks to higher fees. Hence, an increase in c leads to a drop in s, and an increase in platform vulnerability (δ), while fees remain constant.

We also highlight the ambiguous sign of how the attack intensity a^* changes in c. When platforms respond to higher marginal security cost by decreasing security investment, the impact on attack intensity is concave: a c closer to the threshold $2rc = 1$ leads the attacker to take advantage of the lower security investment by increasing its intensity. As c reaches higher values, however, the marginal gain δ^* from additional attack intensity is lower than the marginal cost of attacking, leading the attacker to reduce their total investment in attacking, as the marginal gains to lower attack spending exceed the foregone attack effectiveness.

2.2 Welfare and Regulation

We next compare the allocation of the competitive equilibrium to welfare benchmarks. We consider a social planner who maximizes utilitarian welfare that aggregates all agents' profits and utilities, yielding

$$W \equiv \sum_{i=1}^{N} \left((1 - \delta_i) \, Q_i - c s_i \right), \tag{12}$$

where we assume that the planner does not take into account the utility of the attacker. This assumption can proxy for the attacker being a foreign agent, where the planner intends only to maximize the welfare of its domestic constituents, thereby placing no value on the welfare of agents outside of the home country. Alternatively, this may also proxy for the social value that a society places on the welfare of those who gain through criminal means.

Second Best. In the second-best problem, the planner cannot choose the actions of the attacker. That is, a (constrained) planner takes the attacking strategy of the attacker as given (as do private agents). We use the second-best allocation as the main benchmark throughout the paper. One can show that the second-best outcome (SB) is identical to the competitive allocation:

$$s^{SB} = s^*, \qquad a^{SB} = a^*, \quad \delta^{SB} = \delta^*. \tag{13}$$

It is intuitive that the constrained efficient outcome follows the competitive equilibrium, given the nature of competitive platforms' optimization problem. The competitive platform's optimal security choice can be seen as the value which maximizes consumer utility, subject to its own zero profit condition. This is similar in nature to the planner's problem, who optimizes the expected value of transactions, minus the cost of security investment.

This result allows us to evaluate a common regulatory requirement for financial institutions: minimum industry security protocols and standards. A good example is the requirement for all communications to be secured with 256-bit encryption technologies. The previous encryption technology, 128-bit, would take roughly $1.02 * 10^{18}$ years to crack and 256-bit encryption would take $3.06 * 10^{56}$ years to crack, with today's computational power.[16]

In the context of our model with perfect information, when clients are able to view platforms' security investment, industry standards or minimum required investments are either irrelevant or inefficient. In the case where the required minimum investment is below the equilibrium investment, the minimum investment is irrelevant. If enforcing and monitoring the minimum investment is costly, then it could increase the costs of providing security without improving or increasing security. In the case where the industry standard is above the equilibrium investment, the minimum investment requirement is inefficient as it leads to

[16] See https://www.ubiqsecurity.com/blog/128bit-or-256bit-encryption-which-to-use/, "...even a 128-bit key is secure against attack by modern technology. At its peak, the Bitcoin network ... performed approximately $150 * 10^{18} \approx 2^{67}$ operations per second. ...it would take the Bitcoin network over 70,000,000,000,000,000,000,000,000 years to crack a single AES-128 key."

over-investment in security. The minimum investment is above what customers are willing pay given the value they derive from transacting and their potential losses.

In sum, the industry standard 256-bit encryption may be an example of over-investment, as neither 128-bit nor 256-bit encryption is likely to be cracked within the useful lifetime of the information being secured. However, we show in Section 4 that when clients are unable to observe platform security investment, minimum security standards can play a useful role.

Having established the equilibrium and its constrained-efficiency property, we can express welfare as a function of parameters in this economy:[17]

$$W^{SB} = \begin{cases} M(1-rc) & \text{if } 2rc \leq 1, \\ \frac{M}{4rc} & \text{if } 2rc > 1. \end{cases} \tag{14}$$

Welfare is independent of the number of platforms N and a regulator gains no benefit from increasing competition. Welfare is, however, decreasing in both r and c. Thus, a regulator may be able to influence these parameters in the long-run. For example, regulators could encourage investments in cyber security research or competition within the IT sector. Investments in IT research, or encouraging competition within the IT sector may decrease the cost of security, c, in the long-run. Such a decrease results in welfare gains for the economy. Decreasing the value that an attacker can extract from the transaction r, can also increase welfare. An decrease in r could represent a number of factors which discourage attacks, for example, anti-money laundering controls that verify transactions, increased international enforcement which removes criminal safe havens, or transactions which are reversible within a specified period of time.

To conclude, a government which is able to influence both r and c in the long-run decreases realized attacks, accompanied by smaller total spending on IT security by financial platforms.

[17] Welfare is the same in the competitive equilibrium and under monopoly, as we outline in Section 5.

3 Ransomware

In this section we extend the model to ransomware attacks. High profile ransomware attacks have proliferated in recent years, both inside and outside of the financial system. In March of 2021, U.S. insurer CNA was forced to pay a ransom of $40M to regain control of their systems. Similarly, ransom attack on Colonial Pipeline led to fuel shortages on the U.S. east coast until a $4.4M ransom was paid.[18]

In this section, the attacker has access to ransomware (on top of the conventional attacking method) at an initial date, $t = -1$, at which the attacker decides on either a ransom attack or a conventional attack. This choice is based on the attacker's expected profits. For simplicity, we assume that the choice of attack mode is observable and that the attacker commits to it (e.g., by incurring some unmodelled setup costs for either technology). Real world examples of a commitment could include purchasing exploit kits or establishing an infrastructure to receive ransom payments.[19]

At $t = 2$ the attacker again conducts their attack. If the attack is conventional, the outcome is as presented in Section 2. In a ransom attack, however, the attacker can extract a ransom of ρQ_i from each platform i if their attack is successful, where $\rho \in [0, 1]$ represents how much of a ransom an attacker can successfully extract as a share of the platform's business. A higher value of ρ could represent (i) platforms with a high value of cyber insurance; (ii) platforms with large cash reserves; or (iii) platforms with low bargaining power against ransom attackers. Thus, an attacker's expected profit from a ransom attack is

$$\pi_A = \rho \sum_i (\delta_i Q_i - a_i) . \tag{15}$$

Clients facing a ransom attack have a modified utility function, depending on whether the platform will pay a ransom or not. If the platform is successfully attacked and does not pay

[18] See https://www.bloomberg.com/news/articles/2021-06-04/hackers-breached-colonial-pipeline-using-compromised-password

[19] See https://www.bloomberg.com/features/2020-dark-web-ransomware/ for a narrative example from the press of the steps in establishing a ransomware setup.

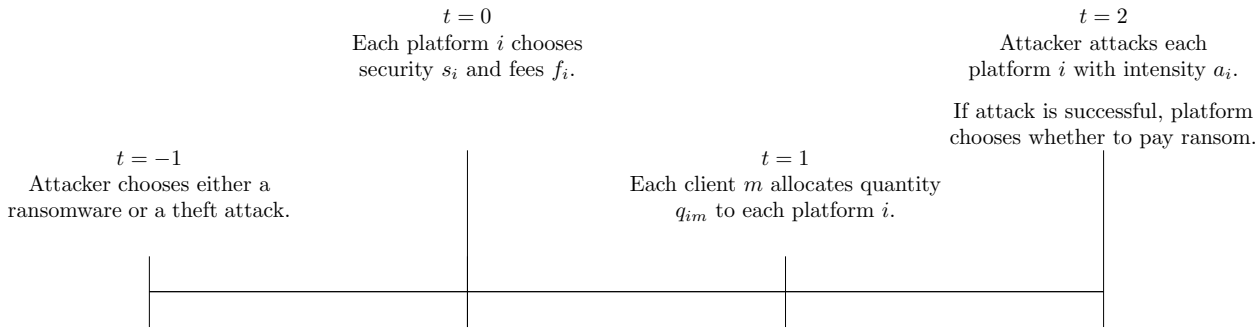
a ransom, clients do not realize the value of their transaction. Thus, clients at platforms that do not pay ransoms have utility identical to the model in Section 1. Alternatively, clients at platforms that pay ransoms to attackers realize their transaction's utility regardless of whether an attack is successful. These clients have a utility of

$$U = (1 - f_i)Q_i, \tag{16}$$

which is independent of δ_i. That is, when the platform pays a ransom, the clients' transactions occur regardless of whether the attack is successful. Thus, clients choose to pool at the platform(s) with the lowest fees.

When faced with a ransom attack, the platform faces two choices (Figure 6). The first is the choice of security s_i and fees f_i at $t = 0$. The second is the choice of whether to pay the ransom if successfully attacked at $t = 2$. A platform that pays a ransom earns the fees from its transaction, f_iQ_i, but loses the value of the ransom, ρQ_i. Alternatively, a platform that does not pay the ransom does not earn fees, but also incurs a loss term LQ_i. The loss term L represents several costs borne by a platform above and beyond the fees earned on transactions. These include costs associated with a long-term shutdown of its systems, reputational costs, the loss of all organizational data, and the costs of rebuilding IT systems.

Figure 6: Timeline of events with two attack types.



The nature of the equilibrium depends on the relative magnitudes of L and ρ. For $L < \rho$, there exists an equilibrium such that it is not optimal for platforms to pay ransoms. In

this equilibrium, were the attacker to select ransoms, the platform would select $s^*_{Ransom} = 0$ and $f^*_{Ransom} = 0$. While the attacker could breach the platform, the platform would pay no ransom, and thus the attacker would select $a^*_{Ransom} = 0$, resulting in $\pi_{A_{Ransom}} = 0$. In this case no attacker would select ransoms, and only theft attacks would occur, identical to the equilibrium presented in Section 2. Therefore, for the remainder of this section, we assume that $L \geq \rho$, so it is always optimal for a platform to pay a ransom once breached at $t = 2$. Put another way, the platform is not able to credibly commit to not paying a ransom once it has been successfully attacked.

Given that a platform pays the ransom at $t = 2$, it chooses s_i and f_i at $t = 0$ to maximize:

$$\pi_i = f_i Q_i - \delta_i \rho Q_i - c s_i \quad \text{s.t.} \tag{17}$$

$$1 - f_i \geq \zeta. \tag{18}$$

The constraint is similar to the case of a conventional attack, whereby clients pool at platforms that offer the highest utility.

Using backward induction, we solve for the equilibrium attack intensity at $t = 2$:

$$a_{Ransom}(Q_i, s_i) = \begin{cases} \sqrt{s_i \rho Q_i} - s_i & \text{if } s_i \leq \rho Q_i \equiv \widetilde{s}_i \\ 0 & \text{if } s_i > \widetilde{s}_i, \end{cases} \tag{19}$$

where $\widetilde{s}_i$ is similar to $\overline{s}_i$ and captures the value of a successful ransomware attack. We then simplify the probability of a successful attack by inputting $a_{Ransom}(Q_i, s_i)$,

$$\delta_{Ransom}(Q_i, s_i) = \delta(a_{Ransom}(Q_i, s_i), s_i) = \begin{cases} 1 - \sqrt{\frac{s_i}{\rho Q_i}} & \text{if } s_i \leq \widetilde{s}_i \\ 0 & \text{if } s_i > \widetilde{s}_i. \end{cases} \tag{20}$$

At $t = 0$, the optimal security investment and fee of a platform are

$$s^*_{Ransom} = \begin{cases} \frac{\rho M}{N} & \text{if } 2c \leq 1 \\ \frac{\rho M}{4c^2 N} & \text{if } 2c > 1, \end{cases} \tag{21}$$

$$f^*_{Ransom} = \begin{cases} \rho c & \text{if } 2c \leq 1 \\ \rho(1 - \frac{1}{4c}) & \text{if } 2c > 1. \end{cases} \tag{22}$$

The resulting attack intensity and probability of a successful attack are

$$a^*_{Ransom} = \begin{cases} 0 & \text{if } 2c \leq 1 \\ \frac{\rho M(2c-1)}{4c^2 N} & \text{if } 2c > 1. \end{cases} \tag{23}$$

$$\delta^*_{Ransom} = \begin{cases} 0 & \text{if } 2c \leq 1 \\ 1 - \frac{1}{2c} & \text{if } 2c > 1. \end{cases} \tag{24}$$

Finally, at $t = -1$, the attacker chooses its attack mode. The attacker compares the expected profit to a conventional attack,

$$\pi_{A,Conv} = \begin{cases} 0 & \text{if } 2rc \leq 1 \\ \frac{M(2rc-1)^2}{4Nrc^2} & \text{if } 2rc > 1. \end{cases} \tag{25}$$

and the expected profit of a ransom attack,

$$\pi_{A,Ransom} = \begin{cases} 0 & \text{if } 2c \leq 1 \\ \frac{\rho M(2c-1)^2}{4Nc^2} & \text{if } 2c > 1. \end{cases} \tag{26}$$

Proposition 3 (Ransomware Equilibrium.) *(i) Platforms are more often successfully attacked when facing a ransomware attack than when facing a conventional attack,* $\delta_C^* \leq \delta_R^*$. *(ii) Consider the case in which both modes of attack are profitable,* $c > \frac{1}{2r}$. *Platforms invest less in security when facing a ransom attack than when facing a conventional attack.*

Ransomware changes the relationship between clients and platforms. When facing conventional attacks, platforms compete not just over fees, but also over the successful probability of an attack. When facing ransoms, however, clients are indifferent to the probability of a successful attack, as long as they expect that platforms pays ransoms upon being breached. Therefore, when platforms select their security investment against ransomware attacks, they do not consider the competitive effect of increased security, only fees. The result is that, for any values of r and ρ, platforms see more frequent breaches against ransom attacks, compared to conventional attacks. When attackers could profitably attack using either of the attack modes ($c > \frac{1}{2r}$), platforms invest less in security when facing ransomware attacks than when facing conventional attacks (Figure 5b).[20]

This result informs real-world outcomes. Ransomware attacks are very common, and many organizations willingly pay ransoms to regain access to their systems. The model implies that this result is endogenous; organizations invest less in security and expect more successful attacks when they can be resolved by ransoms, than when the attacks steal directly from their clients.

Proposition 4 (Attack mode.) *For* $r \leq \frac{1}{2c}$, *the attacker only conducts ransom attacks, regardless of* ρ. *Otherwise, the attacker conducts ransom attacks as opposed to conventional attacks when* $\rho \geq \frac{(2rc-1)^2}{r(2c-1)^2} \equiv \underline{\rho}$. *The bound* $\underline{\rho}$ *increases in* c.

[20] Alternatively, when platforms choose to fully defend against conventional attacks, the difference in security spending is ambiguous. For some parametrizations, platforms may spend more to fully deter conventional attacks and spend less to only partially deter ransom attacks. Similarly, when platforms fully deter both types of attacks ($c \leq \frac{1}{2}$), security spending depends on the magnitudes of ρ and r. A platform may need to spend more to fully deter conventional attacks than to fully deter ransom attacks ($r > \rho$), or vice versa ($\rho > r$).

From Section 2, the attacker does not engage in conventional attacks when $r \leq \frac{1}{2c}$; the same is not true for ransomware attacks, where the boundary is $1 \leq \frac{1}{2c}$ (see Figure 5a). This simple change in boundary conditions has sharp real-world implications. In conventional attacks, the attacker only targets institutions with high values of r, where they can feasibly steal a large proportion of the transaction. Examples of high r platforms can be limited, as regulators are able to prevent transactions or limit the use of the proceeds of crime in many markets. Thus, conventional attacks where the attacker actively steals transactions are likely to be limited to certain high-value segments of the market.

With ransomware, the relative ability of the attacker to extract a ransom ρ is irrelevant to whether the attack takes place. The only determinant of whether an attack can take place with ransomware is the cost of security, c. Unlike r, it is likely that c is relatively similar across industries and markets, due to common access to security providers and IT talent. Thus, whereas conventional attacks should be concentrated among high r targets, ransomware attacks should affect a broad array of industry participants.

When both ransomware and conventional attacks are potentially profitable, the choice of attack type depends on the relative value of ρ and r. Ransomware attacks should be observed for a broad group of institutions with lower values of r, or who are able to pay large ransoms and have a high value of ρ. Alternatively, conventional attacks should be observed among specific high-value targets with high values of r.

Figure 3b shows the regions under which the attacker optimally select ransomware or conventional attacks, based on the cost of security c. For any values of r and c, there is a minimum value of ρ, denoted $\underline{\rho}$, such that if $\rho \geq \underline{\rho}$, the attacker prefers ransomware. This minimum value $\underline{\rho}$ is increasing in the cost of security c, as long as the attacker possibly willing to engage in conventional attacks ($2rc \geq 1$). That is to say, that if security becomes more expensive and c increases, a broader array of institutions should see conventional attacks. Alternatively, if security becomes cheaper and c decreases, conventional attacks should become more isolated compared to ransom attacks. Figure 3b summarizes the above

discussion on the impact of (r, c, ρ) on the choice to i) attack or not attack, and if attacking, ii) whether the attack is conventional or ransomware.

In an environment where attackers use ransomware attacks, we derive the following comparative statics from equations (21)-(24) and summarize them in Proposition 5 below.

Proposition 5 (Comparative Statics when Ransomware Attacks are possible.) *The impact of marginal increases in parameters* $(\frac{M}{N}, \rho, c)$ *on the equilibrium outcomes* $(\delta^*, a^*, s^*, f^*)$ *are given in the table below, where arrows indicate increasing or decreasing.*

	M/N		ρ		c		
	$2c \leq 1$	$2c > 1$	$2c \leq 1$	$2c > 1$	$2c \leq 1$	$c \in (0.5,1)$	$c \geq 1$
Vulnerability (δ^*)	0		0		0	↑	
Attack intensity (a^*)	0	↑	0	↑	0	↑	↓
Security Investment (s^*)	↑		↑		0	↓	
Fees (f^*)	0		↑		↑		

Proposition 5 shows that the impact of market tightness $(\frac{M}{N})$ on a ransomware equilibrium is similar to a conventional attack equilibrium: investments in security and attack intensity (when positive) scale in market tightness, and this offsetting investment leads fees and platform vulnerability to remain unaffected. Moreover, the cost of security influences the ransomware equilibrium similarly to the conventional attack equilibrium, with one exception: fees in the ransomware equilibrium decrease in c when vulnerability is positive $(2c > 1)$ instead of remaining invariant.

When comparing the impact of the benefit to the attacker from the success of a conventional attack (r) versus a ransomware attack (ρ), we look at their impact on the platform: with a ransomware attack, the platform pays the ransom ρ instead of losing the fee, f^*. Thus, to avoid paying the ransom, security investment increases in ρ; because the attacker also benefits more from a higher ρ, the attack intensity is also increasing in ρ. In equilibrium, we find that the increases in security and attack intensity directly offset, leading vulnerability to remain invariant in ρ.

3.1 Welfare and Regulation

Similar to conventional attacks, the regulator's second-best solution when dealing with ransomware attacks is identical to the competitive allocation. (As before, the difference is that fees within the social planner's solution are an arbitrary transfer.)

A regulator may then wish to influence welfare by modifying the model's parameters in the long run. When facing a ransomware attack, welfare is given by

$$W^{SB}_{Ransom} = \begin{cases} M(1-\rho c) & \text{if } 2c \leq 1, \\ \frac{M(4(1-\rho)c+\rho)}{4c} & \text{if } 2c > 1. \end{cases} \tag{27}$$

As before, competition is irrelevant to welfare, and welfare is decreasing in the cost of security. Welfare is also decreasing in ρ, the value which the attacker can demand in ransom from the platform. A regulator wishing to decrease ρ could undertake a number of interventions that limit the size of ransoms or limit the ability to pay ransoms. Regulators could limit the use of cyber insurance for ransoms, as insurance may encourage platforms to pay out using the insurer's funds rather than their own. Indeed, some insurers are now limiting the use of their insurance products for ransoms.[21]

Regulators could also limit the usefulness of ransom payments, by strengthening international enforcement or creating curbs on the use of anonymized cryptocurrencies. An example of one such action is a September 2021 bulletin from U.S. Department of the Treasury's Office of Foreign Assets Control (OFAC) advising on the implications of paying ransoms to entities who may be sanctioned by the U.S. government.[22] Another is the use of offensive actions by law enforcement and intelligence agencies against those perpetuating ransom attacks.[23]

[21] See https://www.insurancejournal.com/news/international/2021/05/09/613255.htm for an example.

[22] See https://home.treasury.gov/system/files/126/ofac_ransomware_advisory.pdf

[23] See https://www.cbc.ca/news/politics/ransomware-critical-infrastructure-cse-1.6274982 for reporting on on such action.

Another possibility would be to reduce losses from unpaid ransoms (L). If it becomes unprofitable for compromised platforms to pay ransoms ($L < \rho$), it can become non-optimal for attackers to engage in these attacks in the first place. For an example, a platform with a redundant backup of critical data, or who is able to confine the attack to a smaller segment of its systems, may suffer fewer losses if it chooses to not pay a ransom.

A decrease in the value of ρ when facing ransom attacks has different implications than a decrease in r when facing conventional attacks. While welfare increases as ρ declines, the amount invested in security declines but the probability of successful attacks remains constant. That is, even if platforms are only able to pay lower ransoms, there should still be an equal volume of successful attacks. This creates an issue if a government regulator wishes to demonstrate successful intervention against ransomware. The regulator must be cognizant that, for ransomware attacks, they may not see a decrease in successful breaches, even if they reduce ransoms paid and increase welfare. The only way a regulator can reduce the observed frequency of successful ransomware attacks is to decrease the security cost c.

4 Unobservable Security Investment

In this section, we examine a model with security investment which is unobservable by the platforms' clients. In reality, clients often have limited access to reliable information about platforms' cyber security practice. For example, publicly traded companies may report total spending on IT infrastructure or cyber security in their financial reports, but often do not provide more granular data. The model is similar to that of Section 1; however, clients are unable to observe security investment by platforms. Attackers continue to observe the security investments of platforms because of their sophistication.

4.1 Conventional attacks

There are three dates, $t = 0, 1, 2$. At date 0, platforms set a fee for transactions and invest in cyber security. At date 1, clients allocate their transaction needs across platforms. They are able to observe the transaction fee but not the security investment. At date 2, the attacker chooses the intensity with which to attack each platform.

Definition 2 (Equilibrium with unobservable security investment.) *An equilibrium of this game is given by* $a^*_{I,i}$, $s^*_{I,i}$, $f^*_{I,i}$, $\hat{s}(f^*_{I,i})_{I,i}$ *and* $q^*_{I,im}$ *for all* $i = 1, ..., N$ *and* $m \in [0, M]$.

1. *At* $t = 2$, *the attacking strategy on platform* i *is* $a(Q_i, s_i) = \arg\max_{a_i} \pi_A$, *for any* Q_i *and* s_i.

2. *At* $t = 1$, *the clients' beliefs about security investment are* $\hat{s}_{I,i}(\mathbf{f}) = \arg\max_{s_i} \pi_i(\mathbf{f})$ *for any equilibrium in fees* $\boldsymbol{f} \equiv \{f_i\}_{i=1}^N$ *and the attacking strategies* $a(Q_i, s_i)$.

3. *At* $t = 1$, *the transaction allocation strategy is* $q_{im}(\hat{s}_{I,i}(\mathbf{f}), \mathbf{f}) = \arg\max_{q_{im}} U_m$ *subject to* $\sum_{i=1}^N q_{I,im} = Q_m$, *their beliefs regarding security strategies* $\hat{s}_{I,i}(\mathbf{f})$ *and the attacking strategies* $a(Q_i, s_i)$, *for any equilibrium in fees* $\boldsymbol{f}$.

4. *At* $t = 0$, $(\mathbf{s}^*, \mathbf{f}^*)$ *is a Nash equilibrium among the platforms. That is,* $(s^*_i, f^*_i) = \arg\max_{s_i, f_i} \pi_i$, *given the choices of the other platforms* (s_{-i}, f_{-i}), *the allocation strategies and beliefs of clients* $q_{im}(\hat{s}_{I,i}(\mathbf{f}), \mathbf{f})$ *and* $\hat{s}_{I,i}(\mathbf{f})$, *and the attacking strategies* $a(Q_i, s_i)$.

At date 2, the attacker's problem and its solutions are identical to those presented in Sections 1 and 2. The attacker's best response function $a(Q_i, s_i)$ is given by Equation 5, and the probability of a successful attack $\delta(Q_i, s_i)$ is given by equation 6.

At date 1, clients allocate their transactions over all platforms to maximize their utility. They do so depending on their beliefs $\hat{s}_i(f_i)$ regarding security investment at each platform, given fees. Consistent with sequential rationality, clients believe that each platform chose security $\hat{s}_i(f_i)$ such that this value of security would maximize the platform's profit, given

observable fees f_i.[24] Given $\delta(Q_i, \hat{s}_i(f_i))$ and f_i, client utility at each platform is expressed as

$$U_m(f_i) = \begin{cases} (1 - f_i)q_{im} & \text{if } \hat{s}_i(f_i) \geq \overline{s}_i \\ (1 - \delta(Q_i, \hat{s}_i(f_i)))(1 - f_i)q_{im} & \text{if } \hat{s}_i(f_i) < \overline{s}_i. \end{cases} \tag{28}$$

Clients again split their transactions equally over all platforms that offer the highest utility.

At date 0, platforms select fees to maximize their expected profits, given the best response functions of all other agents. A platform's expected profits are given by Equation 1. There exists an equilibrium such that fees are given

$$f_I^* = \begin{cases} 2rc & \text{if } 4rc \leq 1 \\ \frac{1}{2} & \text{if } 4rc > 1. \end{cases} \tag{29}$$

The equilibrium fee depends on security costs, c, and the percentage of the transaction that can be stolen, r. For $4rc > 1$, there exists an equilibrium such that each platform i believes each other platform $-i$ will charge $f_{-i} = \frac{1}{2}$. Platform i then selects $f_i = \frac{1}{2}$ to maximize its own profit. Alternatively, for $4rc \leq 1$ there exists an equilibrium such that each platform i believes each other platform $-i$ will charge $f_{-i} = 2rc$, and in turn charges $f_i = 2rc$.

All M clients then allocate their transactions equally over all N platforms. The resulting security investment is:

$$s_I^* = \begin{cases} \frac{rM}{N} & \text{if } 4rc \leq 1 \\ \frac{M}{16rc^2N} & \text{if } 4rc > 1, \end{cases} \tag{30}$$

which results in a probability of a successful breach of

$$\delta_I^* = \begin{cases} 0 & \text{if } 4rc \leq 1 \\ 1 - \frac{1}{4rc} & \text{if } 4rc > 1, \end{cases} \tag{31}$$

[24] We assume that clients hold these beliefs, regardless of whether they observe an on-equilibrium or off-equilibrium f_i. While many beliefs about off-equilibrium f_i are possible, as Perfect Bayesian Equilibrium does not impose a structure on off-equilibrium beliefs, this is one set that is economically meaningful.

Proposition 6 (Security under unobservable security investment.) *When $4rc > 1$, platforms invest less in security ($s_I^* < s^*$) and experience more breaches ($\delta_I^* > \delta^*$) when clients do not observe the platform investment in security.*

4.2 Minimum Security Standards

Welfare in the case with unobservable security investment is

$$W_I^* = \begin{cases} M(1-rc) & \text{if } 4rc \leq 1 \\ \frac{3M}{16rc} & \text{if } 4rc > 1. \end{cases} \tag{32}$$

The second-best level of welfare is unchanged and continues to be given by Equation (14). Welfare in the competitive equilibrium with unobservable security investment is equal to this value when $4rc \leq 1$. However, when $4rc > 1$ welfare is lower in the unobservable security investment case.

Since competition is unable to achieve the second-best under unobservable security investment, a regulator may be interested in taking actions to improve welfare. One common suggestion is that a regulator establish minimum cyber security standards for financial institutions and other important service providers such as utilities. In practice, though many security standards are voluntary "best practices", government regulators have been able to set standards in some areas. As one example, in November 2021, the U.S. OCC, Federal Reserve Board and FDIC established minimum reporting standards for cyber security incidents at federally-regulated banking institutions.[25]

In our model, a regulator is assumed to be able to observe the security level of platforms and can set minimum security standards. Consider a regulator who seeks to maximize welfare, but is only able to set a minimum security value $\underline{s}$ such that $s_i \geq \underline{s}$ is a constraint of the platforms' choices of security investment at $t = 0$.

[25] See, https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20211118a1.pdf

A regulator is able to improve welfare by setting a value:

$$\underline{s}^* = \begin{cases} [0, \frac{rM}{N}] & \text{if } 4rc \leq 1 \\ \frac{M}{4c^2 rN} & \text{if } 4rc > 1. \end{cases} \tag{33}$$

When $4rc \leq 1$, the regulator is unable to improve welfare above the competitive value. In this case, it does not harm welfare as long as it sets a security minimum no larger than than the competitive value. When $4rc > 1$, by contrast, the regulator sets a security minimum equal to the competitive equilibrium value with perfect information. In this case, each platform selects security equal to the minimum security and the second best is achieved.[26]

4.3 Ransomware

We extend the unobservable security investment setting to an attacker with access to ransomware. Similar to Section 3, attackers choose to attack with a conventional theft, or a ransom attack in $t = -1$. As above, clients are unable to observe the platforms' security investment.

Unlike thefts, unobservable security investment does not impact the security investment choices of platforms facing ransom attacks. Similar to Section 3, platforms with $L \geq \rho$ will optimally pay ransoms, and their clients only consider the fees charged by the platforms. In this case, equilibrium values of s^I_{Ransom}, f^I_{Ransom}, a^I_{Ransom} and δ^I_{Ransom} are identical to those given by Equations (21) through (24). Alternatively, platforms with $L < \rho$ will optimally select $s^I_{Ransom} = 0$ and $f^I_{Ransom} = 0$, and face no probability of a successful ransom attack.

Since unobservable security investment does not change the attacker's expected profits with ransomware, they are identical to those in Equation (26). From Section 4.1, the attacker's expected profit from a conventional attack under unobservable security investment

[26] Alternatively, a regulator can achieve the second best by setting a minimum fee. Such a regulator sets a minimum fee of $\underline{f^*} = 1$ when $4rc > 1$. While constrained efficient, such a policy may be undesirable to a regulator as it transfers the entire client surplus to the platforms.

is

$$\pi^{I}_{A,Conv} = \begin{cases} 0 & \text{if } 4rc \leq 1 \\ \frac{M(4rc-1)^2}{16Nrc^2} & \text{if } 4rc > 1. \end{cases} \tag{34}$$

The attacker's choice of attack mode, shown in Figure 3c, again depends on the relative value of the underlying parameters. For $\frac{1}{2} < c \leq \frac{1}{4r}$, conventional attacks are unprofitable and attackers will only select ransoms. This region is similar in nature to the one under perfect information, but occurs under a smaller parameter space. For $\frac{1}{4r} < c \leq \frac{1}{2}$, ransom attacks are unprofitable and attackers will only select conventional attacks. This region does not occur under perfect information and represents an expansion of the parameter space under which attackers will conduct thefts. For $c \geq max\{\frac{1}{4r}, \frac{1}{2}\}$, both attack modes are potentially profitable. Attackers will select ransoms if and only if

$$\rho \geq \frac{(4rc-1)^2}{4r(2c-1)^2} \equiv \underline{\rho}_I \tag{35}$$

Comparing the value of $\underline{\rho}_I$ to $\underline{\rho}$ under perfect information, we have $\underline{\rho}_I \geq \underline{\rho}$. That is, when both conventional attacks and ransom attacks are profitable, ρ must be higher under unobservable security investment for attackers to select ransoms. Put another way, there is a wider parameter space under which attackers select conventional attackers under unobservable security investment. Figure 3d illustrates this difference by a darker region that indicates a parameter space in which conventional attacks would be profitable, where none would be in the case of observable security investment. Taken together, these insights generate the following Corollary.

Corollary 1 (Welfare under Unobservable Security Investment) *Under unobservable security investment, welfare is (weakly) lower for all* (r, c, ρ) *when compared to the case where security investment is observable.*

Corollary 1 highlights the importance of security investment information towards improving overall welfare. While regions of the (r, c, ρ) parameter space may be identical across

both information environments (i.e., when attack mode does not change across environments), there is a region illustrated by light and dark grey shading in Figure 3c in which the ability of clients to observe of security investment by platforms can improve welfare. In this region, publicizing security investments moves attackers from conventional to ransomware attacks, the response to which by platforms achieves constrained efficiency.

The ability of regulators to increase transparency of platforms' security investments may not be far off, with firms like BitSight and UpGuard offering a numerical score–similar in style to a credit ratings offered from the likes of Equifax–that describes the quality of a firm's cyber security. Moreover, BitSight publishes its industry average scores, which provides clients with the ability to compare platform scores to the industry average, should the platform acquire and publish their score. Thus, mandating security scores may be one such tool for moving the market closer towards fully-observable security investment. Currently, the SEC's consideration to mandate disclosure of cybersecurity practices and incident reports would reflect the spirit of our model's prediction. The extent to which this policy may achieve the intended goals, however, will depend on the ability of the target clientele—investors, in this case—to infer security investment from this type of information.

5 Extensions

In this section we discuss three extensions to the model: (1) a monopolistic financial platform; (2) an attacker which is limited by its own budget constraint; and (3) platforms which face increasing marginal costs of investing in security.

5.1 Monopoly

In this section, we investigate the impact of platform competition on platform vulnerability and fees by considering the setting of monopoly, $N = 1$. Similar to the competitive case, the

monopolist platform chooses security s and fees f to maximize expected profit

$$\pi_M = (1-\delta) f\, Q - cs, \tag{36}$$

where $Q \leq M$ is the transaction volume attracted by the monopolist.

The client utility function remains unchanged but clients can only accept or reject the fee and security investment offered by the monopolistic platform. A client accepts if and only if $(1-\delta_M)(1-f_M) \geq 0$. Hence, a client chooses to accept any fee $f_M \leq 1$ and attack probability $\delta_M \leq 1$. Upon accepting (s_M, f_M), the monopolistic platform receives the total mass of client transactions, $Q = M$.

The attacker's strategy is identical to the competitive case, and is accounted for by the monopolist in its security and fee decisions (see also Appendix A.6 for the proof). Thus, the monopoly level of security is

$$s_M = \begin{cases} rM & \text{if } 2rc \leq 1, \\ \frac{M}{4rc^2} & \text{if } 2rc > 1 \end{cases} \tag{37}$$

In comparison to the competitive equilibrium, s_M equals the sum of s_i^* (Equation (8)) across all N venues for all parameter values (r, c). Hence, vulnerability under monopoly is

$$\delta_M = \begin{cases} 0 & \text{if } 2rc \leq 1, \\ 1 - \frac{1}{2rc} & \text{if } 2rc > 1 \end{cases} \tag{38}$$

Equation (38) yields that $\delta_i^* = \delta_M$, where δ_i^* is the vulnerability at each platform i in the competitive case. Hence, the security of transactions do not depend on the level of competition, nor does "putting all one's eggs in one basket" impact a client's transaction vulnerability in our model. Similar to the competitive case, a monopolist invests to fully secure the platform when product of marginal cost of security c and the value of the transaction to the attacker r is sufficiently low; for higher joint values, platform vulnerability increases in

r and c. Unlike the competitive case, the monopolist has the incentive to extract maximum rent from its clients, $f_M = 1$. We summarize the results in the following proposition.

Proposition 7 (Monopoly.) *Monopoly* ($N = 1$) *features the same attacking intensity,* $a_M = a^*$*, security investment,* $s_M = s^*$*, and success probability,* $\delta_M = \delta^*$*, as the competitive equilibrium but fees are higher,* $f_M > f^*$*. Thus, the allocation is constrained efficient.*

In monopoly clients intuitively face higher fees, as the monopolist extracts rents when compared to the competitive case. Moreover, clients suffer pure rent extraction regardless of the cost of security c or the value of the transaction to the attacker r, as the number of realized security breaches is identical to the competitive case. The results of Propositions 1 and 7 imply that competition among platforms plays a key role in improving the expected utility of clients by minimizing platform rent extraction without impacting vulnerability.

Proposition 7 also highlights that creating a "fortress-like platform" where a monopoly invests to guard all client transactions does not benefit clients: the monopoly charges higher fees while offering the same level of protection to each transaction. Since the monopolistic allocation is constrained efficient, there is no role for regulation, however. In fact, an authority who regulates maximum levels of fees or minimum levels of security investment either is ineffective (when these bounds do not bind) or reduces welfare in the economy.

5.2 Attacker Budget Constraint

In this section we consider an attacker who is subject to a budget constraint for attacking. This attacker represents an adversary who must strategically divide limited attention or limited resources across potential targets. The model is identical to the one presented in Section 1, with the exception that the attacker is subject to the constraint:

$$\sum_{i=1}^{N} a_i \leq \bar{a}. \tag{39}$$

In symmetric equilibrium, there is an equal quantity at each platform, $Q_i = \frac{M}{N}$. For a given security investment, the attacker's attacking strategy for platform i now depends on other platforms as well and is given by

$$a_i(s_i) = \begin{cases} \sqrt{\frac{s_i r M}{N}} - s_i & \text{if } \sum_{i=1}^{N} \sqrt{\frac{s_i r M}{N}} - s_i \leq \bar{a} \text{ and } s_i \leq \bar{s_i} \\ \frac{\sqrt{s_i}\bar{a} + \sqrt{s_i}\sum_{j \neq i} s_j - s_i \sum_{j \neq i} \sqrt{s_j}}{\sqrt{s_i} + \sum_{j \neq i} \sqrt{s_j}} & \text{if } \sum_{i=1}^{N} \sqrt{\frac{s_i r M}{N}} - s_i > \bar{a} \text{ and } s_i \leq \bar{s_i} \\ 0 & \text{if } s_i > \bar{s_i}, \end{cases} \tag{40}$$

which has three segments. The first and third are identical to the equilibrium in Section 2 and represent cases where the attacker is unconstrained or where the attacker does not attack. The second case is new and represents the case of scarce aggregate attention or resources.

Given the attacker's strategies, the platforms' profit maximization is the same as in Section 2. In a symmetric equilibrium, $s_i^* = s^*$, the optimal security investment is

$$s^* = \begin{cases} \frac{M}{4c^2 r N} & \text{if } 1 \leq 2rc \text{ and } M < \frac{4\bar{a}rc^2}{2rc-1} \\ \frac{\sqrt{\frac{M}{N}(\frac{M}{N}(N-1)^2 - 16\bar{a}c)} + \frac{M}{N}(N-1) - 4\bar{a}c}{4cN} & \text{if } 1 \leq 2rc \text{ and } M \geq \frac{4\bar{a}rc^2}{2rc-1} \\ \frac{rM}{N} & \text{if } 1 > 2rc \end{cases} \tag{41}$$

and the optimal fee is

$$f^* = \begin{cases} \frac{1}{2} & \text{if } 1 \leq 2rc \text{ and } M < \frac{4\bar{a}rc^2}{2rc-1} \\ \frac{\frac{M}{N}(N-1) + \sqrt{\frac{M}{N}(\frac{M}{N}(N-1)^2 - 16\bar{a}c)}}{4M} & \text{if } 1 \leq 2rc \text{ and } M \geq \frac{4\bar{a}rc^2}{2rc-1} \\ rc & \text{if } 1 > 2rc. \end{cases} \tag{42}$$

The resulting response from the attacker is

$$a^* = \begin{cases} \frac{M}{2cN}\left(1 - \frac{1}{2rc}\right) & \text{if } 1 \leq 2rc \text{ and } M < \frac{4\bar{a}rc^2}{2rc-1} \\ \frac{\bar{a}}{N} & \text{if } 1 \leq 2rc \text{ and } M \geq \frac{4\bar{a}rc^2}{2rc-1} \\ 0 & \text{if } 1 > 2rc \end{cases} \tag{43}$$

and the resulting probability of a successful attack is

$$\delta^* = \begin{cases} 1 - \frac{1}{2rc} & \text{if } 1 \leq 2rc \text{ and } M < \frac{4\bar{a}rc^2}{2rc-1} \\ \frac{4\bar{a}c}{\frac{M}{N}(N-1)+\sqrt{\frac{M}{N}\left(\frac{M}{N}(N-1)^2-16\bar{a}c\right)}} & \text{if } 1 \leq 2rc \text{ and } M \geq \frac{4\bar{a}rc^2}{2rc-1} \\ 0 & \text{if } 1 > 2rc \end{cases} \tag{44}$$

An interesting implication is that the level of competition, via N in reduced form, and market tightness, $\frac{M}{N}$, can matter for the allocation when the aggregate attention $\bar{a}$ is limited.

5.3 Increasing Marginal Costs of Security

In this section we consider an increasing marginal cost of security, so the platforms' new profit function is

$$\pi_i = (1 - \delta_i) f_i Q_i - \frac{c s_i^2}{2}. \tag{45}$$

(Everything else remains as in the model in Section 1.) The platforms are subject to the same constraint as in Section 2, so the attacker's strategy and the probability of successful attack continue to be given by Equations (5) and (6). The platforms' optimal security investment is

$$s^* = \begin{cases} \left(\frac{M}{4c^2rN}\right)^{\frac{1}{3}} & \text{if } 1 \leq \frac{2r^2cM}{N} \\ \frac{rM}{N} & \text{if } 1 > \frac{2r^2cM}{N}. \end{cases} \tag{46}$$

The optimal fee is

$$f^* = \begin{cases} \frac{1}{4} & \text{if } 1 \leq \frac{2r^2cM}{N} \\ \frac{cr^2M}{2N} & \text{if } 1 > \frac{2r^2cM}{N}. \end{cases} \tag{47}$$

The optimal security investment is such that the optimal attack intensity is

$$a^* = \begin{cases} \left(\frac{rM^2}{2cN^2}\right)^{\frac{1}{3}} - \left(\frac{M}{4rcN}\right)^{\frac{1}{3}} & \text{if } 1 \leq \frac{2r^2cM}{N} \\ 0 & \text{if } 1 > \frac{2r^2cM}{N}, \end{cases} \tag{48}$$

and the probability of a successful attack is

$$\delta^* = \begin{cases} 1 - \left(\frac{N}{2r^2cM}\right)^{\frac{1}{3}} & \text{if } 1 \leq \frac{2r^2cM}{N} \\ 0 & \text{if } 1 > \frac{2r^2cM}{N}. \end{cases} \tag{49}$$

Again, the degree of competition (and market tightness) affect the equilibrium allocations in this extension, which contrasts with the main model.

6 Conclusion

With many in the financial sector still working from home, millennials trading using only a smart phone, and seniors doing their banking in their dining room, the security of financial platforms is critical. Large scale thefts of cryptocurrencies and platforms paying millions in ransoms to cyber attackers highlights the problems. Thus far, the real risk and destabilizing potential of the cyber security of the financial system has received insufficient attention.

To fill this gap, we construct a model of the financial system in which clients delegate security decisions to platforms. The model allows us to understand the theft of assets, as well as the more recent case of disruptions caused by ransomware attacks. When cyber attackers seek to steal assets, the economic problem is a principal-agent-type problem between clients and platforms: clients delegate security decision to an agent, and bear the losses if hacked.

Ransomware changes the relationship between client and platform. When platforms bear losses because their services are being disrupted until ransom is paid, they paradoxically under invest in security. This leads to more security breaches because clients focus *only* on fees and not on potential losses due to theft. Moreover, if security investment by platforms is unobservable, this worsens welfare, and widens the scope of firms susceptible to attacks.

Our paper has a number of implications. First, regulators cannot improve welfare by targeting security investment when security investment is observable, as the equilibrium outcome is constrained efficient. Instead, regulators can affect welfare by reducing the cost of security and by reducing the value of stolen data to attackers. For example, regulators can create curbs on the use of cryptocurrencies or forbid the payment of ransoms to sanctioned entities. If security investment is unobservable, a regulator may improve welfare by setting minimum security standards, or targeting information opacity by requiring firms to obtain and publish third-party security ratings (i.e., from cyber security rating firms such as BitSight or UpGuard). We highlight that cyber insurance policies that insure against ransomware attacks may cause unintended harm. While intuitively appealing, our results suggest that making it more likely that high-value ransoms will be paid may reduce overall welfare and security. The payment of ransoms make all platforms more susceptible to attacks and increase the probability of an attack.

Our testable implications provide several directions for further study. For instance, does an increase in publicly available information on platform security change the nature of, or reduce the instances of attacks in certain industries? Does banning insurers from reimbursing ransom payments lead to noticeably fewer (ransomware) attacks? Future work can also consider the risk of contagion: is the security of each individual institution sufficient to protect the financial system? Conversely, does a market with more platforms provide benefits to diversification with respect to cyber attackers? With the importance of cyber security to the functioning of the financial system, and the lack of work in the area, more academic research is warranted.

References

Acemoglu, Daron, Azarakhsh Malekian, and Asu Ozdaglar, 2016, Network security and contagion, *Journal of Economic Theory* 166, 536–585.

Anderson, Ross, 2001, Why information security is hard-an economic perspective, in *Proceedings 17th Annual Computer Security Applications Conference (ACSAC), 2001.*, 358–365, IEEE.

Anderson, Ross, Chris Barton, Rainer Böhme, Richard Clayton, Michel JG Van Eeten, Michael Levi, Tyler Moore, and Stefan Savage, 2013, Measuring the cost of cybercrime, in *The economics of information security and privacy.*, 265–300.

Anderson, Ross, and Tyler Moore, 2006, The economics of information security, *Science* 314, 610–613.

August, T., and T.I Tunca, 2006, Network software security and user incentives, *Management Science* 52, 1703–1720.

August, Terrence, Duy Dao, and Marius Florin Niculescu, 2022, Economics of ransomware: Risk interdependence and large-scale attacks, *Management Science* (forthcoming).

Becker, Gary S., 1968, Crime and punishment: An economic approach, *Journal of Political Economy* 76, 169–217.

Biancotti, Claudia, 2017, The price of cyber (in) security: evidence from the italian private sector, *Questioni di Economia e Finanza (Occassional Papers) 407* .

Bier, Vicki, Santiago Oliveros, and Larry Samuelson, 2007, Choosing what to protect: Strategic defensive allocation against an unknown attacker, *Journal of Public Economic Theory* 9, 563–587.

Böhme, R., and G. Schwartz, 2010, Modeling cyber-insurance: Towards a unifying framework., *Workshop on the Economics of Information Security (WEIS)*. Harvard University.

Crosignani, Matteo, Marco Macchiavelli, and André F. Silva, 2022, Pirates without borders: The propagation of cyberattacks through firms' supply chains, Technical report.

Duffie, Darrell, and Joshua Younger, 2019, *Cyber runs* (Brookings).

Dynes, Scott, Eric Goetz, and Michael Freeman, 2007, Cyber security: Are economic incentives adequate?, in *International Conference on Critical Infrastructure Protection*, 15–27.

Dziubiński, Marcin, and Sanjeev Goyal, 2013, Network design and defence, *Games and Economic Behavior* 79, 30–43.

Eisenbach, Thomas M., Anna Kovner, and Michael Junho Lee, 2021, Cyber risk and the u.s. financial system: A pre-mortem analysis, *Journal of Financial Economics* (forthcoming).

Florackis, Chris, Christodoulos Louca, Roni Michaely, and Michael Weber, 2022, Cybersecurity risk, Technical report.

Foley, Sean, Jonathan R Karlsen, and Tālis J Putniņš, 2019, Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?, *The Review of Financial Studies* 32, 1798–1853.

Gordon, Lawrence A, and Martin P Loeb, 2002, The economics of information security investment, *ACM Transactions on Information and System Security (TISSEC)* 5, 438–457.

Goyal, Sanjeev, and Adrien Vigier, 2014, Attack, Defence, and Contagion in Networks, *The Review of Economic Studies* 81, 1518–42.

Grossklags, J., N. Christin, and J. Chung, 2008, Secure or insure? a game-theoretic analysis of information security games., *Proceedings of the 17th International Conference on World Wide Web,* Association for Computing Machinery, New York, NY., 209–218.

Hoyer, Britta, and Kris de Jaegher, 2016, Strategic network disruption and defense, *Journal of Public Economic Theory* 18, 802–830.

Johnson, B., R. Böhme, and J. Grossklags, 2011, Security games with market insurance., *Decision and Game Theory for Security* Berlin, Heidelberg, 117–130.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M. Stulz, 2021, Risk management, firm reputation, and the impact of successful cyberattacks on target firms, *Journal of Financial Economics* 139, 719–749.

Kotidis, Antonis, and Stacey L. Schreft, 2022, Cyberattacks and financial stability: Evidence from a natural experiment, Technical report.

Kovenock, Dan, and Brian Roberson, 2018, The optimal defense of networks of targets, *Economic Inquiry* 56, 2195–2211.

Laszka, Aron, Sadegh Farhang, and Jens Grossklags, 2017, On the economics of ransomware, *Decision and Game Theory for Security* (GameSec 2017), 397–417.

Manshaei, Mohammad Hossein, Quanyan Zhu, Tansu Alpcan, Tamer Bacşar, and Jean-Pierre Hubaux, 2013, Game theory meets network security and privacy, *ACM Computing Surveys* 45, 1–39.

Massacci, Fabio, Joe Swierzbinski, and Julian Williams, 2017, Cyberinsurance and public policy: Self-protection and insurance with endogenous adversaries .

Moore, Tyler, 2010, The economics of cybersecurity: Principles and policy options, *International Journal of Critical Infrastructure Protection* 3, 103–117.

Paquet-Clouston, Masarah, Bernhard Haslhofer, and Benoit Dupont, 2018, Ransomware payments in the bitcoin ecosystem, *arXiv preprint arXiv:1804.04080* .

Parlour, Christine, 1998, Price dynamics in limit order markets, *Review of Financial Studies* 11, 789–816.

A Proofs

A.1 Competitive equilibrium

We solve for an equilibrium through backwards induction. First, at $t = 2$ the attacker chooses a_i to maximize π_A that can be written as

$$\pi_A = r \sum_{i=1}^{N} \left(\frac{a_i}{a_i + s_i} Q_i - a_i \right). \tag{50}$$

The first derivatives with respect to a_i are set equal to zero, resulting in:

$$\frac{\partial \pi_A}{\partial a_i} : \frac{s_i}{(a_i + s_i)^2} r Q_i - 1 = 0. \tag{51}$$

The resulting $a(Q_i, s_i)$ and $\delta(Q_i, s_i)$ are given by Equations (5) and (6).

Second, at $t = 1$ the clients allocate their transactions q_{im} such that $q_{im}(\mathbf{s}, \mathbf{f}) = \arg\max_{q_{im}} U_m$ with $\sum_{i=1}^{N} q_{im} = Q_m$ for any s_i and f_i and given $a(Q_i, s_i)$. The clients' utility is

$$U_m = \sum_{i=1}^{N} (1 - \delta(Q_i, s_i))(1 - f_i) q_{im}. \tag{52}$$

A client's utility is maximized by allocating $q_{im} > 0$ to any group of platforms with the highest $(1 - \delta(Q_i, s_i))(1 - f_i)$. Clients equally allocate q_{im} amongst these platforms.

Third, each platform i assumes that amongst the other platforms $-i$, a different platform j offers the highest value of $(1-\delta(Q_j, s_j))(1-f_j)$, where $(1-\delta_j)(1-f_j) = \zeta$ and $\zeta > 0$. Each platform chooses s_i and f_i to maximize its profits, taking the actions of the other platforms (s_{-i}, f_{-i}) as given, such that $(1 - \delta_i)(1 - f_i) \geq \zeta$. The first-order conditions with respect to s_i and f_i are

$$\frac{\partial \pi_i}{\partial s_i} : \sqrt{\frac{Q_i}{s_i r}} \frac{f_i}{2} - c + \sqrt{\frac{1}{s_i r Q_i}} \frac{\lambda (1 - f_i)}{2} = 0, \tag{53}$$

$$\frac{\partial \pi_i}{\partial f_i} : \sqrt{\frac{s_i Q_i}{r}} - \lambda \sqrt{\frac{s_i}{r Q_i}} = 0, \tag{54}$$

$$\lambda \left(\sqrt{\frac{s_i}{r Q_i}} (1 - f_i) - \zeta \right) = 0, \tag{55}$$

where λ is a Lagrange multiplier. It can be shown that $\lambda = 0$ implies $\zeta < 0$, which violates the clients' participation constraints. Thus, $\lambda > 0$ and $\sqrt{\frac{s_i}{r Q_i}}(1 - f_i) = \zeta$, which results in:

$$s_i = \begin{cases} \frac{Q_i}{4c^2 r} & \text{if } a_i^* > 0 \\ rQ_i & \text{if } a_i^* = 0, \end{cases} \tag{56}$$

$$f_i = 1 - \zeta \sqrt{\frac{rQ_i}{s_i}}. \tag{57}$$

Finally, we impose symmetry among all platforms and invoke a zero profit condition to solve for ζ^*. The equilibrium functions are piece-wise, depending on whether $a^* > 0$ or $a^* = 0$, and are shown in Equations (8) through (11).

A.2 Proposition 2

Recall the functions for $(s^*, f^*, a^*, \delta^*)$ from (8)-(11) respectively. Most of the comparative statics follow by inspection. For s^*, for example, we see that, for $2rc \leq 1$, s^* is independent of c, and increases in M/N and r. For $2rc > 1$, s^* increases in M/N but decreases in c and r. Consider f^* next. For $2rc \leq 1$, we can see that f^* is independent of M/N, and increases in c and r. For $2rc > 1$, f^* is independent of all parameters.

Next, for a^*, for $2rc \leq 1$, we can see that a^* is independent of all parameters. For $2rc > 1$, a^* increases in M/N and in r. To study how a^* changes in c, we take the first derivative:

$$\frac{\partial a^*}{\partial c} = -\frac{M(rc-1)}{2c^3 rN} \tag{58}$$

Hence, $\frac{\partial a^*}{\partial c}$ increases in c for $rc \in (1/2, 1)$, and decreases in it for $rc > 1$.

Finally, for δ^*, we can see that, for $2rc \leq 1$, δ^* is independent of all parameters. For $2rc > 1$, δ^* is independent of M/N, but increases in r and c.

A.3 Proposition 3

The are three regions, $0 < c \leq \frac{1}{2}$, then $\frac{1}{2} < c \leq \frac{1}{2r}$, and then $\frac{1}{2r} < c$. When $0 < c \leq \frac{1}{2}$, $\delta^* = 0$ and $\delta^*_{Ransom} = 0$. When $\frac{1}{2} < c \leq \frac{1}{2r}$, $\delta^* = 0$ and $\delta^*_{Ransom} = 1 - \frac{1}{2c} > 0$. Finally, when $\frac{1}{2r} < c$, $s^* = \frac{M}{4c^2 rN}$ and $s^*_{Ransom} = \frac{\rho M}{4c^2 N}$. Since $0 < r, \rho \leq 1$, then $s^* \geq s^*_{Ransom}$. Further, $\delta^* = 1 - \frac{1}{2rc}$ and $\delta^*_{Ransom} = 1 - \frac{1}{2c}$, since $0 < r \leq 1$, then $\delta^* \leq \delta^*_{Ransom}$.

A.4 Proposition 5

In this proof, we show the comparative statics directions listed in the table of Proposition 5. Recall the equilibrium equations for $(a^*, s^*, f^*, \delta^*)$. First, we examine s^*:

$$s^*_{Ransom} = \begin{cases} \frac{\rho M}{N} & \text{if } 2c \leq 1 \\ \frac{\rho M}{4c^2 N} & \text{if } 2c > 1, \end{cases} \tag{59}$$

By inspection, for all $2c \leq 1$, s^* is invariant in (c, ρ) and increasing in M/N. For $2c > 1$, s^* is invariant in ρ, increasing in M/N, and decreasing in c, as c^2 is in the denominator.

Proceeding with f^* :

$$f^*_{Ransom} = \begin{cases} \rho c & \text{if } 2c \leq 1 \\ \rho(1 - \frac{1}{4c}) & \text{if } 2c > 1. \end{cases} \tag{60}$$

M/N is not present in Equation 60, so f^* is invariant to M/N. Moreover, f^* is increasing in ρ for all c, and i) increasing in c for $2c \leq 1$, and ii) increasing in c for $2c > 1$, as the first-order derivative equals $\rho/4c^2$.

For a^*, we have:

$$a^*_{Ransom} = \begin{cases} 0 & \text{if } 2c \leq 1 \\ \frac{\rho M(2c-1)}{4c^2 N} & \text{if } 2c > 1. \end{cases} \tag{61}$$

a^* is invariant to $(M/N, c, \rho)$ for $2c \leq 1$, and increasing in $(\rho, M/N)$ for $2c > 1$. We then take the piecewise derivative with respect to c for $2c > 1$, and find that it equals $\frac{\rho M(1-c)}{2c^3 N}$. Hence, for all $c \in (0.5, 1)$, f^* is increasing in c, and (weakly) decreasing in c for $c \geq 1$.

Finally, for δ^*, we have:

$$\delta^*_{Ransom} = \begin{cases} 0 & \text{if } 2c \leq 1 \\ 1 - \frac{1}{2c} & \text{if } 2c > 1. \end{cases} \tag{62}$$

δ^* is invariant to $(M/N, c, \rho)$ at $2c \leq 1$ because $a^* = 0$. For $2c > 1$, only c is present, and by inspection, δ^* is increasing in c.

A.5 Unobservable Security Investment

The attacker's problem at $t = 2$ and its solution are identical to those in the competitive case.

The clients' problem in $t = 1$ requires beliefs regarding platforms' security investment. As clients observe fee from each platform, we condition these beliefs on fees and denote them $\hat{s}(f_i)_i$ for each platform i. We assign clients the beliefs that each platform selects

$\hat{s}(f_i)_i = \arg\max_{s_i} \pi_i(f_i)$. Clients then allocate transactions across platforms to maximize their utility function, given fees and beliefs regarding security investment.

In $t = 0$, platforms select s_i and f_i to maximize their profit function, given by Equation 1, such that,

$$(1 - \delta(\hat{s}_i))(1 - f_i) \geq \zeta \tag{63}$$

Note that platforms do not control $\hat{s}_i$ in the constraint directly, unlike how they control s_i. There then exists an equilibrium among platforms given by Equations 29 and 30.

We note that, under the beliefs that each platform selects $\hat{s}(f_i)_i = \arg\max_{s_i} \pi_i(f_i)$, clients who observe an off-equilibrium f_i have no incentive to allocate transaction volume to the deviating platform. Consider clients who observe some off-equilibrium fee, denoted f_i' and where $4rc > 1$. These clients believe that the security value s_i' is such that $s'(f_i')_i = \arg\max_{s_i} \pi_i(f_i')$. The on-equilibrium $f_I^* = \frac{1}{2}$ results in an expected client utility of $U_m(f_I^*) = \frac{1}{8rc}$. For any observed f_i', the inferred $s'(f_i')_i = \frac{Q_i f_i'^2}{4rc^2}$ results in $U_m(f_i') = \frac{f_i'(1-f_i')}{2rc}$, which is equal to $\frac{1}{8rc}$ for $f_i' = f_I^* = \frac{1}{2}$ and lower otherwise. The case when $4rc \leq 1$ is similar.

A.6 Monopoly

The monopoly case is also solved through backwards induction. The attacker's problem at $t = 2$ and its solution are identical to those in the competitive case with $N = 1$. At $t = 1$, the clients' utility functions are identical to the competitive case with $N = 1$. We assume that the client has a reservation utility of 0, and therefore accepts any fee $f \leq 1$ and attack probability $\delta \leq 1$. Finally, the monopolist platform chooses s and f at $t = 0$ such that

$$(s_M, f_M) = \arg\max_{s,f} \pi_M = (1 - \delta) f Q - cs \quad \text{s.t.} \quad 0 \leq (1 - \delta)(1 - f) \tag{64}$$

The first-order conditions are

$$\frac{\partial \pi}{\partial s} : \sqrt{\frac{Q}{sr}} \frac{f}{2} - c + \sqrt{\frac{1}{srQ}} \frac{\lambda(1 - f)}{2} = 0, \tag{65}$$

$$\frac{\partial \pi}{\partial f} : \sqrt{\frac{sQ}{r}} - \lambda \sqrt{\frac{s}{rQ}} = 0, \tag{66}$$

$$\lambda \left(\sqrt{\frac{s}{rQ}} (1 - f) \right) \geq 0. \tag{67}$$

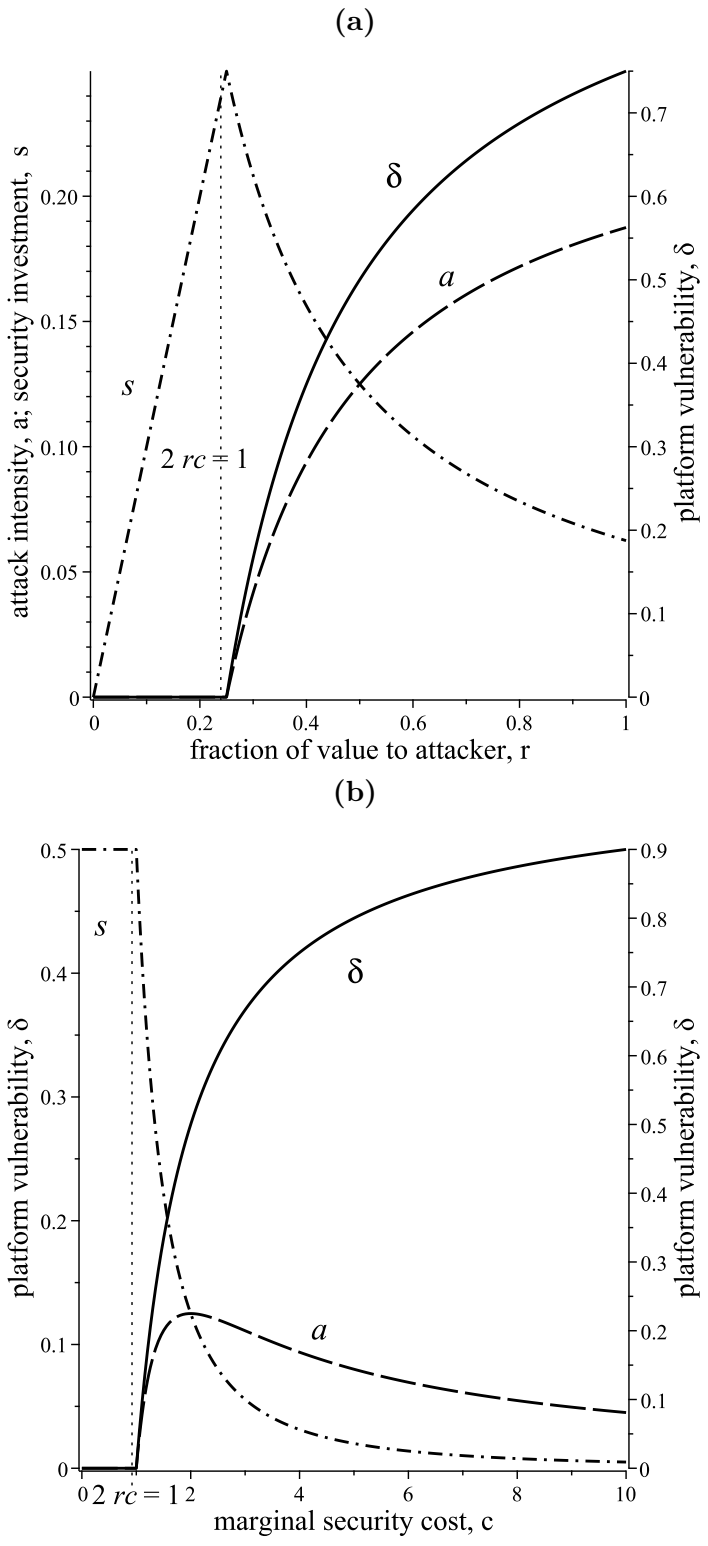
$\frac{\partial \pi}{\partial f} = 0$ implies $\lambda_M = Q$, which in turn leads to the optimal security investment

$$s_M = \begin{cases} \frac{Q}{4rc^2} & \text{if } 1 \leq 2rc \\ rQ & \text{if } 1 > 2rc. \end{cases} \tag{68}$$

The inequality condition then results in a fee $f_M = 1$.

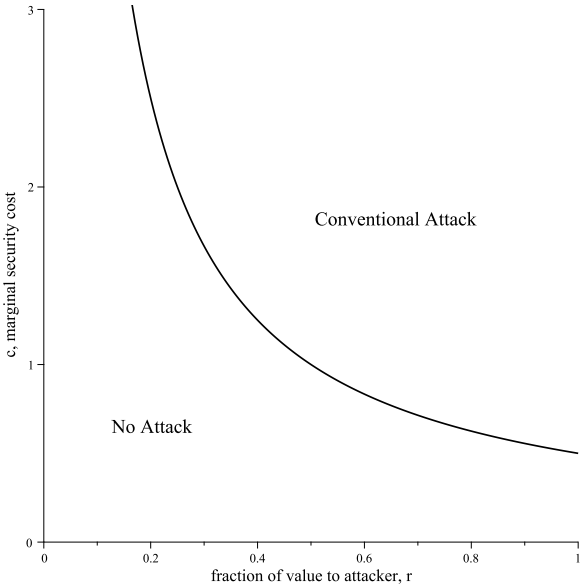
B Figures

Figure 2: Comparative Statics

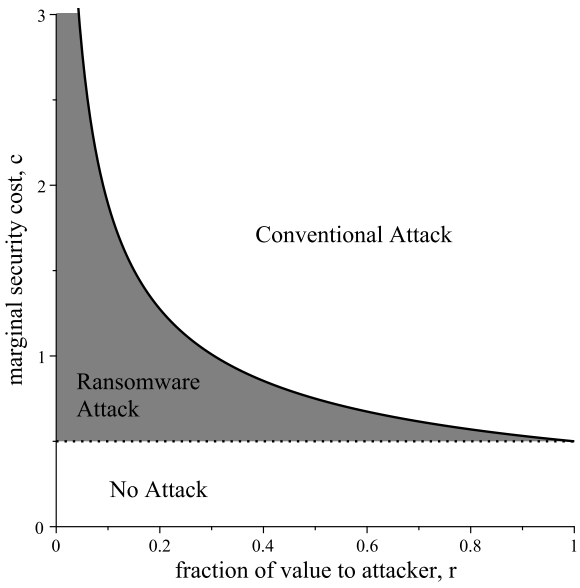


The figures describe the equilibrium behavior of a^* (dash line), s^* (dash-dot line), and δ^* (solid line) in parameters r (top) and c (bottom). Parameter values set $M = N$, $c = 2$ (top) and $r = 0.5$ (bottom). A vertical dotted line indicates the attack/no attack threshold, $2rc = 1$.

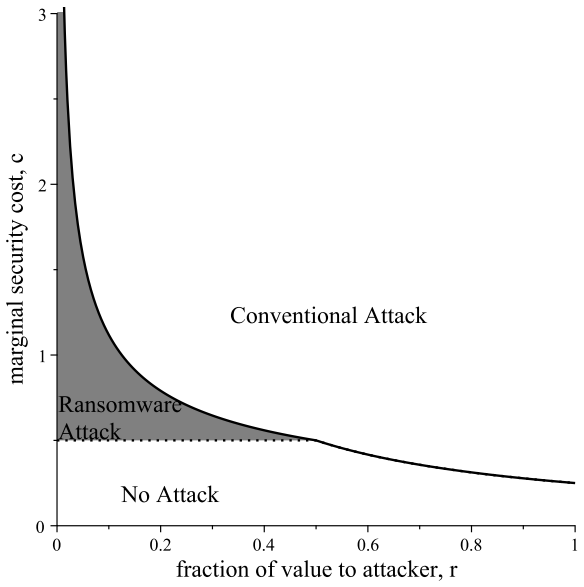
Figure 3: Attack Incidence and Type



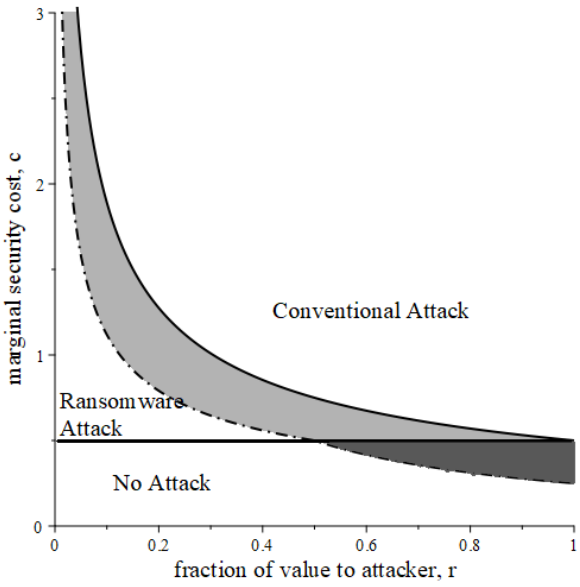
(a) Theft vs No Attack



(b) Theft vs Ransomware (Observable)

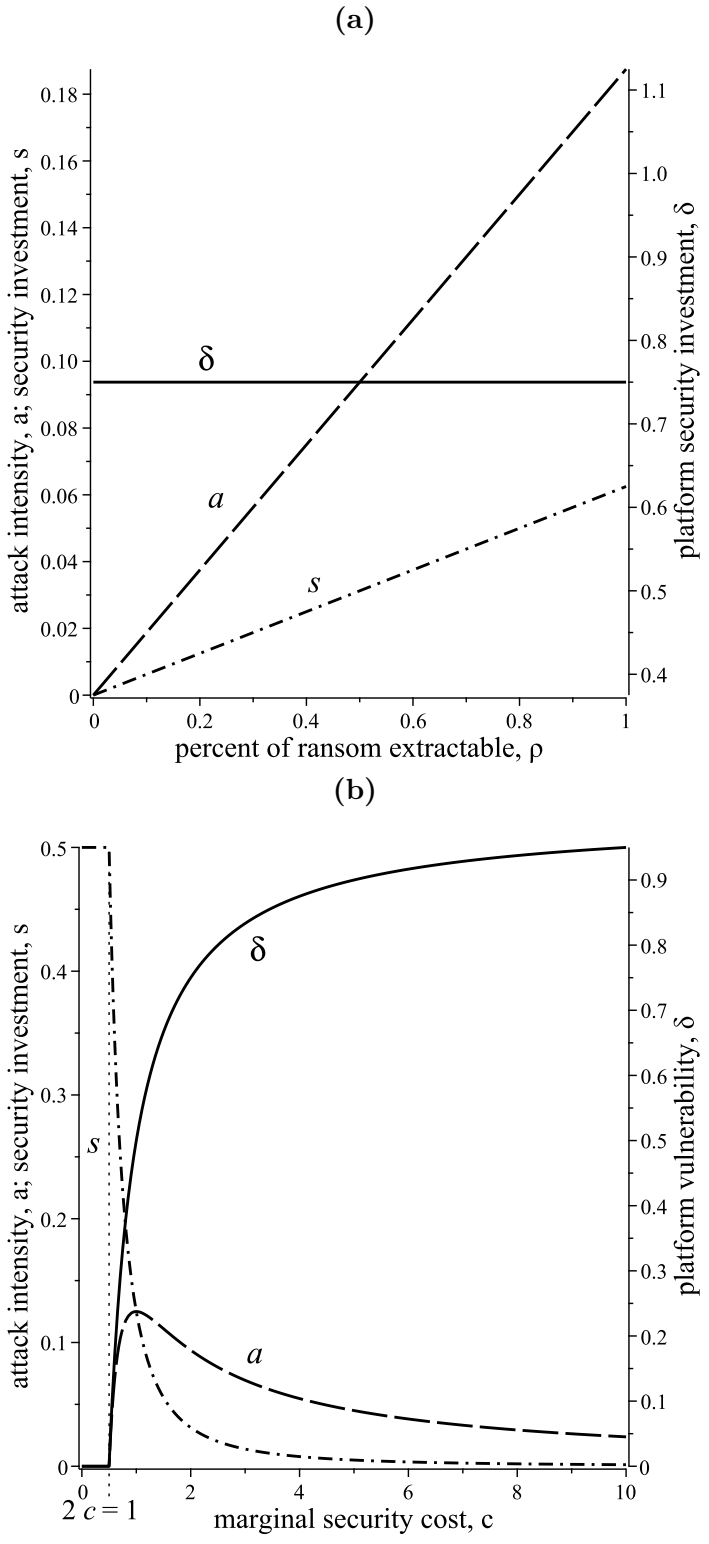


(c) Theft vs Ransomware (Unobservable)



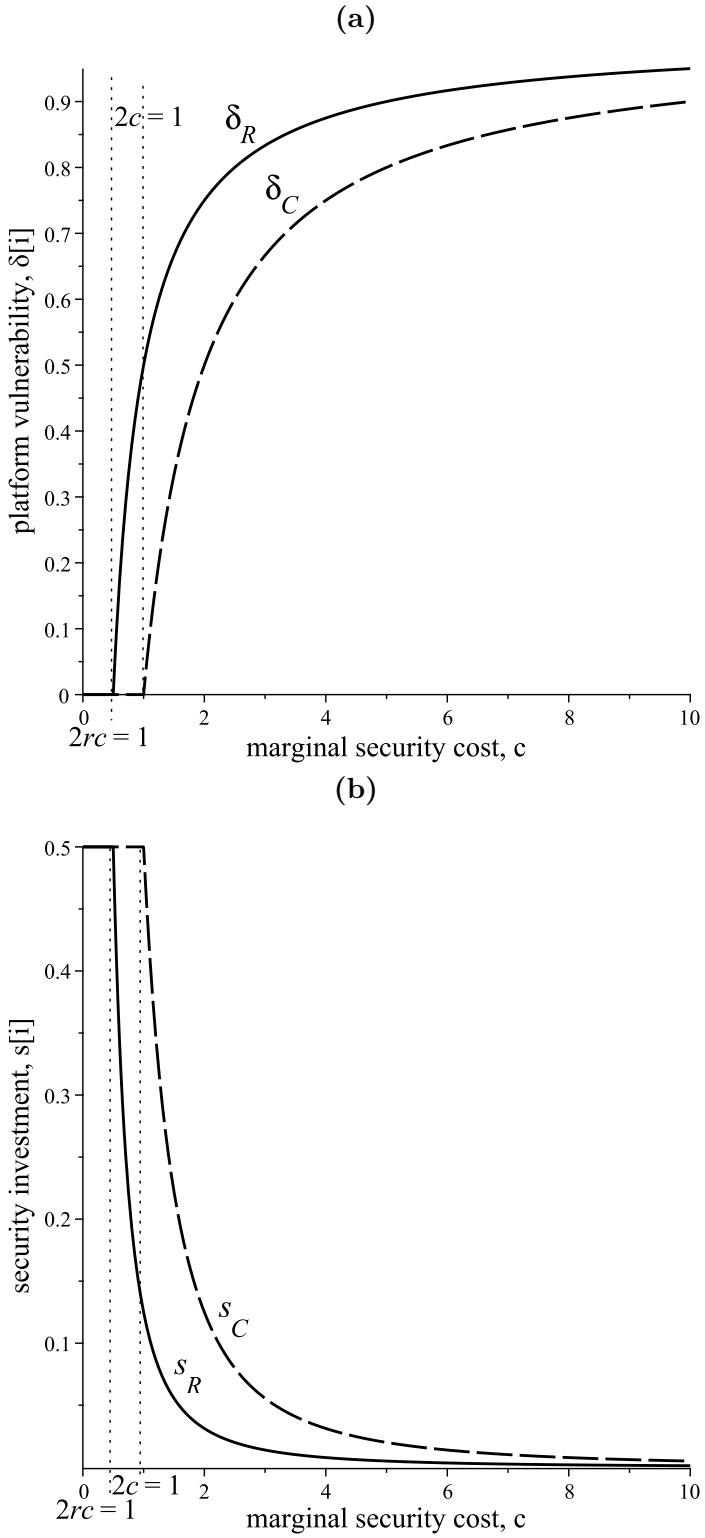
(d) Observable vs Unobservable Security

Figure 4: Ransomware Equilibrium Comparative Statics



The figures above describe the equilibrium behavior of a^* (dash line), s^* (dash-dot line), and δ^* (solid line) in parameters ρ (top) and c (bottom). Parameter values set $M = N$, $c = 2$ (left) and $\rho = 0.5$ (right). A vertical dotted line indicates the attack/no attack threshold, $2c = 1$.

Figure 5: Vulnerability and Security Comparative Statics



The figures above compare the equilibrium behavior of s^*, and δ^* in c across conventional (dash line) and ransomware (solid line) equilibria. Parameter values set $(M, r, \rho) = (N, 0.5, 0.5)$. A vertical dotted line indicates the attack/no attack thresholds, $2rc = 1$ (conventional) and $2c = 1$ (ransomware).