# The Impact of Cryptocurrency on Cybersecurity

Terrence August*
Rady School of Management
University of California, San Diego

Duy Dao†
Haskayne School of Business
University of Calgary

Kihoon Kim‡
Korea University Business School

Marius Florin Niculescu§
Scheller College of Business
Georgia Institute of Technology

Forthcoming in *Management Science*

Version: Jan 3, 2025

**Abstract**

Cryptocurrencies have prompted a shift away from classic security attacks toward ransomware-based extortion. To better understand the impact of cryptocurrencies on the cybersecurity landscape, we conduct a comparative analysis of cybersecurity metrics prior to and after the adoption of cryptocurrency using a series of connected software use models in the presence of security externalities. In the proposed framework, we endogenize the actions of both heterogeneous consumers and attackers, with entry of the latter being driven by both the size of the unpatched consumer population and, as a subset of it, the size of the ransom-paying consumer population. We first examine users' adoption and patching behavior under both security scenarios. We explore how changes in attacker entry costs impact outcomes under both conventional and post-crypto ransomware threat landscapes. We show that ransomware scenarios may be more desirable than conventional ones when attacker entry costs are low, provided that the gains from entering with standard attacks instead under the ransomware scenario are not too high. However, under such scenarios, social welfare can increase under the same conditions that lead to larger ransoms being demands and a higher expected total ransom being paid, which presents a conundrum to policymakers. We also examine the impact of market parameters associated with security losses from conventional attacks and residual losses when victims pay in ransomware attacks.

*Rady School of Management, University of California, San Diego, La Jolla, CA 92093-0553. e-mail: taugust@ucsd.edu

†Haskayne School of Business, University of Calgary, Calgary, AB T2N 1N4. e-mail: duy.dao@ucalgary.ca

‡Korea University Business School, Seoul, Korea, 136-701. e-mail: kihoon@korea.ac.kr

§Scheller College of Business, Georgia Institute of Technology, Atlanta, Georgia 30308. e-mail: marius.niculescu@scheller.gatech.edu

# 1 Introduction

Cryptocurrencies (crypto), the most prominent type of blockchain-based digital asset, have been heralded as a game changer for the payments industry, marking the beginning of an era of decentralized finance (DeFI), with promise of near-instant secure transactions, storage of funds in secure digital wallets, and accessibility via Internet from anywhere (Levy 2022). But for all the sung praise and promise, it is undeniable that there is a real dark side related to the adoption of crypto; namely, cryptocurrencies have become the go-to instrument for transactions associated with cybercriminal extortion-type activity due to the pseudonymity of crypto transactions (Black 2022). Such extortion-based cyberattacks, commonly referred to as *ransomware* attacks, have rapidly grown in frequency and impact as seen in the Financial Crimes Enforcement Network (FinCEN) 2011-2021 data reproduced in Figure 1.

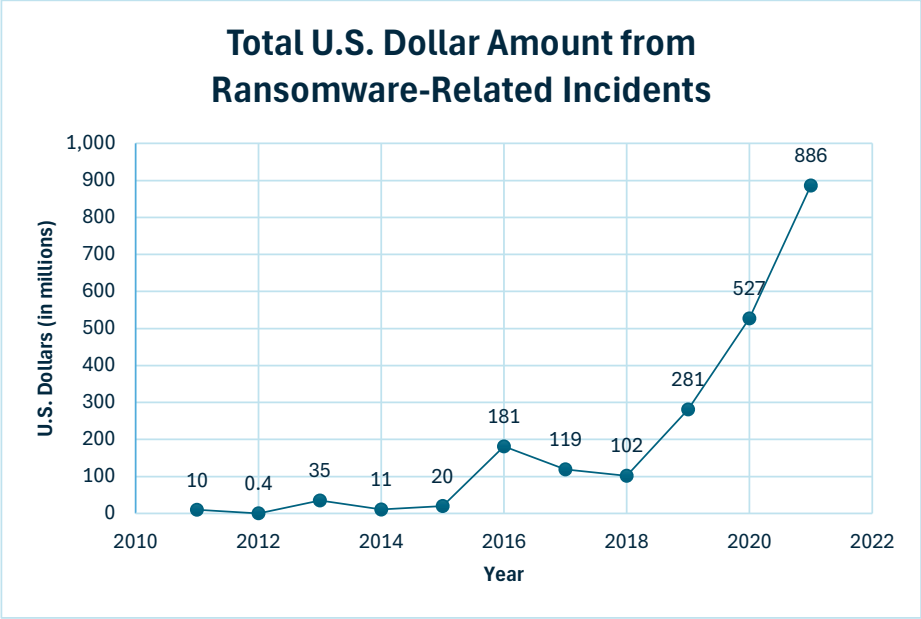


Figure 1: Data sourced from FinCEN ransomware trends report (FinCEN 2022). Values are computed as of "incident date" which reflects the date associated with payments.

Ransomware did exist prior to the introduction of Bitcoin in 2009 (even going back to 1989),

but its financial impact in earlier decades was negligible.[1] Once Bitcoin adoption gathered steam and crypto transaction tools and exchanges evolved into the mainstream, the stage was set for malicious actors to take ransomware to a whole new impact level, which occurred in 2013 when CryptoLocker became the first major ransomware strain combining advanced encryption locking with ransom demands in cryptocurrency (Crowdstrike 2022). Going beyond FinCEN's 2011–2021 data, researchers estimate that ransom payments reached $1.1 billion in 2023 (Chainalysis 2024).

While modern crypto-fueled ransomware attacks initially targeted proprietary software, we also see a large volume of these attacks on open-source software (OSS), particularly post 2021.[2] More broadly, cybercriminals are attacking OSS across industries using a variety of attack vectors and payloads even going beyond ransomware (Statista 2024). In fact, according to a 2024 study by Sonatype, 704,102 malicious packages targeting OSS have been identified since 2019 (Maudrill 2024). The reason for malicious hackers' interest in exploiting OSS vulnerabilities is simple: nowadays, leveraging OSS is a critical aspect of most firms' IT strategies. With OSS, a firm maintains significant flexibility with regard to customization, less lock-in with software vendors, lower software costs, and more. What is particularly noteworthy is the extent to which OSS components are being leveraged. The number of OSS contributions to GitHub (one of the most prominent repositories of OSS projects) reached 301 million in 2023 alone and nearly 30% of Fortune 100 companies started their own open source program offices to coordinate OSS strategies (GitHub 2024). RedHat's *The*

[1]The first documented ransomware attack, the AIDS Trojan (also called the "P.C. Cyborg attack") occurred in 1989 (Murphy-Kelly 2021).

[2]The Apache Log4j (Log4Shell) Java logging vulnerability, which exposed a large swath of web servers and services (including those using VMware Horizon) has been extensively used in the deployment of ransomware attacks starting at the end of 2021 and throughout 2022 (Tung 2022, Arctic Wolf Labs 2022). DarkRadiation, written in Bash script that targeted Red Hat, CentOS, and Debian Linux distributions as well as Docker cloud containers was spread via an SSH worm (Lakshmanan 2021, Zahravi 2021). Attackers such as the notorious REvil gang who was responsible for shutting down the operations of JBS, the world's largest meat supplier, and extorting $11 million in ransom, are now porting their ransomware to Linux (Vanian 2021). Being able to attack bare metal hypervisors like VMware's ESXi as well as Linux-based network storage devices is a lucrative opportunity for hackers; indeed, ransomware strains DarkSide and RansomEXX have evolved with OSS in their sights (Winder 2021, Spring 2021). Year 2022 saw an acceleration of Linux encryptors being developed and deployed by the most prominent ransomware gangs including, in addition to the aforementioned examples, Conti, LockBit, HelloKitty, AvosLocker, Babuk, PureLocker, Mespinoza, Hive, IceFire (Gatlan 2023).

*State of Enterprise Open Source* reports that 90% of surveyed IT leaders already use enterprise OSS (Red Hat 2021). According to Synopsis (2024), 96% of commercial codebases leverage to some extent OSS code (with 77% of the total code originating from OSS). Among surveyed codebases, 84% contained OSS security vulnerabilities (with 74% of the codebases including high-risk vulnerabilities), and no less than 91% contained components that were at least ten versions behind the current version of the component. This current industry reality presents hackers with opportunities to land widely disruptive attacks on OSS vulnerabilities, with substantial associated losses.

When examining the economics of open source as it relates to security, certain attributes are particularly salient. OSS is typically available for free and, therefore, can engender significantly larger user populations than its commercially available counterparts. Increased usage often goes hand-in-hand with increased *risky* usage in that a sizeable portion of the user population will find it incentive compatible to remain unpatched. Thus, historically with traditional (non-ransomware) attacks, risk management has centered on mitigation via policies to improve patching behavior (August and Tunca 2006). However, in the face of ransomware, this calculus gets shifted because users have an additional opportunity to contain losses by paying out ransoms instead. To be sure, a user with intent to pay ransom as part of its strategic defense clearly has dampened incentives to patch. Ceteris paribus, possessing an option to pay ransom instead of incurring larger losses can be good for consumer surplus. At the same time, attackers' strategies are shifting as they consider both ransomware and standard attacks. Ransomware can certainly streamline the monetization process by extracting payments directly from the victims; yet aggressive attacks, larger payouts, and publicized successes can instead lead to higher attack volumes and ultimately deter unpatched consumption (Jeffery and Ramachandran 2021).

A common presumption in the economics of cybercrime is that raising hacker costs and risks would be beneficial to cybersecurity. For example, national security strategy in the UK dictates

that "we need to *raise the cost*, raise the risk, and reduce the reward of cyber criminals' activity" and, similarly in the US, aims to "disrupt and dismantle threat actors" and has an approach that specifically includes " targeting the illicit cryptocurrency exchanges on which ransomware operators rely and improving international implementation of standards for combatting virtual asset illicit finance" (UK 2016, White House 2023). Attacker costs are critical in that they can significantly sway attacker entry behavior. However, the shifting of attacker entry costs can have a significant, *differential* impact when examining a pre-crypto environment versus a post-crypto one characterized by the presence of ransomware. Moreover, the presumption that raising costs will be helpful needs further consideration in that the strategic interaction between user behavior (adoption, protection, and ransom-paying strategies) and attacker behavior (entry and attack strategy) is highly complex.

Against this backdrop, in this paper, we perform a comparative analysis of the security landscapes and market outcomes under both a pre-crypto context characterized by traditional attacks and a post-crypto context where ransomware becomes a possibility. Given the enabling role that crypto plays in the execution of digital extortion attacks, this comparison allows us to identify the impact of cryptocurrency on cybersecurity. As part of our research agenda, we aim to shed light on the role of attacker entry costs in shaping cybersecurity market outcomes, while particularly highlighting that increased costs can unexpectedly be detrimental in the post-crypto ecosystem. Thus, reducing attacker entry costs may possibly lead to superior post-crypto outcomes in comparison to pre-crypto ones.

To the best of our knowledge, this is the first study to execute such a comparison while endogenizing the actions of both (a continuum of) heterogeneous consumers and (a continuum of) heterogeneous attackers, with the entry of the latter being driven by both the size of the unpatched consumer population and, as a subset of it, the size of the ransom-paying consumer population. Moreover, by focusing on a relatively new warfront in ransomware attacks, namely attacks on OSS,

we further extend the research agenda on the economics of ransomware.

Specifically, we study how attackers are drawn differently to vulnerable software across pre-crypto and post-crypto scenarios. Post-crypto, we characterize how ransomware-specific metrics such as the ransom-paying population size and expected total ransom paid are impacted by attacker entry costs, a relevant moderating variable for our comparative analysis as discussed above. We explore how consumer and attacker strategic behaviors ultimately impact social welfare, as well as making comparisons before and after cryptocurrencies changed the landscape. There are multiple economic forces at work. Under both settings, lowered entry barriers for attackers can increase pressure on consumers from a security risk perspective but can also dilute attacker returns due to crowding out effects. However, in the post-crypto ransomware setting, entering attackers also strategically choose the attack mode (with or without ransomware payload) and the ransom amount. This provides them some greater flexibility to manage the crowding out effect. Moreover, when some attackers choose ransomware, a portion of the unpatched consumers are able to mitigate, to some degree, the attack losses through ransom payments in the post-crypto setting. These forces shape the equilibrium segmentation of consumers and attackers, with the overall net effect on outcomes hinging on the extent of each segment's presence. Because, in equilibrium, unprotected consumers segment into ransom payers and non-payers, and attackers segment into those who enter with conventional attacks and those who enter with ransomware attacks, there are significantly more complex dynamics under post-crypto ransomware setting relative to the pre-crypto one. A key contribution of our study is that we formally demonstrate that the presence of ransomware can be socially beneficial, even in the presence of strategic attackers making self-interested entry decisions. We show that a *decrease* in attacker entry costs can unexpectedly shift post-crypto welfare outcomes higher relative to the pre-crypto scenario. Moreover, we also highlight that welfare can be increasing in the attacker entry costs on the same region where ransomware-specific metrics (expected ransom

paying population, expected ransom demand, and expected total ransom paid) are also increasing. For example, in that one might expect that decreasing expected total ransom paid would be helpful, rather than harmful (which we establish is a possibility), to social welfare, our results can help policymakers better understand how to approach the security problem in this domain and avoid policies that might focus in the wrong direction on certain metrics. We also conduct a numerical analysis that includes additional layers of attacker heterogeneity to demonstrate the robustness of our findings. Finally, we discuss how our exploration can help inform policymakers working in this space.

## 2 Literature Review

This study builds on several streams of research: (i) economic impact of cryptocurrencies, (ii) IT security, and (iii) economics of ransomware.

A lot of the recent literature on cryptocurrencies focuses on their technical aspects (as a direct application of blockchains). Only recently are we starting to see an emerging body of work on the economic impact of this type of digital money while the vast majority of extant work focuses on financial aspects of cryptocurrencies, as is summarized in recent research literature surveys by Halaburda et al. (2020), Bariviera and Merediz-Solà (2021), and Yue et al. (2021). While there is palpable excitement about the potential of cryptocurrencies (and decentralized finance in general) to bring significant change to the financial sector, there are some problematic effects of the adoption of cryptocurrencies that warrant closer scrutiny. First, the impact of mining for cryptocurrencies has been examined from an energy sustainability and environmental perspective in several studies, with particular focus on the Bitcoin protocol, which relies on a computationally intensive proof-of-work implementation (Badea and Mungiu-Pupazan 2021). Also, due to various degrees of pseudonymity, cryptocurrencies became the preferred go-to payment tools for illegal and

criminal activities (Foley et al. 2019). It is within this latter category that our study makes a novel contribution, by focusing on the impact of adoption of cryptocurrencies on the cybersecurity landscape.

As cryptocurrency payments represent the main enabler of ransomware attacks at such unprecedented scale, in order to assess the role cryptocurrencies have in shaping software adoption, security losses and overall social welfare, it is insightful to compare security outcomes across scenarios in the absence of ransomware threats (pre-crypto) and in the presence of ransomware threats (post-crypto). The extant literature on the economics of IT security has explored at depth pre-crypto (traditional) cyberthreat landscapes, tackling a wide spectrum of research questions on topics including interdependent network security risks (e.g., Gal-Or and Ghose 2005, August and Tunca 2011, Zhao et al. 2013), patching incentives and management (e.g., Ioannidis et al. 2012, Dey et al. 2015, August et al. 2019), management of vulnerability disclosure (e.g., Cavusoglu et al. 2007, Choi et al. 2010, Mitra and Ransbotham 2015), open-source software security (e.g., Witten et al. 2001, Swire 2005, Schryen and Rich 2010), cyber insurance (e.g., Böhme and Schwartz 2010).

On the other hand, the literature on economics of IT security in the presence of ransomware (post-crypto) has just recently gained traction. Ransomware attacks differ from conventional attacks in that, in the wake of the attack, some of the victims may be able to moderate the magnitude of the loss through the payment of a ransom. Laszka et al. (2017), Cartwright et al. (2019), and Ryan et al. (2022) explore interactions between ransomware attackers and victims in the context of bargaining games, a setup fitting the realities of targeted, sparse attacks. Vakilinia et al. (2021) propose a smart-contract solution to implement a mechanism to ensure attackers share the decryption keys after a ransomware attack. Galinkin (2021) explores options to decrease ransomware attacker incentives. Fang et al. (2022) study the decision to pay ransom from the perspective of a Bayesian game with incomplete information for both the hacker and victim. Cartwright and

Cartwright (2019) and Li and Liao (2020) explore the role of attacker reputation in the context of multiple-period repeated ransomware attacks. Dey and Lahiri (2021) analyze how banning ransom payments or executing corporate bailouts would affect welfare in a steady-state multi-period market with two participants. Balasubramanian (2021) and Yin et al. (2023) explore how cyber insurance impacts the frequency and severity of ransomware attacks as well as defender efforts. Hernandez-Castro et al. (2020) study how consumer market and ransomware attack parameters affect welfare in targeted attacks without interdependent risk. Ahnert et al. (2022) explore the impact of ransomware in financial markets. Zhao et al. (2021), August et al. (2022), and Cartwright and Cartwright (2023) explore ransomware in the context of interdependent security risks in networked environments.

The aforementioned ransomware studies consider either a single, endogenous attacker or an exogenously modeled threat. While there exists an established literature within the more general attacker-defender modeling domain accounting for multiple, endogenous attackers (Hausken and Bier 2011, Garcia et al. 2019, Xu and Zhuang 2019, DuBois et al. 2023), considerably less exploration has been done on this front in the ransomware threat context (which, as mentioned above, is different from the conventional threats context). In a related work in the context of economics of kidnapping, Selten (1988) propose a model of multiple attackers who contemplate perpetrating at most one attack per period and update their attack pattern over time. In their model, the size of the attacker population is exogenously determined (all attackers are active). Li and Whinston (2020) propose a multi-period Bayesian game with multiple attackers who strategically build reputation over time. In their study, the attackers' ransom choice is exogenous but the choice of attacking (entering) at each period is endogenized. We make a significant contribution to this line of work by constructing a model with multiple heterogeneous ransomware attackers who strategically decide whether to enter the market, which attack mode to employ, as well as set their ransom demand.

Raising the complexity but better capturing their interaction, we incorporate indirect network externalities between attackers and victims such that users' patching and ransom-paying decisions affect attackers' incentives, and attackers' entry decisions affect users' incentives. In addition, we capture direct crowding-out effect that attackers have on one another.

Importantly, there is a dearth of research *directly comparing* and *contrasting* traditional (pre-crypto) vs ransomware (post-crypto) cybersecurity outcomes. This is where our main contribution lies. As discussed in the introduction, there is a direct connection between the adoption of cryptocurrencies as payment methods and the proliferation of significant ransomware attacks. Thus, through a comparison of traditional cybersecurity scenarios and cybersecurity scenarios with ransomware, we capture the nuanced impact of cryptocurrencies on cybersecurity. August et al. (2022) compare and contrast traditional and ransomware risk scenarios in the context of proprietary software and strategic software vendors. Their model includes an exogenous ransomware threat and does not account for the role of the ransom paying population on the attack rate. However, in this work, we endogenize both consumer and attacker decisions, and consider a multi-attacker setting with strategic entry driven by expectations of both the size of the unpatched consumer population and, as a subset of it, the size of the ransom-paying consumer population. Moreover, we extend that exploration by focusing on a relatively new area in ransomware attacks, namely attacks on OSS. Ahnert et al. (2022) also compare traditional vs ransomware attacks, but their model considers a single malicious actor attacking multiple homogeneous platforms, each serving multiple homogeneous clients (with assumed symmetric multi-homing). The security investment decision is delegated to the platforms. The authors show how ransomware changes the relationship between client and platform. In contrast, we consider both multiple heterogeneous attackers and multiple heterogeneous consumers whose strategic decisions are endogenized. This allows us to model explicitly one of the anecdotal observations in the ransomware landscape, that more unpatched or

paying consumers encourage more attackers to join the landscape, and the ensuing market effects associated with such dynamics.

## 3 Model Description and Consumer Market Equilibrium

In order to highlight the impact cryptocurrencies have on the software security ecosystem, we develop a unifying modeling framework that facilitates a comparative analysis across two scenarios. The initial specification centers on the state of affairs prior to the cryptocurrency-fueled ransomware problem of today. We refer to this modeling context as *pre-crypto benchmark* ($BM$) setting. The second specification focuses on the context of the current times and it incorporates the characteristics of extortion attacks. We refer to this modeling context as *post-crypto ransomware setting* ($RW$).[3] First, we introduce a primary model that captures core interdependencies at play (in Section 3), and we explore equilibrium market behavior across these settings, as well as the potential harm and benefit of ransomware attacks on society. Subsequently, in Appendices C and D, we examine a generalized model with additional attacker heterogeneity to explore robustness of our results.

Our study centers on open-source software but the applicability of the model extends to a broader range of technologies that are offered for zero price and possess similar protective action opportunities and risk interdependence characteristics. Moreover, through a reframing of the setup, our model and results can also be used to generate insights in a more general context inclusive of other types of software.[4]

[3]Throughout the exposition, we use these terms interchangeably.

[4]For example, our model can inform in-use software with sunk prices. Focusing on the population of *existing* users, we can explore how, based on residual value of the product, consumers decide whether to abandon the product or continue to use it and adjust their security strategy based on the perceived threat landscape.

### 3.1 Pre-Crypto Benchmark Model ($BM$)

There is a unit mass continuum of consumers whose valuations of the software lie uniformly in $\mathcal{V} = [0, 1]$.[5] Users incur an adoption cost of $\kappa > 0$ if they choose to use the software.[6] Moreover, the software is used in a network setting, thus exposing consumers to security risks associated with its use. A security vulnerability can arise in the software in which case the provider makes a patch available. Users who do not apply the security patch are at risk. If a user decides to patch the software, the user will incur an expected cost of patching denoted $c_p > 0$ to immunize against the attack.[7] If the user decides not to patch the software, the probability that the user is randomly hit by a security attack depends upon the number of attackers who choose to enter the market, which we discuss next after introducing attacker agents. Ultimately, each user makes a decision to either *adopt*, $A$, or *not adopt*, $NA$. Similarly, the patching decision is either *patch*, $P$, or *not patch*, $NP$. The consumer action space is then given by $S^c_{BM} = \{(A, P), (A, NP), (NA, NP)\}$.

There is a unit mass continuum of attackers, each characterized by their skill $\theta$ distributed uniformly on $\Theta = [0, 1]$. In particular, an attacker with skill $\theta$ will incur a fixed cost of $\tau(1 - \theta)$, with $\tau > 0$, to enter the threat landscape, configure attack tools, initiate attack campaigns, risk being caught, etc. Thus, attackers with greater skill incur lower costs. The parameter $\tau$ captures the state of technology available to facilitate attacks and evade prosecution, and hence the cost burden for hackers.[8] Each attacker decides whether to *enter* the market, $E$, or *not enter*, $NE$. The

[5] In general, our analysis applies to all users, individual and business, in that decisions in the model are made on a per system basis. Business users can be expected to have multiple systems, and can make separate adoption and patching decisions on each of these systems separately. As long as business users only make decisions for a countable set of systems, our model can capture such complexities in business where many systems are managed by a corporate entity. For example, even a year after WannaCry ransomware worm was unleashed, Boeing corporation still had not patched a small number servers in their Commercial Airplanes division, and these systems ended up being compromised in March 2018 in a fairly isolated incident (Gates 2018).

[6] The adoption cost accounts for piloting, configuring and integrating the software with business processes.

[7] The patching cost accounts for the money and effort that a consumer must exert in order to verify, test, and roll-out patched versions of existing systems.

[8] We assume that the variable cost to replicate the attacks on multiple victim systems is negligible compared to the one-time cost to develop/configure needed malware tools and learn how to utilize them to perpetrate such attacks (e.g., some malware are designed with worm-like self-replicating capabilities which automate spread).

attacker action space is then given by $S^a_{BM} = \{E, NE\}$.

Let $\gamma : \Theta \rightarrow S^a_{BM}$ and $\sigma : \mathcal{V} \rightarrow S^c_{BM}$ denote the strategy functions of the attackers and the consumers, respectively, based on the heterogeneity of each group. Moreover, $\Phi = <\gamma, \sigma>$, denotes an *attacker-consumer strategy profile*.[9] Next, we lay out how attacker and consumer decisions are interdependent. For example, attackers' entry decisions affect consumer losses, and consumer patching behavior affects attackers' entry decisions. Under a candidate $\Phi$, the size of the entering attacker population is given by:

$$a(\Phi) \triangleq \int_{\Theta} \mathbb{1}_{\{\gamma(\theta) = E\}} \, d\theta \, . \tag{1}$$

We denote the probability that a consumer is randomly hit by a security attack by $q(a(\Phi))$, where $a(\Phi)$ is endogenously determined in equilibrium. The function $q(\cdot)$ is assumed to be increasing and concave, with $q(0) = 0$, $q(1) \leq 1$, $q'(0) < \infty$, and the attacker elasticity of risk is decreasing, i.e., $\eta'_q(a) < 0$ where $\eta_q(a) = (dq/q)/(da/a)$ such that there is a diminishing marginal increase in risk as more attackers enter, both in absolute and relative terms. If successfully attacked, an adopting consumer will incur expected security losses that are positively correlated with their valuation $v$. That is, consumers with high valuations will suffer higher losses than consumers with lower valuations due to opportunity costs, higher criticality of data and loss of business.[10] Consistent with related literature, we assume that the correlation is of first order, i.e., the loss that a consumer with valuation $v$ suffers if hit by an attack is $\alpha v$ where $\alpha > 0$ is a constant. Therefore, the expected

[9] We assume imperfect information in both pre-crypto and post-crypto contexts such that agents only know their own types as well as the overall type distributions. Attackers cannot directly observe valuations, and therefore can only make strategic decisions based on inferred consumer thresholds in equilibrium.

[10] For simplicity, we assume that all attacked victims will address the vulnerability to prevent future attacks, such that the probability of successful repeat attacks on the same system is negligible.

utility for a consumer with valuation $v$ is given by:

$$U^c_{BM}(v,\Phi) \triangleq \begin{cases} v-\kappa-c_p & if \quad \sigma(v)=(A,P)\,; \\ v-\kappa-q(a(\Phi))\times\alpha v & if \quad \sigma(v)=(A,NP)\,; \\ 0 & if \quad \sigma(v)=(NA,NP)\,. \end{cases} \tag{2}$$

Just as the entering attackers affect consumer utility as in (2), unpatched users attract attackers. We similarly denote the size of the *unpatched* consumer population in the market, contingent on profile $\Phi$, as

$$u(\Phi) \triangleq \int_{\mathcal{V}} \mathbb{1}_{\{\sigma(v)=(A,NP)\}}\, dv\,. \tag{3}$$

For notational simplicity, we omit subscript $BM$ on $a(\cdot)$ and $u(\cdot)$; however, we point out that quantities $a(\cdot)$ and $u(\cdot)$ will be different under $BM$ and $RW$ models as they are endogenously determined in equilibrium.

The total expected mass of compromised systems by attackers is $\Omega(\Phi) = q(a(\Phi)) \times u(\Phi)$. Each compromised system yields value $\rho > 0$ for the attacker, and each attacker has an equal chance to be the first to successfully compromise any unpatched consumer. Thus, each attacker expects to gain $\rho \times \Omega(\Phi)/a(\Phi) = \frac{q(a(\Phi))\times\rho\times u(\Phi)}{a(\Phi)}$ when entering. Thus, for an attacker with type $\theta$, the expected utility is given by:

$$U^a_{BM}(\theta,\Phi) \triangleq \begin{cases} \frac{q(a(\Phi))\times\rho\times u(\Phi)}{a(\Phi)} - \tau(1-\theta) & if \quad \gamma(\theta)=E\,; \\ 0 & if \quad \gamma(\theta)=NE\,. \end{cases} \tag{4}$$

Note that in (4) the number of entering attackers $a(\Phi)$ impacts attacker utility in two ways, capturing the strategic interactions among attackers. First, $a(\Phi)$ has a direct impact through $q(a(\Phi))/a(\Phi)$. This captures the direct crowding-out effect of attackers in that $\frac{d}{da}\left[\frac{q(a)}{a}\right] < 0$ under

the assumptions on $q(\cdot)$. Second, $a(\Phi)$ has an additional, indirect impact from $u(\Phi)$ (via the strategies $\Phi$). This is because $u(\Phi)$ is endogenously determined based on consumers maximizing utility in the face of security risk. Specifically, in the consumer's utility function in (2), $a(\Phi)$ is seen in the loss term should a consumer elect to adopt but not patch. Therefore, the size of attackers entering the threat landscape influences the equilibrium mass of unpatched consumers, as seen in (4).

We study a game of incomplete information where consumers and attackers only know the type distributions of others. In the benchmark game, all decisions are made simultaneously. For all agents, the equilibrium will satisfy a threshold structure as in the following.

**Lemma 1** *In the pre-crypto benchmark setting, the equilibrium attacker-consumer strategy profile* $\Phi^*_{BM} = <\gamma^*_{BM}, \sigma^*_{BM}>$ *is characterized by thresholds* $\bar{v}_u$, $\bar{v}_p$, $\bar{\theta} \in [0,1]$. *For a consumer of type* $v \in \mathcal{V}$, *the optimal strategy* $\sigma^*_{BM}$ *satisfies*

$$\sigma^*_{BM}(v) = \begin{cases} (A,P) & if \quad \bar{v}_p < v \leq 1\,; \\ (A,NP) & if \quad \bar{v}_u < v \leq \bar{v}_p\,; \\ (NA,NP) & if \quad 0 \leq v \leq \bar{v}_u\,. \end{cases} \tag{5}$$

*For an attacker with type* $\theta \in \Theta$, *the optimal strategy* $\gamma^*_{BM}$ *satisfies*

$$\gamma^*_{BM}(\theta) = \begin{cases} E & if \quad \bar{\theta} < \theta \leq 1\,; \\ NE & if \quad 0 \leq \theta \leq \bar{\theta}\,. \end{cases} \tag{6}$$

In the market equilibrium, high valuation consumers patch and protect their systems whereas mid valuation consumers adopt but remained unpatched giving rise to cybersecurity risk. The equilibrium unpatched population $u(\Phi^*_{BM}) = \bar{v}_p - \bar{v}_u$ provides an incentive for attackers to enter the market and strike unpatched systems as seen in the utility function expressed in (4). As a result, attackers can attain positive utility from entering provided that their costs are covered, i.e.,

the attacker skill is greater than $\bar{\theta}$. This gives rise to an equilibrium attacker population of size $a(\Phi^*_{BM}) = 1 - \bar{\theta}$.

### 3.2 Post-Crypto Ransomware Model ($RW$)

In the post-crypto context, in the presence of ransomware, the threat landscape changes. An attacker may still elect not to enter ($NE$). However, if they decide to enter, they choose the attack mode - standard attack similar to the pre-crypto benchmark setting ($E$) or ransomware attack ($ER$). In the case of the latter, the attacker also selects a ransom amount, thus making two decisions. The attacker action space in this setting becomes $S^a_{RW} = \{\{ER\} \times [0, \infty)\} \cup \{E\} \cup \{NE\}$.

Consistent with the pre-crypto benchmark setting, we assume an attacker entering with a standard attack ($E$) and skill $\theta$ will incur a fixed cost of $\tau(1-\theta)$, with $\tau > 0$, to enter the threat landscape, configure attack tools, and initiate the attack campaign. In the case of entry with ransomware ($ER$), there may be additional costs such that an attacker with skill $\theta$ incurs a fixed cost of $\tau(1+\omega)(1-\theta)$ where $\omega > 0$.[11] Hence, the additional cost of $\tau(1-\theta)\omega$ represents the skill-based, ransomware-specific *cost differential*. In this case, if a victim chooses not to pay ransom, they are monetized in a similar way as in a standard attack. We let $\tilde{\rho} > 0$ denote the return to an attacker under the ransomware scenario when either (i) the attacker enters with a standard attack or (ii) the attacker enters with ransomware, but the victim elects not to pay. Here, we permit $\tilde{\rho}$ under ransomware to differ from $\rho$ under the benchmark because the existence of ransomware as well as differences in the post-crypto landscape can alter the cybercrime economy in ways that give rise to distinct returns under the two scenarios.

On the consumer side, unpatched consumers that are successfully attacked face a post-attack choice between paying the ransom to mitigate the loss or not paying the ransom and incurring the

[11] There are additional costs associated with setting up and configuring systems to support the interactions of the ransom process, including the handling of crypto transactions, the encryption of data, and the generation and passing of decryptor tools, in addition to costs to reduce the risk of being caught.

full loss. Following August et al. (2022), we assume that a victim that agrees to pay ransom will incur an additional residual loss $\delta\alpha v$ with $\delta \in [0,1)$ from the attack.[12] Suppose an attacker with skill $\theta$ sets a ransom amount of $R$. Thus, if unpatched and hit by this attacker, a consumer of type $v$ incurs a loss of $R + \delta\alpha v$ if they pay the ransom. However, they incur a loss of $\alpha v$ if they decide not to pay the ransom. Thus, the consumer pays ransom if and only if $R \leq \alpha v(1-\delta)$. In that this depends on $v$, we define $R_{max}(v) = \alpha v(1-\delta)$ as the maximum ransom amount that a consumer of type $v$ would be willing to pay if hit with ransomware. If they are hit by an attacker of type $\theta$ such that $R > R_{max}(v)$, then they would not pay that ransom. If they are hit by an attacker of type $\theta$ such that $R \leq R_{max}(v)$, then they would pay the ransom. We define a ransom mapping $R(\theta) = R$ for all $\theta \in \Theta$ to represent a candidate set of ransoms.

Let $\tilde{\gamma} : \Theta \to S^a_{RW}$ and $\tilde{\sigma} : \mathcal{V} \to S^c_{RW}$ denote the strategy functions of the attackers and the consumers, respectively, based on the heterogeneity in each group. For the given attacker-consumer profile $\tilde{\Phi} = <\tilde{\gamma}, \tilde{\sigma}>$, the expected utility for a consumer with valuation $v$ is given by:

$$U^c_{RW}(v, \tilde{\Phi}) \triangleq \begin{cases} v - \kappa - c_p & if \;\; \tilde{\sigma}(v) = (A, P)\,; \\ v - \kappa - q(a(\tilde{\Phi})) \times L(v) & if \;\; \tilde{\sigma}(v) = (A,\, NP)\,; \\ 0 & if \;\; \tilde{\sigma}(v) = (NA,\, NP)\,, \end{cases} \tag{7}$$

where the attacker population is given by $a(\tilde{\Phi}) \triangleq \int_\Theta \mathbb{1}_{\{\tilde{\gamma}(\theta) \neq NE\}}\, d\theta$ and the conditional expected loss if hit is $L(v) \triangleq L_1(v) + L_2(v)$. Here, $L_1(v)$ represents the conditional expected loss in the event the victim ends up paying ransom and is given by $L_1(v) = \int_\Theta \Big[\mathbb{1}_{\{\tilde{\gamma}(\theta) = ER \cap R(\theta) \leq R_{max}(v)\}} \times (R(\theta) + \delta\alpha v)\Big] \times \frac{1}{a(\tilde{\Phi})} d\theta$. This loss is induced by attackers who enter with ransomware attacks and choose ransoms below the consumer of type $v$'s maximum willingness to pay ransom. On the other hand, $L_2(v)$ represents the conditional expected loss in the event the victim is hit with a ransom

[12] Research indicates that total ransom attack losses can be as high as seven times the ransom demand (Blosil 2022). A study by IBM Security (2022) reports victims incurring average ransomware costs of $4.54 million in excess of the ransom demand. Residual losses can be due to investigation and remediation, lawsuits, reputation impact, etc.

amount higher than their maximum willingness to pay or is hit with a standard attack and can be expressed as $L_2(v) = \int_\Theta \left[ \mathbb{1}_{\{(\tilde{\gamma}(\theta)=ER \cap R(\theta) > R_{max}(v)) \cup \tilde{\gamma}(\theta)=E\}} \times \alpha v \right] \times \frac{1}{a(\tilde{\Phi})} d\theta$.

Ransomware attacks and conventional attacks often exploit a vulnerability to gain access to a system and activate a backdoor. For example, a recent report by IBM (2023) identifies the deployment of backdoors as the most common approach across all types of cyberattacks in 2022. Consistent with these observations and to maintain parity across contexts, as in the $BM$ scenario, we assume $q(a(\tilde{\Phi}))$ is the probability that a consumer is randomly hit by a security attack. Above, $1/a(\tilde{\Phi})$ is the conditional density of it originating from each type of entering attacker. As seen in (7), a user that adopts but does not patch ($\tilde{\sigma}(v) = (A, NP)$) incurs an expected security loss that accounts for the likelihood it falls victim to any of the attackers that enter the landscape (either paying ransom or incurring standard losses).

Again, $u(\tilde{\Phi}) \triangleq \int_{\mathcal{V}} \mathbb{1}_{\{\tilde{\sigma}(v) = (A,NP)\}} \, dv$ denotes the size of the unpatched adopter population. Similar to model $BM$, here we also omit subscript $RW$ for quantities $u(\cdot)$ and $a(\cdot)$ for ease of exposition. However, we remind the reader that these quantities are endogenously determined in equilibrium and hence, potentially different under the two models. Given strategy profile $\tilde{\Phi}$, if an attacker sets a ransom amount of $R$, then the size of the unpatched adopter population that is willing to pay this ransom is given by:

$$r(R, \tilde{\Phi}) \triangleq \int_{\mathcal{V}} \mathbb{1}_{\{\tilde{\sigma}(v) = (A,NP) \cap R_{max}(v) \geq R\}} \, dv \, . \tag{8}$$

For any of the unpatched systems, each attacker that enters the landscape has the same chance to be the one successfully compromising that system, at the rate $\frac{q(a)}{a}$. Hence, an attacker with type

$\theta$ receives expected net utility:

$$U^a_{RW}(\theta, \tilde{\Phi}) \triangleq \begin{cases} \frac{q(a(\tilde{\Phi}))\times\left[\tilde{\rho}\times(u(\tilde{\Phi})-r(R,\tilde{\Phi}))+R\times r(R,\tilde{\Phi})\right]}{a(\tilde{\Phi})} - \tau(1-\theta)(1+\omega) & if \quad \tilde{\gamma}(\theta) = (ER, R)\,; \\ \frac{q(a(\tilde{\Phi}))\times\tilde{\rho}\times u(\tilde{\Phi})}{a(\tilde{\Phi})} - \tau(1-\theta) & if \quad \tilde{\gamma}(\theta) = E\,; \\ 0 & if \quad \tilde{\gamma}(\theta) = NE\,. \end{cases} \tag{9}$$

Notably, (9) is consistent with (4) when an attacker enters with standard attacks. However, for an attacker with type $\theta$ entering with ransomware and choosing ransom amount $R$, their payoff can come from two streams: (i) unpatched consumers who are willing to pay $R$ (i.e., a group of size $r(R, \tilde{\Phi})$) and (ii) unpatched consumers who are not willing to pay $R$ and get monetized in the standard way (i.e., the remainder of the unpatched population $u(\tilde{\Phi}) - r(R, \tilde{\Phi})$).

In equilibrium, when attackers choose to enter with ransomware, they have the flexibility to set a type-dependent ransom amount. However, attacker heterogeneity appears only at the entry cost level and does not impact the chance of a successful attack. Then conditional on entry with ransomware, all attackers face the same objective function when determining their optimal ransom. Consequently, in equilibrium, ransomware attackers will optimally set their ransom demand to a common endogenously determined amount $R(\theta) = R^*$, which is to say that their optimal ransoms do not depend on their type. In Appendices C and D, we present a generalized model with attacker heterogeneity impacting both revenues and costs which, in turn, leads to type-dependent ransom demands, and we illustrate the robustness of our results under this more general setting. To streamline the exposition, we utilize the simplified model here in the main body of the paper.

Similar to the pre-crypto benchmark setting, under post-crypto ransomware setting, the market equilibrium has a threshold structure as follows:

**Lemma 2** *In the post-crypto ransomware setting, the equilibrium attacker-consumer strategy profile* $\tilde{\Phi}^*_{RW} = <\tilde{\gamma}^*_{RW}, \tilde{\sigma}^*_{RW}>$ *is characterized by thresholds* $\tilde{v}_u$, $\tilde{v}_p$, $\tilde{\theta}, \hat{\theta} \in [0, 1]$ *and ransom demand* $R^*$.

*For a consumer of type $v \in \mathcal{V}$, the optimal strategy $\tilde{\sigma}^*_{RW}$ satisfies*

$$\tilde{\sigma}^*_{RW}(v) = \begin{cases} (A,P) & if \quad \tilde{v}_p < v \leq 1\,; \\ (A,NP) & if \quad \tilde{v}_u < v \leq \tilde{v}_p\,; \\ (NA,NP) & if \quad 0 \leq v \leq \tilde{v}_u\,. \end{cases} \tag{10}$$

*For an attacker with type $\theta \in \Theta$, the optimal strategy $\tilde{\gamma}^*$ satisfies*

$$\tilde{\gamma}^*_{RW}(\theta) = \begin{cases} (ER,R^*) & if \quad \tilde{\theta} < \theta \leq 1\,; \\ E & if \quad \hat{\theta} \leq \theta \leq \tilde{\theta}\,; \\ NE & if \quad 0 \leq \theta < \hat{\theta}\,. \end{cases} \tag{11}$$

High valuation consumers continue to patch and protect their systems. On the other hand, mid to high valuation consumers adopt and remain unpatched, but elect to pay ransoms if ever hit. Low to mid valuation consumers also adopt and remain unpatched but instead elect not to pay ransoms when compromised. Finally, low valuation consumers choose not to adopt. In this case, the equilibrium unpatched population $u(\tilde{\Phi}^*_{RW}) = \tilde{v}_p - \tilde{v}_u$ (of which a subset will pay ransom, as described in (8), and a subset will not and incur standard losses instead) provides an incentive for attackers to enter the market and strike unpatched systems as seen in the utility function expressed in (9). Similar to the pre-crypto benchmark setting, attackers with types higher than $\hat{\theta}$ enter. This gives rise to the attack likelihood $q(a(\tilde{\Phi}^*_{RW}))$ where $a(\tilde{\Phi}^*_{RW}) = 1 - \hat{\theta}$, as seen in the consumer utility function expressed in (7).

### 3.3 Focal Region

In the following, we will narrow our attention to a subset of the parameter space, calling this our *focal region*. While in the entirety of the parameter space, we can enumerate and characterize all

of the different market structures that obtain in equilibrium, we focus on the more salient ones that are typically observed in these practical software contexts. In particular, our focal region consists of the parameter conditions under which, in both pre-crypto benchmark ($BM$) and post-crypto ransomware ($RW$) settings, we have *all* consumer segments present with positive mass in equilibrium. Also, some attackers will choose to enter while some will stay out of the market. Under $BM$, we have some consumers who opt out of the software, some who use the software but do not patch, and some who use the software and patch. Moreover, under $RW$, consumers that use the software but stay unpatched are split into those who pay ransom and those who do not.

The focal parameter region resembles market realities in that all consumers segments are in play. More specifically, we observe in practice the existence of both patched and unpatched user segments, with a portion of the latter segment paying ransom. In 2022, the majority of ransomware attacks exploited old, known bugs (Vijayan 2023). In fact, 76% of those flaws were from 2019 or before - thus, over three years old at that time. All users had an option to protect but some chose not to do so. Consumers with greater valuation at risk prioritize patching soon after patch release, often using patch management solutions (Kaseya 2020, Polaris Market Research 2024).

In terms of the magnitude of the impact of the cyberincidents on business consumers, as of 2024, according to analysis by Sophos, ransomware attacks exploiting *unpatched vulnerabilities* have the greatest business impact (Adam 2024). In particular, attacks that started with an exploited vulnerability succeeded in 75% of the instances in compromising backups and in 67% of the cases were followed by data encryption. With backups compromised and data encrypted, no less than 71% of these organizations chose to pay the ransom when the attack vector was an unpatched vulnerability (Adam 2024). Thus, in reality, the ransom paying population represents a sizeable segment of the unpatched population.

The parameter conditions that give rise to equilibria matching these outcomes are: a low soft-

ware adoption cost ($\kappa < \bar{\kappa}$), a low residual loss to consumers who pay ransom ($\delta < \bar{\delta}$), an intermediate security loss factor ($\underline{\alpha} < \alpha < \bar{\alpha}$), an intermediate attacker entry cost parameter ($\underline{\tau} < \tau < \bar{\tau}$), an intermediate additional attacker entry cost associated with ransomware ($\underline{\omega} < \omega < \bar{\omega}$), and a sufficient attacker value from executing standard attacks in both benchmark ($\rho > \underline{\rho}$) and ransomware ($\tilde{\rho} > \underline{\tilde{\rho}}$) scenarios. The essence of the conditions is that parameter values need to be within a reasonable range such that consumer segments do not disappear and there is partial entry of attackers into the market. It is important to note that our focal region only serves to simplify the paper while still establishing the main insights; our insights can be shown to hold over broader regions and other market structures. We demonstrate the existence and characterization of the bounds denoted above in Appendix A.

# 4 Equilibrium Behavior Comparative Analysis

Users who choose to remain unpatched in equilibrium attract attackers and thereby contribute to losses across the market due to this security externality. Hence, consumers' patching behavior is influenced by attacker entry behavior and vice versa. Before diving into the comparative analysis of the benchmark and ransomware scenarios, it is important to note that lower adoption costs tend to exacerbate the security landscape.

**Remark 1** *Under both pre-crypto benchmark and post-crypto ransomware settings, a decrease in consumer adoption cost* ($\kappa$) *leads to more attackers entering in equilibrium, and hence higher security risk.*

Taking $\kappa$ smaller reflects the lower adoption costs associated with OSS, and Remark 1 emphasizes that higher security risk will result for such applications. Because lower adoption costs incentivize more consumers to use the software and some elect to do so in a risky, unpatched man-

ner, a greater number of attackers find it beneficial to enter the market and cause this risk. In contrast, with proprietary software, the price serves dual purposes. Not only does it drive revenue for the vendor but, importantly, the price helps to shrink the user population (by increasing adoption costs) which provides less incentive for attackers to enter and reduces risk in turn.

Therefore, in the OSS world, it is important to understand how patching and attacking behaviors vary across the pre-crypto benchmark and post-crypto ransomware settings to understand how to better manage this critical driver of risk. In the following, we discuss how attacker entry and the equilibrium level of risk are impacted by cost changes associated with attacker technology.

**Proposition 1** *An increase in attacker entry costs affects the size of the equilibrium attacker population differently under pre-crypto benchmark and post-crypto ransomware settings:*

*(i) In the pre-crypto benchmark setting, an increase in attacker entry costs always leads to fewer attackers in equilibrium and hence reduced security risk.*

*(ii) In the post-crypto ransomware setting, an increase in attacker entry costs can lead to either fewer or more attackers in equilibrium. Specifically, the size of the equilibrium attacker population and corresponding security risk increase when $\tau$ is intermediate and decrease (consistent with the pre-crypto benchmark setting) when $\tau$ is either small or large.*

In this proposition, we study how the equilibrium and related metrics vary due to changes in attacker entry costs ($\tau$). First, it is not simple to directly argue that there are less attackers when entry costs rise because any pressure that reduces attacker entry also reduces the risk faced by consumers, which, in turn, increases their incentive to use the software product in an unpatched manner, which indirectly increases the incentive for attackers to enter again. We begin by discussing the pre-crypto benchmark setting using a contradiction argument. Suppose that the attacker population $a$ increases in the entry costs for some interval of $\tau$. The probability a consumer

is randomly hit, $q(a)$, increases in $a$; therefore, consumers bear more risk if choosing to be an unpatched user. By (2), the marginal consumer who is indifferent between being unpatched and not using will no longer use. Similarly, the marginal consumer who is indifferent between patching and being unpatched will shift toward patching. Altogether, $u$ shrinks as a result. Examining the attacker utility in (4), both $q(a)/a$ and $u$ are decreasing in $a$. Thus, the attacker's revenue decreases in tandem with an increase in entry cost, which would effectively reduce the attacker population in equilibrium, contradicting the supposition that $a$ can be increasing in $\tau$.

However, in the post-crypto ransomware setting, the decision problem is more complex for both consumers and attackers, which can lead to a different relationship between the equilibrium number of attackers and changes to the entry cost. First, note that an attacker always prefers to set a ransom that satisfies $R \geq \tilde{\rho}$ if it enters as a ransomware attacker because it can alternatively enter as a standard attacker and generate $\tilde{\rho}$ per victim. From a consumer standpoint, by (7) and because $R_{max}(v) = \alpha v(1-\delta)$, it is necessary that $\alpha v(1-\delta) > \tilde{\rho}$ for some consumers who choose to remain unpatched in equilibrium, otherwise an attacker will prefer standard attacks.

We begin with $\tau$ near the lower end of the focal region. In this region, as $\tau$ declines, there is an incentive for a lot of attackers to enter at low cost which, in turn, is associated with a higher $q$ and hence higher risk that all consumers face. As a result, consumers with higher valuations will have a greater incentive to patch and protect their systems which is to say that the threshold valuation consumer who is indifferent between patching and not patching will have a tendency to decline. This behavior puts a downward pressure on the ransom price in that the attackers need consumers to be unpatched in equilibrium for attackers to thrive. In order to counter the decline of the aforementioned threshold valuation, attackers will reduce ransom price; however attackers are willing do so to the extent that their ransom prices remain above $\tilde{\rho}$, as this is the outside option. Therefore, as $\tau$ gets small, at some point attackers run out of room to lower price and can no

longer provide sufficient incentives for consumers to pay ransom. Because attackers vary in their additional cost of entry particular to ransomware, as $\tau$ gets small, more and more attackers switch from entering with ransomware to entering as standard attackers, and the landscape converges to the benchmark case.

When $\tau$ is in the middle range, the dynamics in play are significantly different because of attacker strategic behavior as they sharply modify ransom prices and switch from standard attack to ransomware attack strategies. In panel (e) of Figure 2, we depict how the ransom price increases as the cost of entry increases through the focal region. Notably, in the middle of the range, attackers choose in equilibrium to increase ransom prices nonlinearly. This occurs because $\tau$ reaches a level where a sufficient amount of attackers have exited the market due to the same driving effects discussed in the lower $\tau$ region. This smaller attacker population induces an increase in the unpatched population at the higher portion of the valuation space as consumers strategically switch from being patched to being unpatched. This, in turn, permits higher ransom levels to be charged due to the higher valuations being present in the unpatched population. As $\tau$ increases, attackers simultaneously accelerate both the increase in ransom price and their transition away from standard attacks toward ransomware attacks as they benefit from these higher prices. This can be seen in panels (e) and (b), respectively, of Figure 2. This shift in attacker strategy toward ransomware attacks affects the composition of consumers' losses, with a significantly higher probability of being hit by ransomware attacks than standard attacks. However, losses in ransomware attacks are partially mitigated in that consumers can pay the ransom price whereas with standard attacks, victims can incur substantial losses as they are valuation dependent (see the bracketed expression in (7)); in fact, consumers can be better off with a greater proportion of ransomware attacks even if the ransom prices are somewhat higher. As such, unpatched usage also increases at an accelerated rate in $\tau$ as can be seen in panel (a) of Figure 2. This drastic increase in $u$ encourages more

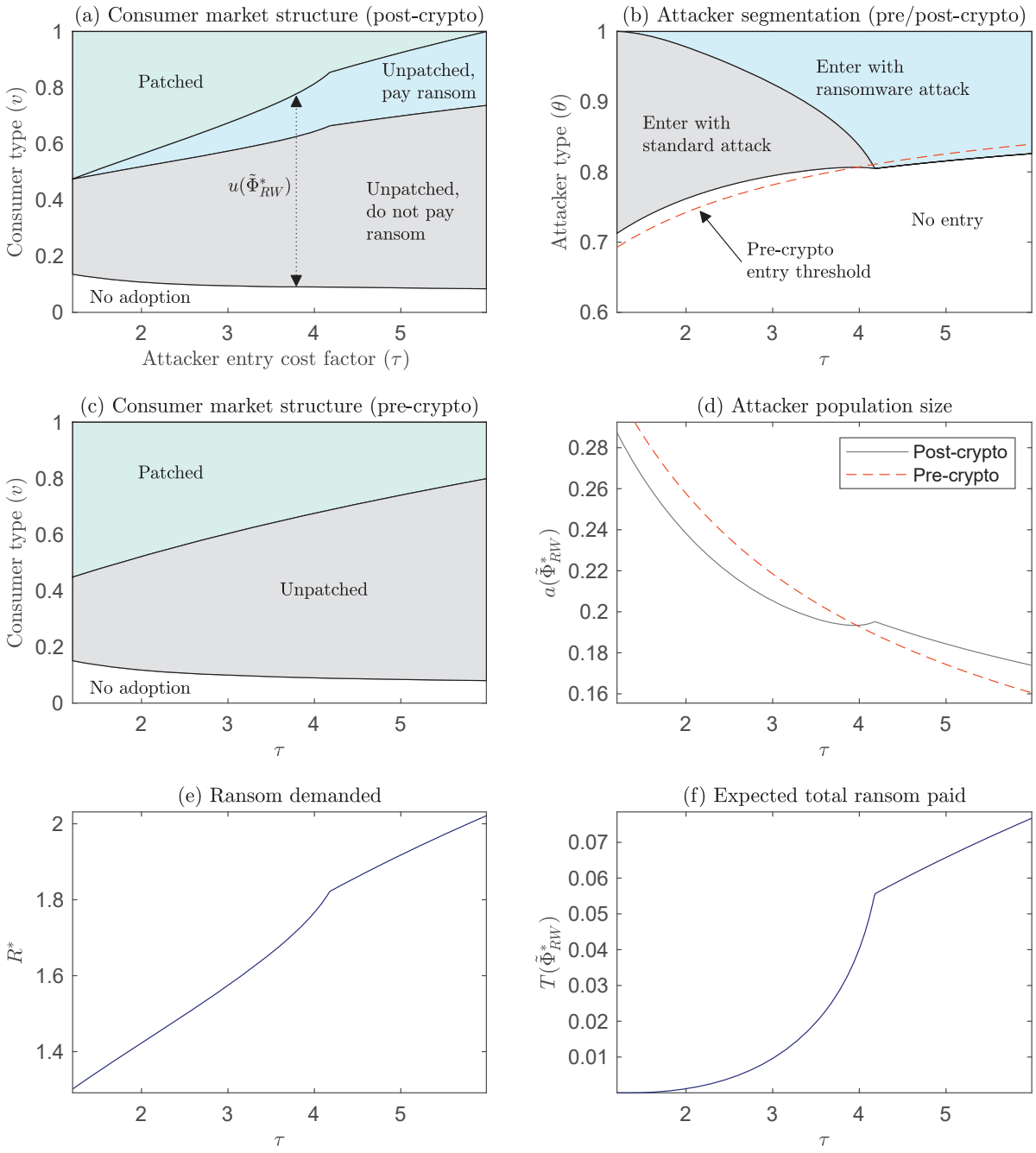


Figure 2: The impact of the attacker entry cost factor ($\tau$) on equilibrium market outcomes under pre-crypto benchmark and post-crypto ransomware settings. The parameter values are: $c_p = 0.3$, $\delta = 0.02$, $\kappa = 0.05$, $\alpha = 2.8$, $\rho = 1.6$, $\tilde{\rho} = 1.3$, $\omega = 0.1$, and $q(a) = 0.9a - 0.4a^2$.

attackers to enter the market as standard attackers as it dominates the negative impact of higher entry cost ($\tau$) and crowding out effects on revenue (decreasing $q(a)/a$ in $a$). This increase in the

attacker population can be seen in panel (d) of Figure 2.

As $\tau$ further increases, a point is reached where standard attacks are no longer sustained in equilibrium (e.g., this behavior can be seen around $\tau = 4.2$ in panel (b) of Figure 2 for this specific numerical illustration). Because attackers are no longer switching from standard attacks to ransomware attacks at this higher range of $\tau$, the effect discussed above where the attacker population ultimately increases as a result is no longer present. This is because the unpatched population cannot be rapidly increased (as $\tau$ increases) without attackers significantly transitioning from entering with standard attacks to entering with ransomware. Solely manipulating the ransom price is insufficient to generate this effect. Instead, the attacker population will decrease in $\tau$ throughout this region which, in turn, leads to a relatively flatter increase (in comparison to the intermediate $\tau$ region) in the unpatched consumer population, which can be seen in panels (d) and (a), respectively, of Figure 2. Moreover, under higher $\tau$, the attacker population decreases at a slower rate than seen in the benchmark case because attackers can leverage an ability to charge ransoms to higher valuation consumer types to preserve revenue.

**Proposition 2** *Under the post-crypto ransomware setting, there exists $\hat{\tau} > 0$ such that, under the focal region, when $\tau > \hat{\tau}$, the ransom-paying population size, equilibrium ransom charged, and expected total ransom paid all strictly increase in $\tau$, despite the attacker population size decreasing in $\tau$.*

The consumer market equilibrium structure is characterized by threshold values $\tilde{v}_u$ and $\tilde{v}_p$ marking the indifference threshold between being an unpatched user and being a non-user, and between being a patched and an unpatched user, respectively. Moreover, these thresholds satisfy $\tilde{v}_u < \frac{R^*}{\alpha(1-\delta)} < \tilde{v}_p$. Thus, we can define $\tilde{v}_r = \frac{R^*}{\alpha(1-\delta)}$ as the threshold, unpatched consumer indifferent between paying ransom and incurring losses. Therefore, the mass of consumers who are unpatched and prefer to pay ransom in equilibrium is $r(\cdot) = \tilde{v}_p - \tilde{v}_r$. Also, when $\tau > \hat{\tau}$ (in the numerical

illustration in Figure 2, $\hat{\tau} \approx 4.2$), the attacker market equilibrium structure is characterized by a single threshold $\tilde{\theta}$ marking the attacker type who is indifferent between entering with ransomware and not entering the market.[13] Therefore, the mass of attackers entering in equilibrium is given by $a^* = 1 - \tilde{\theta}(\tau)$. For a given $\tau$, the expected total ransom paid can then be computed as $T = q(1 - \tilde{\theta}(\cdot)) \times R^*(\cdot) \times r(\cdot)$.

Taking these three components one at a time, we begin with $q(\cdot)$. As $\tau$ reaches a higher level, by part (ii) of Proposition 1, the equilibrium number of attackers decreases in $\tau$. As such, $q(a^*)$ decreases. For the second component, the equilibrium ransom $R^*(\cdot)$, with fewer attackers, there is an opportunity for more consumers to be remain unpatched. This, in turn, provides attackers with an ability to charge higher ransoms in that there are more higher valuation consumers remaining unpatched which can be monetized. As a result, attackers will set an optimal ransom that is gradually increasing in $\tau$. Moving to the last component, $r(\cdot)$, a higher ransom can make some consumers less willing to pay this ransom, and this intuition is true for consumers near the $\tilde{v}_r$ threshold who will switch to not paying ransom. However, it turns out that consumers near the $\tilde{v}_p$ threshold have an incentive to switch *toward paying ransom* instead of patching due to the reduction in the number of attackers and reduced risk, despite the higher ransom. Examining the thresholds $\tilde{v}_r$ and $\tilde{v}_p$ in panel (a) of Figure 2, notably $\tilde{v}_p$ increases more sharply than $\tilde{v}_r$. Therefore, overall, the ransom paying population size $r$ is increasing in $\tau$ throughout this region. Turning our attention to the expected total ransom paid, as $\tau$ increases, $q(\cdot)$ decreases while $R^*(\cdot)$ and $r(\cdot)$ increase. The multiplicative effect of these latter two components dominate the decrease in risk, and hence the expected total ransom paid ultimately increases in $\tau$. We illustrate this net effect in panel (f) of Figure 2.

Interestingly, this region is characterized by a higher ransom, a larger unpatched population (as

[13] This is because $\hat{\theta} = \tilde{\theta}$ in equation (11) of Lemma 2 in this particular region.

well as a larger ransom-paying subset), and a higher expected total ransom being paid as $\tau$ grows. Nevertheless, the attacker population still decreases in aggregate. It is important to further explore why this is the case and examine how attackers are differentially impacted by this shift in entry costs, which we study in the following proposition.

**Proposition 3** *When the attacker entry cost parameter $\tau$ is at the higher end of the focal region, under both pre-crypto benchmark and post-crypto ransomware settings, an increase in attacker entry costs benefits some entering attackers while hurting others. Specifically, among attackers who enter in equilibrium, higher skilled attacker utilities increase in $\tau$, whereas lower skilled attacker utilities decrease in $\tau$.*

*Technically, under the post-crypto ransomware setting, given $\tau \in (\hat{\tau}, \bar{\tau})$, there exists $\theta' \in (\hat{\theta}(\tau), 1)$ such that $\frac{dU^a_{RW}(\theta, \tilde{\Phi}^*)}{d\tau} < 0$ for all $\theta \in (\hat{\theta}(\tau), \theta')$ while $\frac{dU^a_{RW}(\theta, \tilde{\Phi}^*)}{d\tau} > 0$ for all $\theta \in (\theta', 1]$, where $\hat{\theta}(\tau)$ represents the type of the marginal attacker in the market as defined in Lemma 2. The statement under the pre-crypto benchmark setting is similar.*

For the post-crypto ransomware setting, as we already established above, in this region we have increasing equilibrium ransoms, increasing unpatched segment and as a subset, increasing ransom paying population, together with an increasing expected total ransom paid in association with an increase in attacker entry costs. As all these metrics increase, it follows from (9) that attacker revenue (the first part of the utility equation) is also increasing in $\tau$ for all remaining attackers in the market. Moreover, as previously discussed, attacker revenue is homogeneous across attackers (with heterogeneity manifested only at the entry cost level). Attackers with highest skills ($\theta$ close to 1) will incur small entry costs which are offset entirely by the increase in revenue, resulting in such attackers being better off as attacker entry costs increase and additional less-skilled attackers drop of the market, lowering competition. On the other hand, for attackers still in the market but with lower skills, the increase in entry costs takes a bigger toll on the overall utility, erasing the

increase in revenue. This results in less-skilled attackers that are still in the market being negatively impacted by an increase in entry costs (with some attackers gradually peeling off). To sum up, while the overall attacker population declines in the attacker entry costs in this region, the impact is in fact positive for some attackers and negative for others.

For the pre-crypto benchmark setting, the argument is similar but simpler in that the revenue portion of the attacker utility function in (4) is more straightforward. Specifically, by part (i) of Proposition 1, $a$ decreases in $\tau$, thus $q(a)/a$ increases in $\tau$. Moreover, the unpatched population increases due to the reduced risk from attackers. Thus, the revenue portion of the utility function increases in $\tau$ for all attackers who remain in the market. Similar to the ransomware scenario, the heterogeneity in entry cost leads to the higher skilled attacker ultimately benefitting from the higher $\tau$ and the lower skilled attackers being worse off.

# 5 Welfare Comparative Analysis

In this section, we conduct welfare sensitivity analysis and welfare comparisons across scenarios with respect to pertinent model parameters. Since cybercriminals engage in actions prohibited by criminal legislation and, moreover, they often operate outside the sovereign jurisdiction covering their victims, we exclude their criminal gains in our welfare analysis consistent with extant literature (see Lewin and Trumbull 1990 for an elaborate discussion on this topic). Therefore, welfare analysis in this context centers on consumer surplus. We use $W_{BM}$ and $W_{RW}$ to denote the equilibrium welfare metric under pre-crypto benchmark and post-crypto ransomware settings.

## 5.1 Impact of Market Primitives of Attackers ($\tau$ and $\tilde{\rho}$)

In this subsection, we discuss how attacker entry costs and gains from traditional attacks impact welfare.

**Proposition 4** *In the focal region, the sensitivity of welfare to attacker entry costs is structurally different under pre-crypto benchmark and post-crypto ransomware settings. Under the pre-crypto benchmark setting, welfare always increases as attacker entry costs rise. However, under the post-crypto ransomware setting, welfare actually decreases in attacker entry costs when the latter are within an intermediate range. Outside of that range, the behavior is similar to that under the pre-crypto benchmark setting.*

The essence of Proposition 4 is illustrated in Figure 3. At first thought, one might expect social welfare to (weakly) increase in $\tau$ as increased efforts lead to less attackers, and, hence, lowered risk of a breach. Our result confirms this expectation in the benchmark scenario. However, as it turns out, in the post-crypto scenario, social welfare can actually decrease in $\tau$. By part (ii) of Proposition 1, the size of the equilibrium attacker population and corresponding security risk increase when $\tau$ is intermediate. As discussed in Section 4 and illustrated in Figure 2, the effects are complex as this increase in attacker population occurs in tandem with the unpatched consumer population also unexpectedly increasing more steeply in $\tau$ in this subregion. This can be seen in panel (a) of Figure 2. This behavior, in turn, leads to a steeply increasing ransom being charged by attackers to opportunistically leverage the higher valuation consumers who are electing to go unpatched which can be seen in panel (e) of Figure 2.

Unpacking the aggregate equilibrium movement in these metrics, the impact of a shift in $\tau$ on individual consumer decisions varies considerably depending on the type of consumer and their strategic security choices. For example, if a consumer has a high valuation and finds it optimal to patch in equilibrium, then an increase in $\tau$ has no effect on the surplus of the consumer provided that they continue to patch. If a consumer prefers to patch but switches to being unpatched and paying ransom due to an increase in $\tau$, then their surplus improves, by revealed preference, at the point where they switch. Next, for a consumer who does not change their strategy and still

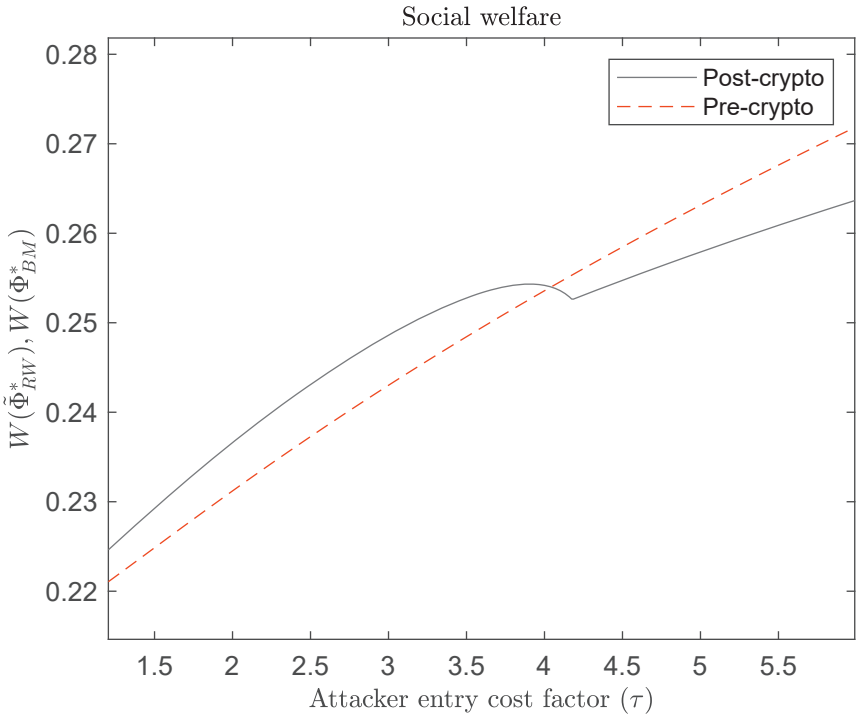


Figure 3: The impact of the attacker entry cost factor ($\tau$) on social welfare outcomes under pre-crypto benchmark and post-crypto ransomware settings. The parameter values are: $c_p = 0.3$, $\delta = 0.02$, $\kappa = 0.05$, $\alpha = 2.8$, $\rho = 1.6$, $\tilde{\rho} = 1.3$, $\omega = 0.1$, and $q(a) = 0.9a - 0.4a^2$.

prefers to be unpatched and pay ransom as $\tau$ increases, the analysis is more complex. Consumers' surpluses can either increase or decrease in $\tau$ depending on their type and the rate at which attackers transition from standard attack to ransomware attack strategies. Next, for a consumer who pays ransom and then switches to not paying due to an increase in $\tau$ and the associated increase in ransom, their surplus decreases in that the size of the attacker population increases in $\tau$. The same situation unfolds for consumers who already preferred to be unpatched and not pay ransom even under the lower $\tau$. Finally, some unpatched consumers who were not paying ransom are forced out of the market due to the increased risk which negatively affects consumer surplus. Taking all of these disparate impacts into account, Proposition 4 formally establishes that the overall welfare ultimately declines as the attacker entry costs increase in this subregion, largely due to the stark increase in risk facing lower valuation consumers who remain in the market and the loss of

consumers from the market.

Outside of the intermediate range of $\tau$, the behavior in the post-crypto ransomware setting is similar in nature to that under the pre-crypto benchmark setting. As can be seen in panel (d) of Figure 2, the decrease in the attacker population in equilibrium is much steeper in the lower and upper $\tau$ regions in comparison to the intermediate $\tau$ region. As a result, the effect that stems from a decrease in security risk associated with the declining attacker population dominates the overall effect on welfare. Therefore, welfare increases in $\tau$ in the outer regions, consistent with Proposition 4 and Figure 3.

Next, we continue our exploration of the potential benefits of ransomware attacks on society by directly comparing welfare outcomes in the post-crypto ransomware setting to the pre-crypto benchmark setting. An important question is whether the addition of ransomware to the threat landscape, despite giving strategic attackers multiple ways to monetize victims, can benefit society. In the following proposition, we formally examine this question as moderated by the returns to post-crypto standard attacks and varying levels of attacker entry costs.

**Proposition 5** *The societal impact of the presence of ransomware hinges on the returns to standard attacks under the post-crypto ransomware setting ($\tilde{\rho}$), the level of attacker entry costs ($\tau$), and ransomware-specific cost differential ($\omega$). Specifically, when $\omega$ is not too high:*

*(i) When $\tilde{\rho}$ is low, $W_{RW} > W_{BM}$ holds under both low and high $\tau$;*

*(ii) When $\tilde{\rho}$ is moderate, $W_{RW} > W_{BM}$ continues to hold under low $\tau$. However, $W_{RW}$ and $W_{BM}$ eventually cross, with $W_{BM} > W_{RW}$ obtaining under high $\tau$;*

*(iii) When $\tilde{\rho}$ is high, $W_{BM} > W_{RW}$ holds under both low and high $\tau$.*

First, we examine the most complex case which is part (ii) of Proposition 5. In this sub-case, $\omega$ is low such that, post-crypto, attackers can enter with ransomware or standard attacks with a

fairly similar cost structure. Moreover, the returns to standard attacks in the post-crypto setting, $\tilde{\rho}$, are at an intermediate range such that the benefits to attackers under the two settings (pre- and post-crypto) are more comparable. As attacker entry costs decrease toward the lower end of the focal region, the equilibrium outcome under the post-crypto ransomware setting is converging to only having standard attacks. This is because the risk is too high due to the size of the attacker population that enters, which drives higher valuation consumers to patch and leaves little ability for attackers to charge ransoms high enough to justify entering with ransomware. As $\tau$ increases from this boundary into an intermediate range, high skill attackers now find it optimal to enter with ransomware with a ransom in excess of $\tilde{\rho}$ such that some consumers prefer to reduce costs by paying ransom instead of patching. In that the highest skilled attackers have minimally greater costs from entering with ransomware in comparison to entering with standard attacks, these attackers begin to enter with ransomware instead. As patched consumers shift over to being unpatched and pay ransoms, the unpatched population size increases relatively faster in the post-crypto ransomware setting compared to the benchmark. As a result, the attacker population also increases relatively faster due to the greater incentive. This leads to decreasing consumers under $RW$ setting due to consumers with lower valuations exiting at the bottom of the market due to the increased risk associated with a greater attacker population. However, this negative impact on surplus is less than the positive impact associated with higher valuation consumers reducing costs by paying ransom instead of patching. Therefore, in aggregate, welfare is higher under $RW$ than under $BM$ when $\tau$ is low.

As the attacker entry cost moves to a high level, the equilibrium outcomes in the post-crypto ransomware and pre-crypto benchmark settings diverge. As discussed earlier, as $\tau$ gets high, attackers shift greatly toward ransomware strategies and away from standard attacks. As a result, the attacker population can be considerably larger under post-crypto ransomware setting than under

benchmark. Comparatively, substantially fewer consumers will elect to use the software under $RW$ setting than under $BM$ setting due to the greater risk associated with this attacker population. Notably, the consumers who remain unpatched and do not pay ransom are also exposed to much greater risk, thereby incurring larger losses that negatively impact surplus. Altogether, under high attacker entry costs, welfare under post-crypto ransomware setting is much lower than under the benchmark setting due to these effects. In Figure B.1 in Appendix B, we numerically illustrate the cutoff value for $\tau$ that dictates whether welfare is higher under the post-crypto setting. We explore how the cutoff is impacted by patching cost, security losses, and the attacker gains pre-crypto.

Turning part (iii) of Proposition 5, as $\tilde{\rho}$ becomes larger, it further encourages the entry of attackers under the post-crypto ransomware setting due to higher returns even with standard attacks. This significantly shifts the trade-offs discussed above in the direction of the latter effects described where welfare is worse off under $RW$ setting, ultimately dominating for both low and high $\tau$. For part (i) of Proposition 5, the effects shift in the opposing direction in that a low level of $\tilde{\rho}$ will ultimately shift more attackers toward entering with ransomware instead of standard attacks under the post-crypto setting. Moreover, post-crypto, attackers have a more difficult time justifying entering with a standard attack under the lower $\tilde{\rho}$. Therefore, some of these attackers with lower types will prefer to no longer enter the market. As a result, post-crypto, welfare can benefit both from fewer attackers in the market and from consumers who have more options to mitigate loss by paying ransom due to attackers who shifted attack modes.

A similar finding has been established in the context of an exogenous ransomware threat in August et al. (2022). Proposition 5 confirms the nuanced nature of this debate in the context of endogenous attackers. *Depending on the attacker primitives of the security landscape, the post-crypto ransomware setting may be more socially desirable than the pre-crypto benchmark setting from a security standpoint even in the presence of strategic attackers who are endogenously moti-*

*vated by ransoms being paid and can choose an attack mode that maximizes their own gains.*

## 5.2 Impact of Market Primitives for Consumers ($\alpha$ and $\delta$)

Next, we discuss the impact of the security loss factor versus the residual loss factor on welfare. Notably, security losses as dictated by parameter $\alpha$ exist under both pre-crypto benchmark and post-crypto ransomware settings. However, the residual loss parameter $\delta$ is only relevant under *RW* since it accounts for the *additional* losses sustained by consumers even though they paid the ransom that was demanded.

**Proposition 6** *In the focal region, when $\tau$ is not too low, welfare decreases in $\alpha$ under both pre-crypto benchmark and post-crypto ransomware settings. However, welfare increases in $\delta$ under the post-crypto ransomware setting.*

The benchmark scenario can be reasoned as follows. As $\alpha$ increases, there is a pressure on consumers at both ends of the market. Lower valuation consumers exit due to the increased risk, and higher valuation consumers shift toward patching. These two effects make the market less attractive to attackers, thus the attacker population also decreases. This slightly reduces risk for consumers, but the small decrease in attackers does not compensate for the overall higher losses associated with $\alpha$. The net effect is a decrease in welfare.

The post-crypto scenario is more complex. In the higher range of $\alpha$, attackers enter predominantly with ransomware attacks. In that range, the losses are so high that consumer demand shrinks, with exits at the bottom of the unpatched population (among consumers who were not willing to pay ransom). However, the difference between standard and residual losses ($(1-\delta)\alpha v$) also increases for remaining unpatched consumers, with increased pressure to patch for those with higher valuations. At the same time, attackers take advantage of an opportunity to increase ransom (given that paying ransom can mitigate a higher loss). Altogether, the compounded effect of

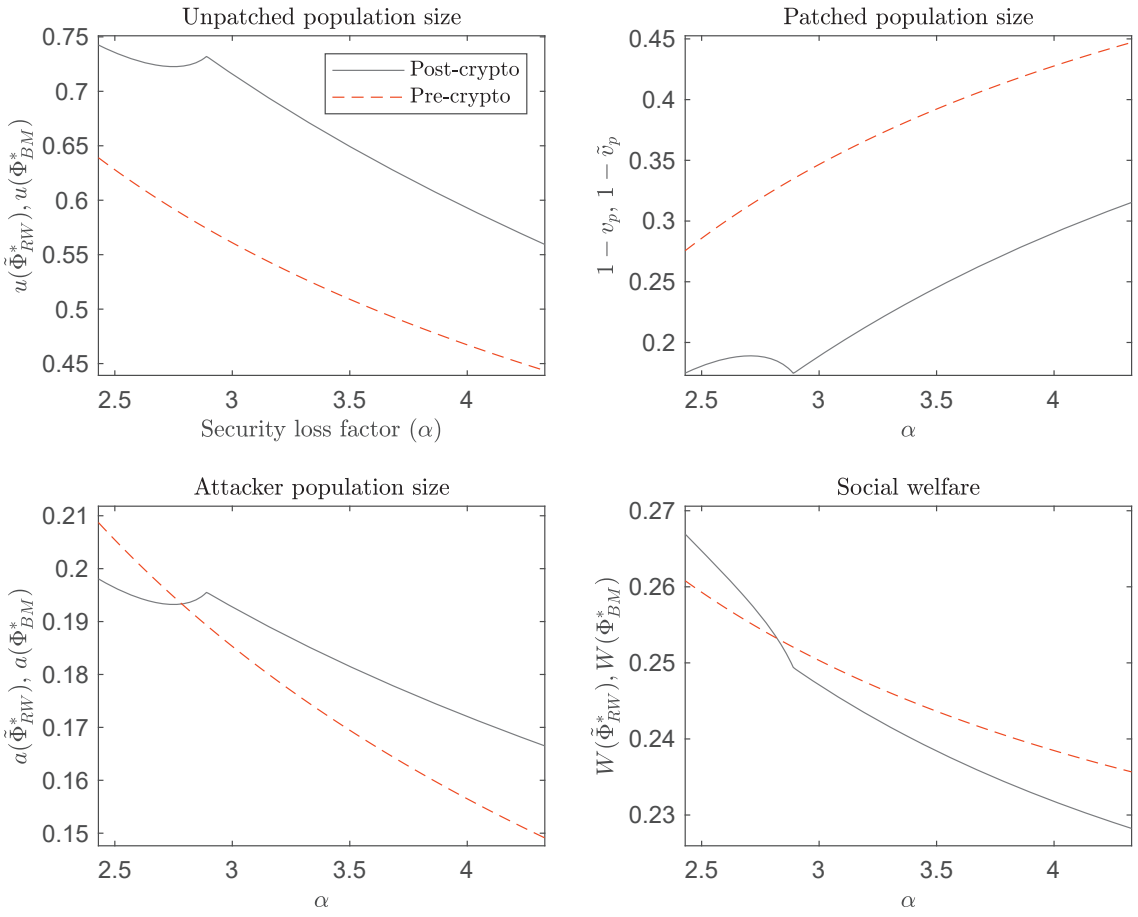


Figure 4: The impact of the security loss factor ($\alpha$) on equilibrium outcomes under pre-crypto benchmark and post-crypto ransomware settings. The parameter values are: $c_p = 0.3$, $\delta = 0.02$, $\kappa = 0.05$, $\tau = 4$, $\rho = 1.6$, $\tilde{\rho} = 1.3$, $\omega = 0.1$, and $q(a) = 0.9a - 0.4a^2$.

higher losses and higher ransoms drives an increase in patching, depressing the unpatched segment. As a result, there is a decrease in the attacker population; similar to the benchmark setting, that effect is not sufficient to counter the increase in losses on the consumer side, leading to an overall decrease in welfare under post-crypto ransomware setting too. For low $\alpha$ we see similar outcomes - post-crypto, in that range, most attackers enter with standard attacks. As such, the increase in security losses has a strong impact on consumer welfare as there is a very limited chance to be attacked by ransomware and be able to mitigate such losses.

However, for $\alpha$ in the intermediate range of the focal region, the dynamics at play are more com-

plex, similar to those captured under Propositions 1 and 4 with respect to changes in attacker entry costs ($\tau$). Specifically, attackers will gradually shift from standard attacks towards ransomware attacks as $\alpha$ increases as consumer willingness to pay ransom increases. As the chance to be able to mitigate risk via paying ransom increases at an accelerated rate (i.e., victims are increasingly more likely to be hit with a ransomware rather than a standard attack), we see that both the attacker and the unpatched consumer populations actually increase in $\alpha$, which represents a structural departure in sensitivity of such metrics compared to the pre-crypto scenario (as seen in panels (a) and (c) of Figure 4). While fewer consumers patch at the top of the valuation spectrum, at the bottom consumers still exit the market; hence overall we have market shrinkage. For similar reasons as in the higher $\alpha$ region, attackers still find it optimal to jack up the ransom demanded. In this intermediate $\alpha$ region, while direct losses to consumers are more moderate, the added risk from the increase in attackers leads to a compounded negative effect of higher losses on welfare post-crypto. At first it might seem that the post-crypto welfare impacts of a shift in $\tau$ in the intermediate range and a shift in $\alpha$ in intermediate range are different in nature as only in the former case the welfare changes monotonicity. However, that is not necessarily the case - in both instances (moving the attacker entry costs or the security losses) the momentary incentives for more attackers to join add even more stress to consumer welfare. As seen in Proposition 4 and Figure 3, that stress was enough to induce a decline in post-welfare in intermediate $\tau$, whereas in other regions an increase in $\tau$ benefits consumers. With sensitivity in $\alpha$, post-crypto, for large and small losses regimes, consumers suffer if losses further increase and this trend is actually *amplified* in the intermediate $\alpha$ range (as see in panel (d) of Figure 4, through a very fast decline in post-crypto welfare).

Now focusing solely on the post-crypto scenario, Proposition 6 reveals an interesting impact of residual losses on welfare. While an increase in security losses (through $\alpha$) has a negative impact on welfare, an isolated increase in residual losses after paying ransom (through $\delta$), surprisingly,

has a positive impact on welfare. As consumers are less able to mitigate losses through paying ransom, fewer will end up choosing that route. That lowers the ability of attackers to dip into the more lucrative revenue stream (via extortion), which, in turn, leads to a lot of attackers exiting the market. As such, the overall security risk in the market drops, which in turn benefits all unpatched consumers and even leads to an expansion of the market through entry at the low end of the valuation spectrum (as those consumers would not pay ransom and are unaffected by residual losses but they do benefit from an overall more secure market). For this reason, opportunities for very targeted policy interventions that shift $\delta$ higher and increase consumer burden when paying ransom may hold promise, as discussed in Section 6.

We wrap up this this welfare sensitivity analysis by emphasizing that, post-crypto, shifts in $\tau$, $\alpha$, and $\delta$ induce different correlations between movements in welfare and movements in attacker population, ransom demands, and expected total ransom paid. Shifts in $\tau$ in general move welfare and attacker population in opposite directions (e.g., achieving higher welfare comes with the bonus of fewer attackers), but higher welfare is associated with higher ransom amounts and total expected ransoms paid. Moreover, in the intermediate range of $\tau$, higher attacker entry costs can actually shift welfare lower. On the other hand, shifts in $\alpha$, for the most part, move welfare and attacker population in the same direction, which may seem to pose a dilemma depending on which metric lawmakers try to impact (e.g., higher welfare is associated with more attackers in the market). However, there is an intermediate region where increasing $\alpha$ leads to greater attacker entry but only serves to sharply decrease welfare due to the accelerated shift toward ransomware attacks away from standard attacks. Last, shifting $\delta$ seems to align the effects in a way that may more easily justify policy action. Specifically, increases in welfare are correlated with decreases in attacker population, ransom demanded, and expected total ransom paid.

# 6 Potential Policy Implications

In our study, we focused on exploring the impact of market conditions on consumer and attacker behavior, and how, in turn, that affects security and welfare metrics. In this section, we comment on potential policy implications by connecting particular policies to the specific market parameters they can impact. The aim of our economic model is to highlight where policy may be fruitful by isolating specific parameters and shifting them to determine the aggregate impact of the shift on welfare. As policy makers can use multiple levers that shift multiple parameters concomitantly, the effects we demonstrate as theoretically possible in isolation can help to inform subsequent research that can substantiate the magnitude of these effects (in separate or combined form) and account for other policy-related concerns and roll-out costs which altogether can lead to sharp policy recommendations.

## 6.1 Ransomware Legislative Context

While some more digitally savvy governments such as the U.S. government are well-versed at crypto tracing (House 2021), disruption of a currency explicitly designed to elude governmental control is a challenging task requiring extensive coordination and diplomacy (Geller 2021). Some proposed or enacted regulations require increased transparency around the operation of crypto exchanges and crypto operations in general. While cryptocurrency exchanges in the U.S. are required to implement *anti-money laundering* (AML) measures including reporting of large transactions and *know-your-customer* (KYC) data collection, recently proposed bipartisan legislation in the U.S., called the Digital Asset Anti-Money Laundering (DAAML) Act, advocates, among other things, for the extension of such regulations to the space of decentralized finance (DeFi) protocols and decentralized autonomous organizations (DAOs), effectively closing a loophole and bringing under AML oversight a significant and until-now largely unregulated part of the crypto industry (Lemire

2022, Lowe et al. 2023). In the European Union, updated AML (including KYC) standards around crypto transactions and the entities facilitating those will come into effect under the Markets in Crypto Assets (MiCA) regulation at the end of 2024, replacing the dated AMLD5 initiatives (Moody's 2022). In tackling the ransomware boom, international consensus is emerging that "consistent international scrutiny of cryptocurrencies will be key" (Uberti 2021), with many territories following suit (Thompson Reuters 2022). Also some general security policies require reporting of ransom payments. Such legislation will help flag faster to a greater extent transactions related to ransomware attacks. A much brighter spotlight will be shined on *both* victims and ransomware cybercriminals.

Enhanced focus of law enforcement on ransomware operators and facilitators is also palpable. After the U.S. Department of Justice recovered $2.3 million in cryptocurrency following the DarkSide's extortion of Colonial Pipeline, Deputy Attorney General Lisa Monaco commented, "Ransom payments are the fuel that propels the digital extortion engine, and today's announcement demonstrates that the United States will use all available tools to make these attacks more costly and less profitable for criminal enterprises. We will continue to target the entire ransomware ecosystem to disrupt and deter these attacks" (U.S. DOJ 2021a). In addition to going directly after the ransomware cybercriminals, broader targeting of ecosystem facilitators is illustrated by the recent sanctions issued by United States Department of the Treasury's Office of Foreign Assets Control against cryptocurrency exchanges and such as OTC Suex, Chatex, Garantex and Binance, as well as darknet market portals such as Hydra (Gkritsi 2021, US Department of Treasury 2022, 2023). Utilizing a more coordinated approach to combat ransomware, in October 2021 the U.S. Department of Justice created the National Cryptocurrency Enforcement Team (NCET) "to tackle complex investigations and prosecutions of criminal misuses of cryptocurrency, particularly crimes committed by virtual currency exchanges, mixing and tumbling services, and money laundering

infrastructure actors (U.S. DOJ 2021b).

### 6.2 Pre-crypto versus Post-crypto Contexts

Other government and law enforcement measures target the ransomware cyberthreat explicitly. In the U.S., while some radical measures such as a complete ban on ransomware payments have been floated and even partially implemented in a couple of states (Brumfield 2021, Ikeda 2022), such initiatives have been deemed infeasible, impractical, or suboptimal to fully implement at a federal level and hence momentarily abandoned (CNN 2021, Segal 2021, Wheeler and Martin 2021, Kapko 2022). However, suppose we entertain for a moment the possibility that such restrictive policies were feasible and that lawmakers were able to implement measures that effectively and completely disincentivize such attacks. An important question in such a case would be the following: *is it socially optimal to get rid of the ransomware security threat altogether*? It turns out that the answer is complex. In Proposition 5, we formally establish that in a context with strategic attackers making entry and attack mode decisions and strategic consumers making security decisions, social welfare can still be higher in broad regions under the post-crypto scenario with the presence of ransomware.

However, recognizing the potential infeasibility of the most drastic approach, legislators and law enforcement are taking a more pragmatic approach by focusing on disrupting the processes and requirements around such transactions. In other words, *in the near future, instead of completely eliminating the ransomware threat altogether, the policy makers aim to strategically contain it*, which we next discuss.

### 6.3 Attacker Entry Costs ($\tau$ and $\omega$)

It is not a stretch to envision today's world more in the range of moderate rather than extreme (low or high) attack costs, consistent with our definition of the focal region. Realistically, policies to dial up or down the attack costs have only so much bite to them (i.e., can move $\tau$ only a limited amount), due in many cases to policies operating within territorial jurisdictions whereas attackers operate across physical borders (and, even then, as discussed in Proposition 4, policymakers should be aware of the directional impact as they dial the costs).

One way to increase attacker entry costs is to increase law enforcement resources, forge collaboration at both national and international level, encourage public-private partnerships, and initiate bounty programs to disrupt cybercrime and identify/apprehend operators (e.g., INTERPOL 2021, US Department of State 2021, White House 2021, National Cybersecurity Alliance 2022, NSA 2024). Such initiatives increase the risk for cyber criminals to perpetrate the attacks in general, which can be quantified in our models through an increase in $\tau$ (i.e., the expected costs of attack campaigns increase due to higher risk of being caught).

Also, more specific to the post-crypto setting, crypto oversight legislation and ransomware-targeted law enforcement activity directly affect the entry cost of attackers. Closer monitoring of the ecosystem of ransomware facilitators and operators, along with aforementioned updated AML (including KYC) requirements on crypto transactions, as well as legislation that requires the reporting of ransom payments, will flag much faster the activity of ransomware operators and will require cybercriminals to jump through more complex hoops to be able to both collect ransom payments in crypto wallets and also withdraw those funds. The ransomware-specific entry costs are captured in our model with the parameter $\omega$, and the crypto transactional costs can be captured with $s(\theta)$ introduced in a generalized model in the extension in Appendix C.1.

### 6.4 Consumer Security Losses ($\alpha$)

Consumer losses are affected when legislation is enacted that indirectly impacts the IT security threat landscape through disclosure requirements. In Mar 2022, President Biden signed into law in the U.S. the Cyber Incident Reporting for Critical Infrastructure Act (CIRCIA), which mandated reporting of certain categories of cyber incidents by certain entities of national strategic importance (CISA 2022). In addition, in the particular case of data breaches, a swath of federal and state laws mandate various degrees of reporting in the U.S. (National Conference of State Legislatures 2022). In the European Union, data breach reporting requirements are stipulated under the General Data Protection Regulation (Intersoft Consulting 2023) and more general cyber incident reporting requirements for the financial sector are covered by the Digital Operations Resilience Act, which will enter in effect in early 2025 (Morgan Lewis 2024). Such policies can increase the impact of the attack on victims through a trust/reputation cost. In our model, this can be captured through a shift in parameter $\alpha$ (whether the victim incurs the full loss regardless of the attack type, or the residual losses in case of electing to pay ransom).

### 6.5 Consumer Residual Losses under Ransomware ($\delta$)

AML (and KYC) legislation adds regulatory burden and increased reputation risk on the victim side explicitly in the case of crypto ransom payments. The recently proposed Ransomware and Financial Stability Act of 2024 would require further scrutiny and approval by law enforcement for ransomware payments in excess of $100K (US Congress 2024). Hence, in our framework, we can link the impact of such initiatives to shifts in $\delta$, the coefficient of residual losses (which encompass all additional costs incurred by the victims in the wake of making a ransom payment). More precisely, on the victims' side, we postulate that such polices will likely lead to an increase in $\delta$. Taken in isolation (ignoring for the moment the effect on attackers), such an effect can lead to

resistance to related regulation due to perceived added burden on the victims. At first glance, one might intuitively expect that, as residual losses increase, the ability of consumers to mitigate risk via the ransom payment option decreases, which, in turn, might lead to reduced welfare. However, Proposition 6, establishes there are regions under which welfare actually increases in $\delta$ due to the strategic reaction by attackers.

# 7 Concluding Remarks

To better understand the impact of cryptocurrencies on the cybersecurity landscape, we conduct a comparative analysis of cybersecurity metrics and social welfare prior to and after the adoption of cryptocurrency. To that end, our modeling contribution is the development of a series of connected software-use models in the presence of security externalities, where consumers and attackers endogenously determine this risk, specifically with a *continuum of heterogeneous attackers making entry, attack mode (ransomware or traditional attacks), and ransom decisions.* Leveraging these models, we explore how market parameters impact outcomes under both pre-crypto benchmark and post-crypto ransomware threat landscapes. We show theoretically that post-crypto scenarios that include ransomware threats can be more socially desirable than pre-crypto, conventional threat scenarios when it is not too costly for attackers to enter the market. Hence, the existence of crypto-fueled ransomware, in itself, is not necessarily bad for cybersecurity and welfare.

We further show that directional effects (such as increasing attacker entry cost increase welfare) that obtain under conventional attacks do not necessarily obtain under the ransomware scenario. For example, instead of containing the ransomware threat, increasing attacker costs sometimes increases attacker entry and is detrimental to welfare. Moreover, increasing attacker entry costs can have secondary effects including increased demanded ransom, larger ransom paying population segments, and higher expected total ransom paid. We also study in a similar way the impact of

security losses as well as residual losses on these metrics. Altogether, the study of these effects offers great value to policy makers as they consider cybersecurity and crypto policies to drive these market parameters toward more favorable outcomes for society.

Our work's focus is on establishing theoretical possibilities, with future empirical work needed to assess the actual prevalence and magnitude of the outcomes we characterize. Our results are intended as an initial exploration to shed light into how shifting directional forces in isolation (i.e., one market parameter at a time) can alter the dynamics of the cybersecurity landscape, especially in the post-crypto era. Thus, quantities are normalized and a certain level of abstraction is present to permit an in-depth analysis of the particular dynamics in play.

The derived findings serve to inform policy makers of the possible need to consider more complex policies (or packages of multiple policies in tandem) that apply cybersecurity interventions on multiple dimensions concomitantly. Such complex regulations may be able to achieve welfare benefits without some of the secondary aforementioned effects. This presents great opportunities for further exploratory research into how to regulate crypto and its impact on cybersecurity. Our initial exploration looks at welfare effects net of any implementation costs required to induce a directional shift in market primitives (i.e., costs associated with policy roll out). Armed with accurate estimates of these costs, future work can perform comprehensive analyses and provide policy recommendations aimed at inducing the effects we highlight.

Another potential direction of future research is to expand the model framework to examine the interaction between user decisions and the additional options and costs that exist in the cybersecurity ecosystem. In our paper, we focus on users' adoption costs and patching costs as they decide whether or not to use a system and patch it in the face of potential cybersecurity attacks and cognizant of the externality imposed by their patching behavior on attackers' entry incentives. More generally, users also have the option to purchase cyberinsurance which can interact with their

adoption and patching decisions and lead to riskier behaviors. There can also possibly be valuation-independent incident reporting costs mandated by the government when systems are breached that are not captured solely by an increase in our $\alpha$ parameter. We acknowledge that the focal costs captured by our model provide a lower bound to the user costs present in a broader cybersecurity context, and we believe our work can provide a foundation to explore these additional factors.

Our work identifies effects on cybersecurity metrics (such as ransom payments) as well as welfare. Sometimes they move in the same direction as market parameters shift and sometimes not. Because ransom payments can be directly observable on public blockchains yet welfare is more difficult to measure, an interesting extension on the policy front would be to examine scenarios in which planners rely on public blockchains to peek into certain ransomware activity (including volume and destination of transactions), utilizing readily observable signals to inform and manage post-crypto cybersecurity policy agenda. Other interesting avenues for future research include exploring how the subsidization of patching costs or the taxation of ransomware payments (in addition to other regulations such as AML-mandated reporting) can influence the comparison between pre-crypto and post-crypto attack metrics.

## References


Adam, S. (2024). Unpatched vulnerabilities: The most brutal ransomware attack vector. Sophos News. Apr. 3. `https://news.sophos.com/en-us/2024/04/03/unpatched-vulnerabilities-the-most-brutal-ransomware-attack-vector/` (Accessed on Dec 15, 2024).

Ahnert, T., M. Brolley, D. Cimon, and R. Riordan (2022). Cyber security and ransomware in financial markets. CEPR Discussion Paper No. DP17403. Available at SSRN. `https://ssrn.com/abstract=4144735`.

Arctic Wolf Labs (2022). A Log4Shell (Log4j) retrospective. Dec. 5 `https://arcticwolf.com/resources/blog/log4j-retrospective/`.

August, T., D. Dao, and K. Kim (2019). Market segmentation and software security: Pricing patching rights. *Management Science 65*(10), 4451–4949.

August, T., D. Dao, and M. F. Niculescu (2022). Economics of ransomware: Risk interdependence and large-scale attacks. *Management Science 68*(12), 8979–9002.

August, T. and T. I. Tunca (2006). Network software security and user incentives. *Management Science 52*(11), 1703–1720.

August, T. and T. I. Tunca (2011). Who should be responsible for software security? A comparative analysis of liability policies in network environments. *Management Science 57*(5), 934–959.

Badea, L. and M. C. Mungiu-Pupazan (2021). The economic and environmental impact of Bitcoin. *IEEE Access 9*, 48091–48104.

Balasubramanian, A. (2021). Insurance against ransomware. Working paper, `https://ssrn.com/abstract=3846111`.

Bariviera, A. F. and I. Merediz-Solà (2021). Where do we stand in cryptocurrencies economic research? A survey based on hybrid analysis. *Journal of Economic Surveys 35*(2), 377–407.

Black, D. B. (2022). Cryptocurrency fuels growth of crime. Mar. 11. `https://www.forbes.com/sites/davidblack/2022/03/11/cryptocurrency-fuels-explosive-growth-of-crime/`.

Blosil, J. (2022). Measuring the true cost of a ransomware attack. Oct. 22. NetApp. `https://www.netapp.com/blog/ransomware-cost/`.

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

Brumfield, C. (2021, June). Four states propose laws to ban ransomware payments. CSO.

Cartwright, A. and E. Cartwright (2019). Ransomware and reputation. *Games 10*(2), 26.

Cartwright, A. and E. Cartwright (2023, Feb). The economics of ransomware attacks on integrated supply chain networks. *Digital Threats*.

Cartwright, E., J. Hernandez Castro, and A. Cartwright (2019). To pay or not: game theoretic models of ransomware. *Journal of Cybersecurity 5*(1), 1–12.

Cavusoglu, H., H. Cavusoglu, and S. Raghunathan (2007). Efficiency of vulnerability disclosure mechanisms to disseminate vulnerability knowledge. *IEEE Transactions on Software Engineering 33*(3), 171–185.

Chainalysis (2024). Ransomware payments exceed $1 billion in 2023, hitting record high after 2022 decline. Feb 7, `https://www.chainalysis.com/blog/ransomware-2024/` (Accessed on Apr 20, 2024).

Choi, J. P., C. Fershtman, and N. Gandal (2010). Network security: Vulnerabilities and disclosure policy. *Journal of Industrial Economics 58*(4), 868–894.

CISA (2022). Cyber Incident Reporting for Critical Infrastructure Act of 2022 (CIRCIA). Cybersecurity and Infrastructure Security Agency. `https://www.cisa.gov/topics/cyber-threats-and-advisories/information-sharing/cyber-incident-reporting-critical-infrastructure-act-2022-circia`.

CNN (2021). FBI tells Congress ransomware payments shouldn't be banned. Jul. 27.`https://www.cnn.com/2021/07/27/politics/senate-judiciary-ransomware-hearing/index.html`.

Crowdstrike (2022). History of ransomware. Oct. 10. `https://www.crowdstrike.com/cybersecurity-101/ransomware/history-of-ransomware/`.

Dey, D. and A. Lahiri (2021). Should we outlaw ransomware payments? In *Proceedings of the 54th Hawaii International Conference on System Sciences*, pp. 6609.

Dey, D., A. Lahiri, and G. Zhang (2015). Optimal policies for security patch management. *Journal on Computing 27*(3), 462–477.

DuBois, E. B., A. Peper, and L. A. Albert (2023). Interdicting attack plans with boundedly-rational players and multiple attackers: An adversarial risk analysis approach. *arXiv preprint arXiv:2302.01975*.

Fang, R., M. Xu, and P. Zhao (2022). Determination of ransomware payment based on bayesian game models. *Computers & Security 116*(May), 102685.

FinCEN (2022). Financial trend analysis: Ransomware trends in Bank Secrecy Act data between July 2021 and December 2021. US Treasury. Financial Crimes Enformcement Network. Report. Nov. 1. https://www.fincen.gov/sites/default/files/2022-11/Financial%20Trend%20Analysis_Ransomware%20FTA%202_508%20FINAL.pdf.

Foley, S., J. R. Karlsen, and T. J. Putnins (2019, 04). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies 32*(5), 1798–1853.

Gal-Or, E. and A. Ghose (2005). The economic incentives for sharing security information. *Information Systems Research 16*(2), 186–208.

Galinkin, E. (2021). Winning the ransomware lottery: A game-theoretic approach to preventing ransomware attacks. In *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Virtual Event, October 25–27, 2021, Proceedings 12*, pp. 195–207. Springer.

Garcia, E., A. Von Moll, D. W. Casbeer, and M. Pachter (2019). Strategies for defending a coastline against multiple attackers. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 7319–7324. IEEE.

Gates, D. (2018). Boeing hit by WannaCry virus, but says attack caused little damage. Seattle Times, Mar 28, https://www.seattletimes.com/business/boeing-aerospace/boeing-hit-by-wannacry-virus-fears-it-could-cripple-some-jet-production/ (Accessed on Jul 12, 2024).

Gatlan, S. (2023). IceFire ransomware now encrypts both Linux and Windows systems. BleepingComputer, Mar. 9. https://www.bleepingcomputer.com/news/security/icefire-ransomware-now-encrypts-both-linux-and-windows-systems/ (Accessed on Mar 9, 2023).

Geller, E. (2021, August). Global 'whack-a-mole': Why it's so hard for the U.S. to go after hackers' digital wallets. Politico.

GitHub (2024). The state of open source and rise of AI in 2023. Jul. https://github.blog/news-insights/research/the-state-of-open-source-and-ai/.

Gkritsi, E. (2021, November). Biden administration sanctions crypto exchange chatex over ransomware allegations. Coindesk. Available at https://www.coindesk.com/policy/2021/11/09/biden-administration-sanctions-crypto-exchange-chatex-over-ransomware-allegations/.

Halaburda, H., G. Haeringer, J. S. Gans, and N. Gandal (2020). The microeconomics of cryptocurrencies (no. w27477). Technical report, National Bureau of Economic Research (No. w27477).

Hausken, K. and V. M. Bier (2011). Defending against multiple different attackers. *European Journal of Operational Research 211*(2), 370–384.

Hernandez-Castro, J., A. Cartwright, and E. Cartwright (2020). An economic analysis of ransomware and its welfare consequences. *Royal Society Open Science 7*(3), 190023.

House, B. (2021, July). U.S. Plans to Counter Ransomware Attacks Through Crypto Tracing. Bloomberg.

IBM (2023). IBM report: Ransomware persisted despite improved detection in 2022. Feb. 22. https://newsroom.ibm.com/2023-02-22-IBM-Report-Ransomware-Persisted-Despite-Improved-Detection-in-2022.

IBM Security (2022). Cost of a data breach report 2022. IBM Study, https://www.ibm.com/downloads/cas/3R8N1DZJ.

Ikeda, S. (2022). Patchwork of US state regulations becomes more complex as Florida, North Carolina ban ransomware payments. CPO Magazine. https://www.cpomagazine.com/cyber-security/patchwork-of-us-state-regulations-becomes-more-complex-as-florida-north-carolina-ban-ransomware-payments/.

INTERPOL (2021). INTERPOL cybercrime capacity building project in the Americas, phase ii. Jan. https://www.interpol.int/en/Crimes/Cybercrime/Cyber-capabilities-development/Cybercrime-Capacity-Building-in-the-Americas.

Intersoft Consulting (2023). Art. 33 GDPR: Notification of a personal data breach to the supervisory authority. https://gdpr-info.eu/art-33-gdpr/.

Ioannidis, C., D. Pym, and J. Williams (2012). Information security trade-offs and optimal patching policies. *European Journal of Operational Research 216*(2), 434–444.

Jeffery, L. and V. Ramachandran (2021, July). Why ransomware attacks are on the rise — and what can be done to stop them. PBS News Hour, Nation.

Kapko, M. (2022). US government rejects ransom payment ban to spur disclosure. Cybersecurity Dive. Sep. 19. https://www.cybersecuritydive.com/news/government-ransomware-guidance/632136/.

Kaseya (2020). The 2019 Kaseya state of IT operations report for small and mid-sized businesses. Jan. https://www.kaseya.com/wp-content/uploads/dlm_uploads/2020/05/Kaseya-Whitepaper-2019-IT-Operations-Survey-Report.pdf (Accessed on Dec 18, 2024).

Lakshmanan, R. (2021, June). Wormable DarkRadiation Ransomware Targets Linux and Docker Instances. The Hacker News.

Laszka, A., S. Farhang, and J. Grossklags (2017). On the economics of ransomware. In *Conference on Decision and Game Theory for Security (GameSec)*, Vienna, Austria.

Lemire, K. A. (2022). Cryptocurrency and anti-money laundering enforcement. Reuters. Sep. 26. https://www.reuters.com/legal/transactional/cryptocurrency-anti-money-laundering-enforcement-2022-09-26/.

Levy, A. (2022). Why should you use crypto? The Motley Fool. Sep 20. https://www.fool.com/investing/stock-market/market-sectors/financials/cryptocurrency-stocks/benefits-of-cryptocurrency/.

Lewin, J. L. and W. N. Trumbull (1990). The social value of crime? *International Review of Law and Economics 10*(3), 271–284.

Li, X. and A. Whinston (2020). The economics of cyber crime. Working paper, https://ssrn.com/abstract=3603694.

Li, Z. and Q. Liao (2020). Ransomware 2.0: to sell, or not to sell - A game-theoretical model of data-selling ransomware. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*, pp. 1–9.

Lowe, M. S., E. G. Ostroff, and S. F. Rogers (2023). Bank Secrecy Act's crypto expansion is on the horizon. Law360, Feb. 7. https://www.law360.com/articles/1573438.

Maudrill, B. (2024). Sonatype reports 156Infosecurity Magazine, Oct. 11. https://www.infosecurity-magazine.com/news/156-increase-in-oss-malicious/.

Mitra, S. and S. Ransbotham (2015). Information disclosure and the diffusion of information security attacks. *Information Systems Research 26*(3), 565–584.

Moody's (2022). What does the proposed EU markets in Crypto-Assets Act (MiCA) mean for the industry? Dec. 16. https://kyc.moodys.io/content-highlights-section/what-does-proposed-eu-markets-crypto-assets-act-mica-mean-industry.

Morgan Lewis (2024). Preparing for DORA: ESAs publish incident reporting requirements. Aug. https://www.morganlewis.com/blogs/sourcingatmorganlewis/2024/08/preparing-for-dora-esas-publish-incident-reporting-requirements.

Murphy-Kelly, S. (2021). The bizarre story of the inventor of ransomware. CNN. May 16. https://www.cnn.com/2021/05/16/tech/ransomware-joseph-popp/index.html.

National Conference of State Legislatures (2022, Sep). Summary: 2022 security breach legislation. https://www.ncsl.org/technology-and-communication/2022-security-breach-legislation.

National Cybersecurity Alliance (2022). Cybersecurity collaboration as a national imperative. Oct. 6. https://staysafeonline.org/online-safety-privacy-basics/cybersecurity-collaboration-as-a-national-imperative/.

NSA (2024). NSA Cybersecurity Collaboration Center. Aug. https://www.nsa.gov/About/Cybersecurity-Collaboration-Center/.

Polaris Market Research (2024). Patch management market. Jan. https://www.polarismarketresearch.com/industry-analysis/patch-management-market (Accessed on Dec 18, 2024).

Red Hat (2021). The State of Enterprise Open Source. Red Hat.

Ryan, P., J. Fokker, S. Healy, and A. Amann (2022). Dynamics of targeted ransomware negotiation. *IEEE Access 10*, 32836–32844.

Schryen, G. and E. Rich (2010). Increasing software security through open source or closed source development? empirics suggest that we have asked the wrong question. In *2010 43rd Hawaii International Conference on System Sciences*, pp. 1–10. IEEE.

Segal, E. (2021, June). Banning Ransomware Payments Could Create New Crisis Situations. Forbes.

Selten, R. (1988). *Models of strategic rationality*, Chapter A simple game model of kidnapping, pp. 77–93. Theory and Decision Library C. Springer.

Spring, T. (2021, July). Linux Variant of REvil Ransomware Targets VMware's ESXi, NAS Devices. Fortune.

Statista (2024). Number of supply chain attacks on open source software (OSS) from 2019 to 2023. https://www.statista.com/statistics/1268934/worldwide-open-source-supply-chain-attacks/.

Swire, P. P. (2005). A theory of disclosure for security and competitive reasons: Open source, proprietary software, and government systems. *Hous. L. Rev. 42*, 1333.

Synopsis (2024). Open source security and risk analysis report. https://www.synopsys.com/content/dam/synopsys/sig-assets/reports/rep-ossra-2024.pdf.

Thompson Reuters (2022). Cryptocurrency regulations by country. https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2022/04/Cryptos-Report-Compendium-2022.pdf.

Tung, L. (2022). Ransomware: Hackers are using Log4j flaw as part of their attacks, warns Microsoft. ZDNET. Jan. 11 https://www.zdnet.com/article/ransomware-warning-hackers-are-using-log4j-flaw-as-part-of-their-attacks-warns-microsoft/.

Uberti, D. (2021, October). White House ransomware summit eyes tighter global scrutiny for crypto. The Wall Street Journal. Available at https://www.wsj.com/articles/white-house-ransomware-summit-eyes-tighter-global-scrutiny-for-crypto-11634227377.

UK (2016). National cyber security strategy 2016-2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf (Accessed on Dec 19, 2024).

US Congress (2024). H.R.7965 - Ransomware and Financial Stability Act of 2024. Apr. https://www.congress.gov/bill/118th-congress/house-bill/7965/all-info.

US Department of State (2021). Rewards for Justice – Reward offer for information on foreign malicious cyber activity against U.S. critical infrastructure. Media Note. Jul. 15.https://www.state.gov/rewards-for-justice-reward-offer-for-information-on-foreign-malicious-cyber-activity-against-u-s-critical-infrastructure/.

US Department of Treasury (2022). Treasury sanctions Russia-based Hydra, world's largest darknet market, and ransomware-enabling virtual currency exchange Garantex. Apr. https://home.treasury.gov/news/press-releases/jy0701.

US Department of Treasury (2023). U.S. treasury announces largest settlements in history with world's largest virtual currency exchange binance for violations of U.S. anti-money laundering and sanctions laws. Nov. https://home.treasury.gov/news/press-releases/jy1925.

U.S. DOJ (2021a, June). Department of Justice Seizes $2.3 Million in Cryptocurrency Paid to the Ransomware Extortionists Darkside. U.S. Department of Justice Press Release.

U.S. DOJ (2021b, October). Deputy Attorney General Lisa O. Monaco announces National Cryptocurrency Enforcement Team. U.S. Department of Justice Press Release. Available at https://www.justice.gov/opa/pr/deputy-attorney-general-lisa-o-monaco-announces-national-cryptocurrency-enforcement-team.

Vakilinia, I., M. M. Khalili, and M. Li (2021). A mechanism design approach to solve ransomware dilemmas. In *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Virtual Event, October 25–27, 2021, Proceedings 12*, pp. 181–194. Springer.

Vanian, J. (2021, July). Everything to know about REvil, the group behind a big ransomware spree. Fortune.

Vijayan, J. (2023). Majority of ransomware attacks last year exploited old bugs. Dark Reading. Feb. 20. https://www.darkreading.com/cyberattacks-data-breaches/dozens-of-vulns-in-ransomware-attacks-offer-adversaries-full-kill-chain/ (Accessed on Dec 15, 2024).

Wheeler, T. and C. Martin (2021, July). Should ransomware payments be banned? Brookings.

White House (2021, May). Executive order on improving the nation's cybersecurity. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/.

White House (2023). National cybersecurity strategy. Mar. https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf (Accessed on Dec 19, 2024).

Winder, D. (2021, November). New Ransomware Threat Jumps From Windows To Linux - What You Need To Know. Forbes.

Witten, B., C. Landwehr, and M. Caloyannides (2001). Does open source improve system security? *IEEE Software 18*(1), 57–61.

Xu, Z. and J. Zhuang (2019). A study on a sequential one-defender-n-attacker game. *Risk Analysis 39*(6), 1414–1432.

Yin, T., A. Sarabi, and M. Liu (2023). Deterrence, backup, or insurance: game-theoretic modeling of ransomware. *Games 14*(2), 20.

Yue, Y., X. Li, D. Zhang, and S. Wang (2021). How cryptocurrency affects economy? a network analysis using bibliometric methods. *International Review of Financial Analysis 77*, 101869.

Zahravi, A. (2021, June). Bash Ransomware DarkRadiation Targets Red Hat- and Debian-based Linux Distributions. Trend Micro.

Zhao, X., L. Xue, and A. B. Whinston (2013). Managing interdependent information security risks: Cyberinsurance, managed security services, and risk pooling arrangements. *Journal of Management Information Systems 30*(1), 123–152.

Zhao, Y., Y. Ge, and Q. Zhu (2021). Combating ransomware in internet of things: A games-in-games approach for cross-layer cyber defense and security investment. In *Decision and Game Theory for Security: 12th International Conference, GameSec 2021, Virtual Event, October 25–27, 2021, Proceedings*, pp. 208–228. Springer.

**Online Supplement for**
***"The Impact of Cryptocurrency on Cybersecurity"***

# Appendix A

## A.1 Focal Region Conditions

Define the functions $H(y) = \frac{q^{-1}(y)q'(q^{-1}(y))}{y}$ and $G(\omega) = \left(1+\omega+\sqrt{\omega(\omega+1)}\right) H^{-1}\left(\frac{1}{\omega+\sqrt{\omega(\omega+1)}+\frac{3}{2}}\right)$. Let $\hat{\omega} = G^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)$. The parameter conditions for the focal region are then given by:

(i) $0 < c_p < \frac{1}{2}$

(ii) $\tilde{\rho} > \frac{c_p}{q(1)}$

(iii) $\max\left(\tilde{\rho}, \frac{1}{q(1)}\right) < \alpha < \frac{\tilde{\rho}}{c_p}$

(iv) $\rho > \frac{2\alpha\tilde{\rho}q^{-1}\left(\frac{c_p}{\alpha}\right)^2}{(\alpha+\tilde{\rho})q^{-1}\left(\frac{2c_p}{\alpha+\tilde{\rho}}\right)^2}$

(v) $\max(\hat{\omega}, -1 + \frac{(\alpha+\tilde{\rho})\left(q^{-1}\left(\frac{c_p}{\alpha}\right)\right)^2}{2\rho\left(q^{-1}\left(\frac{2c_p}{\alpha+\tilde{\rho}}\right)\right)^2}) < \omega < \frac{(\alpha-\tilde{\rho})^2}{4\alpha\tilde{\rho}}$

(vi) $\frac{c_p\tilde{\rho}}{\alpha\left(q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)\right)^2} < \tau < \frac{c_p(\alpha+\tilde{\rho})}{2\alpha(1+\omega)\left(q^{-1}\left(\frac{2c_p}{\alpha+\tilde{\rho}}\right)\right)^2}$.

The essence of the conditions is that parameter values need to be within an intermediate range such that consumer segments do not disappear and not all attackers enter the market. These conditions are derived from conditions on making sure the consumer thresholds are ordered properly (i.e., ensuring $0 < \tilde{v}_n < \tilde{v}_r < \tilde{v}_p < 1$ (i.e., that no consumer has an incentive to switch strategies), $0 < a^* < 1$, and no attacker has any incentive to deviate to charging a different ransom, switching from conducting ransomware to only conducting standard attacks or vice versa. It is important to note that our focal region only serves to simplify the paper while still establishing the main insights; our insights can be shown to hold over broader regions and other market structures.

## A.2 Properties of $q(a)$

We denote the probability that a consumer is randomly hit by a security attack by $q(a(\Phi))$, where $a(\Phi)$ is endogenously determined in equilibrium. As a function of the attacker population size $a$, the function $q(a)$ is assumed to be increasing and concave, with $q(0) = 0$, $q(1) \leq 1$, $q'(0) < \infty$,

and the attacker elasticity of risk is decreasing, i.e., $\eta_q'(a) < 0$ where $\eta_q(a) = (dq/q)/(da/a)$ such that there is a diminishing marginal increase in risk as more attackers enter, both in absolute and relative terms. This elasticity assumption captures how the relative increase in risk diminishes as the attacker population size increases, which is expected if the marginal risk from additional attackers decreases. Examples of specific functional forms that satisfy this elasticity assumption are $q(a) = 1 - (1-\xi)^a$, $q(a) = \frac{\xi a}{a+1}$, $q(a) = \xi \log(1+a)$, $q(a) = C_1 \frac{1-e^{-C_2 a}}{1-e^{-C_2}}$, and $q(a) = C_1 a + C_2 a^2$ (where the applicable ranges for the parameters $\xi$, $C_1$, and $C_2$ depend on the specific functional form). For illustration purposes throughout the paper, we use $q(a) = C_1 a + C_2 a^2$, with $C_1 \in (0, 2)$ and $C_2 \in (-\frac{C_1}{2}, \min(0, 1 - C_1))$ to ensure $q(0) = 0$, $q(1) \leq 1$, and $q(a)$ is increasing, concave in $a \in (0, 1)$. Although we use this functional form for figures, the proofs are done with a general $q(a)$ (i.e., without assuming a specific $q(a)$) so that the propositions apply more generally for any $q(a)$ meeting the stated general conditions.

# Appendix B:
# Sensitivity Analysis

In this appendix, we examine the threshold in attacker entry costs ($\tau$) that separates whether pre-crypto or post-crypto welfare is superior. In particular, by part (ii) of Proposition 5, post-crypto welfare exceeds pre-crypto welfare for lower $\tau$, whereas pre-crypto welfare exceeds post-crypto welfare for higher $\tau$. We now study how the threshold in attacker entry costs, $\hat{\tau}$, varies in market parameters. There are two cases depending on whether attackers split in their attack modes in the post-crypto context. For the simpler case where they do not, $\hat{\tau}$ satisfies:

$$\begin{aligned}\hat{\tau} =& \frac{c_p\rho}{\alpha q^{-1}\left(\frac{c_p}{\rho(1+\omega)}\right)^2} + \frac{c_p\left(-\rho^2 q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right) - \frac{2c_p(\rho\omega+\rho-\tilde{\rho})}{(\omega+1)^2 q'\left(q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)\right)}\right)}{\rho(\alpha(-c_p)+\rho\omega+\rho)q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)^3} \times \kappa \\ &+ \frac{2c_p^2\left(\rho(\omega+1)(\tilde{\rho}-2\rho(\omega+1))^2 q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right) q'\left(q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)\right) + 2c_p(\rho\omega+\rho-\tilde{\rho})^2\right)}{\alpha\rho^3(\omega+1)^4 q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)^4 q'\left(q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)\right)^2} \times \delta \\ &+ O(\kappa+\delta)^2\end{aligned} \tag{B.1}$$

As can be seen, although the expression can be written, it is too complex to readily derive comparative statics. Therefore, we study both cases numerically and provide illustrations on how $\hat{\tau}$ varies in the market parameters. We utilize the same common parameter values used throughout the paper for consistency.

In panel (a) of Figure B.1, we examine how the threshold moves in patching costs. The threshold $\hat{\tau}$ decreases in $c_p$. However, note that the focal region bounds also adjust when patching costs change, and overall the portion of the space where post-crypto welfare is superior remains fairly consistent. We only vary the cost by plus or minus 10% for illustration purposes, but the spirit of what we describe expands further. In panel (b) of Figure B.1, we examine how the threshold moves in the gains to pre-crypto attacks while holding the post-crypto gains ($\tilde{\rho}$) constant. In this case, the threshold $\hat{\tau}$ increases in $\rho$. Moreover, higher $\rho$ leads to an expansion of post-crypto welfare superiority. Finally, in panel (c) of Figure B.1, we examine how the threshold moves in security losses. The threshold $\hat{\tau}$ decreases in $\alpha$. In this case, an increase in $\alpha$ leads to an expansion of pre-crypto welfare superiority and a simultaneous reduction in the region where post-crypto welfare is higher.

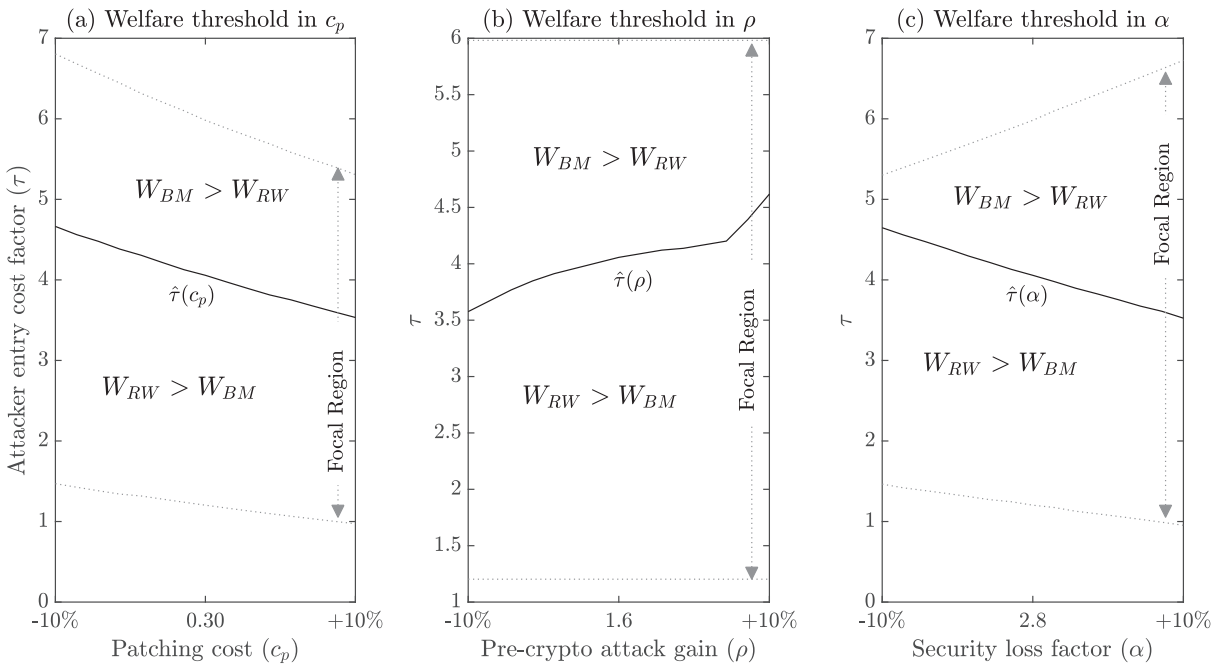


Figure B.1: The impact of patching costs ($c_p$), pre-crypto attack gain ($\rho$), and security loss factor ($\alpha$) on the threshold that designates when the post-crypto context is welfare superior. The baseline parameter values are: $c_p = 0.3$, $\delta = 0.02$, $\kappa = 0.05$, $\alpha = 2.8$, $\rho = 1.6$, $\tilde{\rho} = 1.3$, $\omega = 0.1$, and $q(a) = 0.9a - 0.4a^2$. Panels (a), (b), and (c) vary $c_p$, $\rho$, and $\alpha$, respectively, up to plus or minus 10 percent.

# Appendix C:
# Generalized Model and Proof of Main Results

In the main body of the paper, attacker heterogeneity is captured exclusively in terms of attacker entry cost. However, in a more general setup, one could conceive of multi-dimensional attacker heterogeneity, *both* at entry cost and revenue levels. As such, in this Appendix, we first present the *generalized model* (introduced below in Appendix C.1), with the model in the main body of the paper representing a special case of this general model. To that end, we derive the equilibrium market structure directly using the generalized model and then derive the main results in the paper under the simplified assumptions.

Moreover, we further show in Appendix D that the insights from the main model continue to hold under the generalized attacker heterogeneity model. To streamline the presentation, in the main body of the paper we focus on the simplified model, and discuss the analysis of the generalized model in the Appendix.

## C.1 Generalized Model of Attacker Heterogeneity

As mentioned above, we consider from the beginning a general model that incorporates attacker heterogeneity at both entry cost and revenue levels. More precisely, in this generalized version, the attacker revenues associated with both standard attacks and ransomware are type dependent.

Extending our pre-crypto benchmark model from Section 3.1, if an attacker of type $\theta$ hits a victim with a standard attack, then the attacker will derive $\rho\lambda(\theta)$ from the attack, where $\lambda(\cdot)$ is weakly increasing, differentiable, with $\lambda(\cdot)>0$ and $\lambda(1)=1$. Thus, for an attacker with type $\theta$, the expected utility is given by:

$$U^{a}_{BM}(\theta,\Phi) \triangleq \begin{cases} \frac{q(a(\Phi))\times\rho\lambda(\theta)\times u(\Phi)}{a(\Phi)} - \tau(1-\theta) & if \quad \gamma(\theta)=E\,; \\ 0 & if \quad \gamma(\theta)=NE\,. \end{cases} \tag{C.1}$$

Similarly, under the post-crypto ransomware model, an attacker derives $\tilde{\rho}\lambda(\theta)$. Moreover, if the victim pays the attacker's ransom $R$, then the attacker retains revenue $R\times s(\theta)$, where $s(\theta)$ is a weakly increasing, differentiable function of $\theta$ with $s(\cdot)>0$ and $s(1)=1$. Both of these functions ($\lambda$ and $s$) capture that *less-skilled* attackers (those with lower $\theta$) retain less when extracting revenue (through slippage from one's lower ability in the collection phase of ransom payments and having

a higher chance of being caught across both standard and ransomware attacks). Thus, for an attacker with type $\theta$, generalizing equation (9) from Section 3.2, the expected utility in the post-crypto context is given by:

$$U^{a}_{RW}(\theta, \tilde{\Phi}) \triangleq \begin{cases} \frac{q(a(\tilde{\Phi}))\times\left[\tilde{\rho}\lambda(\theta)\times(u(\tilde{\Phi})-r(R,\tilde{\Phi}))+Rs(\theta)\times r(R,\tilde{\Phi})\right]}{a(\tilde{\Phi})} - \tau(1-\theta)(1+\omega) & if \quad \tilde{\gamma}(\theta) = (ER, R)\,; \\ \frac{q(a(\tilde{\Phi}))\times\tilde{\rho}\lambda(\theta)\times u(\tilde{\Phi})}{a(\tilde{\Phi})} - \tau(1-\theta) & if \quad \tilde{\gamma}(\theta) = E\,; \\ 0 & if \quad \tilde{\gamma}(\theta) = NE\,. \end{cases} \tag{C.2}$$

In this setting, in equilibrium, as long as these type-dependent revenue gain functions satisfy a monotone property (see equation (C.14) in Appendix C.1.2), the equilibrium characterization will yield threshold structures consistent with the ones established in the main model in Lemmas 1 and 2 (but with additional details, as discussed further in Appendix D).

Throughout the paper, we focus on the consumer market equilibrium outcome in which the market structure consists of all possible segments (in particular, some users decide to patch, some decide not to patch and would not pay ransom if hit, some decide not to patch but would pay ransom if hit, and some choose not to adopt). Within this consumer market structure, there are two distinct attacker market structures: one in which some attackers go solely for standard attacks and some attackers develop and deploy ransomware, and another in which all attackers who join the attacker market attack with ransomware. In the next two subsections, we derive the equilibrium $R^*(\theta)$ that arises in the generalized model and the equilibrium consumer and attacker thresholds defining the segmentation for each scenario.

**Model in the main body of the paper (Sections 3-5):**
Special case of the above generalized model with attacker heterogeneity only at the entry cost level (i.e., $s(\theta) = \lambda(\theta) = 1$ for all $\theta$).

Under these simplifying assumptions, equations (C.1) and (C.2) reduce to equations (4) and (9), respectively.

Note: In this simplified model, attacker entry decisions are type dependent but all attackers solve the same optimization problem net of entry costs. As such, they optimally choose the *same* ransom amount in equilibrium.

### C.1.1 Market Segmentation and Equilibrium $R^*(\theta)$ with One Attacker Segment under Post-Crypto Ransomware Setting

In this scenario, no attacker goes for standard attacks in equilibrium. Instead, all attackers who join the attacker market opt to choose to attack with ransomware.

To derive the equilibrium $R^*(\theta)$ and the segment thresholds for this equilibrium, we first characterize how the consumer thresholds respond to the attacker population size and a given attacker strategy profile $R(\theta)$. After understanding how consumers respond, we derive the equilibrium $R^*(\theta)$ and threshold attacker type who joins.

To establish the consumer thresholds, we consider their different options and respective payoffs. Let $\tilde{\gamma} : \Theta \to S^a_{RW}$ and $\tilde{\sigma} : \mathcal{V} \to S^c_{RW}$ denote the strategy functions of the attackers and the consumers, respectively, based on the heterogeneity in each group. For the given attacker-consumer profile $\tilde{\Phi} = <\tilde{\gamma}, \tilde{\sigma}>$, the expected utility for a consumer with valuation $v$ is given in Section 3.2, in equation (7). Given some consumer type $v$, from (7), since the first option protects the user against valuation-dependent losses while the second option requires the user to bear some valuation-dependent loss, this means that higher-valuation consumers would prefer to patch rather than not patch.

In the equilibrium that we construct, not all unpatched users will want to pay some ransom (i.e., for some users, all ransoms charged in equilibrium will be higher than their maximum thresholds for paying a ransom). Of those users who remain unpatched, consider which consumers would pay ransom. If a user were to face the decision of paying ransom amount $R$, then they would trade off $R + \delta\alpha v$ vs. $\alpha v$. They would pay if their loss upon simply paying the ransom is less than their total loss $\alpha v$ (i.e., they would pay if $R + \delta\alpha v \leq \alpha v$). Rearranging, this means that given an $R$, only users of type $v \geq \frac{R}{\alpha(1-\delta)}$ would pay. This implies that if all consumer market segments are present in equilibrium, then the lowest tier of consumers consists of those users who are not willing to pay ransom.

Consider the lowest consumer type who uses the software, denoted $v_u$. We simplify the notation by letting $a$ denote the size of the attacker population and $\underline{\theta}$ denote the attacker type joining the market. Since the threshold consumer is not willing to pay any ransom charged in equilibrium, their expected net payoff would be $v - \kappa - q(a) \int_{\underline{\theta}}^{1} \alpha v \left(\frac{1}{a}\right) d\theta$. Since the threshold type $v_u$ deciding to enter the market has zero net surplus in equilibrium, this means that $v_u$ solves $v - \kappa - q(a) \int_{\underline{\theta}}^{1} \alpha v \left(\frac{1}{a}\right) d\theta = 0$. Since there is only one attacker segment in this scenario and the

higher $\theta$ attackers have lower costs of entry than lower $\theta$ ones, this means that $a = 1 - \underline{\theta}$, or equivalently, that $\underline{\theta} = 1 - a$. Substituting this in and simplifying, we have that $v_u = \frac{\kappa}{1-\alpha q(a)}$.

Next, consider the type indifferent between patching and remaining unpatched. Since the higher tier of consumers willing to remain unpatched would be willing to pay ransoms if hit, this means that this type (denoted $v_p$) would be indifferent between patching and remaining unpatched (paying ransom if hit). In equilibrium, this type is willing to pay any ransom charged by an attacker in equilibrium. Otherwise, if there are some ransoms being charged that are higher than what the highest $v$ willing to pay ransom (i.e., $v_p$) would be willing to pay, then no one with valuations less than $v_p$ would be willing to pay that ransom either (and so the attacker who charges that ransom would have a profitable deviation of lowering their ransom charge). Hence, $v_p$ is willing to pay any ransom amount charged in equilibrium by any attacker. Let $R(\theta)$ denote the ransoms charged as a function of $\theta$. Since $v_p$ is indifferent between patching and remaining unpatched (and being willing to pay any ransom charged in equilibrium by an attacker), $v_p$ solves: $v - \kappa - c_p = v - \kappa - q(a) \int_{\underline{\theta}}^{1} (R(\theta) + \delta\alpha v) \left(\frac{1}{a}\right) d\theta$. Again substituting in $\underline{\theta} = 1-a$ and simplifying, we have that $v_p = \dfrac{1}{\alpha\delta} \left( \dfrac{c_p}{q(a)} - \dfrac{\int_{1-a}^{1} R(\theta) d\theta}{a} \right)$.

Given how $v_u$ and $v_p$ respond to attackers' strategies, we can derive the equilibrium $R^*(\theta)$. Consider an attacker of type $\theta$ entering the attacker market and deciding what $R$ to play. Assuming other attackers are playing according to some strategy profile $\tilde{R}(\theta)$ and consumers are best responding to this, we can find this attacker's optimal $R$.

The unpatched population size is $u = v_p - v_u$. Suppose the attacker sets $R$ as their ransom. Then $v \geq \frac{R}{\alpha(1-\delta)}$ would be willing to pay. Given that $v_p$ is the highest $v$ willing to pay a ransom, the size of the consumer population willing to pay ransom $R$ is $v_p - \frac{R}{\alpha(1-\delta)}$. Then in choosing $R$, the attacker of type $\theta$ maximizes

$$\frac{q(a)}{a} \left( Rs(\theta) \left( v_p - \frac{R}{\alpha(1-\delta)} \right) + \tilde{\rho}\lambda(\theta) \left( \frac{R}{\alpha(1-\delta)} - v_u \right) \right) - \tau(1-\theta)(1+\omega). \tag{C.3}$$

Taking the derivative with respect to $R$, the first-order condition gives $R^* = \frac{1}{2} \left( v_p \alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)} \right)$. The second-order condition is satisfied since the second-derivative with respect to $R$ of the expected profit function above is $-\frac{2q(a)s(\theta)}{a\alpha(1-\delta)}$, which is positive.

Note that from $R^* = \frac{1}{2} \left( v_p \alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)} \right)$, we can substitute in our expression for $v_p$ to get $R^* = \frac{1}{2} \left( \left( \frac{1}{\alpha\delta} \left( \frac{c_p}{q(a)} - \frac{\int_{1-a}^{1} R(\theta) d\theta}{a} \right) \right) \alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)} \right)$. In other words, if attackers between types

$\theta = 1 - a$ and $\theta = 1$ choose ransoms according to the function $R(\theta)$ and all consumers respond accordingly, then the previous expression is what type $\theta$ would optimally choose for their ransom.

To derive what the equilibrium $R^*(\theta)$ is, we need to find a function $R(\theta)$ such that

$$R(\theta) = \frac{1}{2}\left(\left(\frac{1}{\alpha\delta}\left(\frac{c_p}{q(a)} - \frac{\int_{1-a}^{1} R(\theta)d\theta}{a}\right)\right)\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right). \tag{C.4}$$

From the previous expression, we see that $R(\theta)$ can be written in a way that $R(\theta) = \frac{1}{2}\left(\frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right) + C$ for some constant $C$ (i.e., something that is constant in $\theta$). Substituting $R(\theta) = \frac{1}{2}\left(\frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right) + C$ into the right-hand side of (C.4), we have that $R(\theta) = \frac{1}{2}\left(\left(\frac{1}{\alpha\delta}\left(\frac{c_p}{q(a)} - \frac{\int_{1-a}^{1}\frac{1}{2}\left(\frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)d\theta}{a} - C\right)\right)\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)$. Equating this expression to (C.4) and solving for $C$, we have that $C = \int_{1-a}^{1}\frac{R(\theta)}{a}d\theta - \int_{1-a}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2as(\theta)}d\theta$. Let $X = \int_{1-a}^{1}\frac{R(\theta)}{a}d\theta$, so $C = X - \int_{1-a}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2as(\theta)}d\theta$. Substituting this expression for $C$ into $R(\theta) = \frac{1}{2}\left(\frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right) + C$, we have that

$$R(\theta) = \frac{1}{2}\left(\frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right) + X - \int_{1-a}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2as(\theta)}d\theta. \tag{C.5}$$

From (C.4), replacing $\int_{1-a}^{1}\frac{R(\theta)}{a}d\theta$ with $X$ gives $R(\theta) = \frac{1}{2}\left(\left(\frac{1}{\alpha\delta}\left(\frac{c_p}{q(a)} - X\right)\right)\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)$. Solving for $X$ in this equation gives $X = \frac{s(\theta)(c_p(1-\delta) - 2\delta q(a)R(\theta)) + \delta\tilde{\rho}q(a)\lambda(\theta)}{(1-\delta)q(a)s(\theta)}$. Substituting this back into (C.5) and then solving for $R(\theta)$ gives the closed-form solution for $R^*(\theta)$:

$$R^*(\theta) = \frac{1}{2}\left(\frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right) + \frac{(1-\delta)\left(c_p - q(a^*)\left(\int_{1-a^*}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2a^*s(\theta)}\,d\theta\right)\right)}{(1+\delta)q(a^*)}, \tag{C.6}$$

where $a^*$ will be characterized later in (C.8).

From the earlier equation $R^* = \frac{1}{2}\left(v_p\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)$, since this holds for any type $\theta$, we can solve for $v_p$ in terms of $R^*(\theta)$. This gives $v_p = \frac{2R^*(\theta)s(\theta) - \tilde{\rho}\lambda(\theta)}{\alpha s(\theta) - \alpha\delta s(\theta)}$. Substituting in (C.6) into the previous equation, we have

$$v_p = \frac{2c_p - 2q(a^*)\left(\int_{1-a^*}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2a^*s(\theta)}\,d\theta\right)}{\alpha\delta q(a^*) + \alpha q(a^*)}. \tag{C.7}$$

To find the equilibrium attacker population size $a^*$, we look back at (C.3). Substituting in

$v_p = \dfrac{2c_p - 2q(a)\left(\int_{1-a}^{1} \frac{\tilde{\rho}\lambda(\theta)}{2as(\theta)}\, d\theta\right)}{\alpha\delta q(a) + \alpha q(a)}$ and $v_u = \frac{\kappa}{1-\alpha q(a)}$ into this expression, substituting in (C.6) for both $R$ and $R(\theta)$, and noting that the threshold type $\underline{\theta}$ can again be written as $\underline{\theta} = 1 - a$ and has zero net profit, we have that the equilibrium attacker population size $a^*$ solves the equation

$$a\tau(1+\omega) = \Bigg(\frac{4(1-\delta)s(1-a)\left(ac_p - q(a)\left(\int_{1-a}^{1} \frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)}\, d\theta\right)\right)^2}{a(1+\delta)^2 q(a)} + \frac{4\tilde{\rho}\lambda(1-a)\left(q(a)\left(a\alpha(\kappa\delta+\kappa+c_p) + (1-\alpha q(a))\left(\int_{1-a}^{1} \frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)}\, d\theta\right)\right) - ac_p\right)}{(1+\delta)(\alpha q(a)-1)} + \frac{a\tilde{\rho}^2\lambda(1-a)^2 q(a)}{(1-\delta)s(1-a)}\Bigg). \quad \text{(C.8)}$$

We use the above (along with work from subsequent sections in this appendix) as part of the numeric exploration part of the paper.

### C.1.2 Market Segmentation and Equilibrium $R^*(\theta)$ with Two Attacker Segments under Post-Crypto Ransomware Setting

The derivation of this equilibrium is similar to that of the previous section, with the difference being that some attackers go for standard attacks in equilibrium in this scenario. We will show that the higher $\theta$ attackers would choose to attack with ransomware over standard attacks in equilibrium under some assumptions. Then we will skip over some of the details of the derivation since the derivation is very similar to that of the previous section and just provide the expressions for the equilibrium $R^*(\theta)$, consumer thresholds, and the equations the attacker population sizes solve.

Consider the lowest consumer type who uses the software, denoted $v_u$. Let $a$ denote the size of the attacker population in (7) and $\underline{\theta}$ denote the attacker type joining the market. Let $\bar{\theta}$ denote the attacker type indifferent between conducting standard attacks and ransomware attacks. We will show under some assumptions later that higher attacker types prefer to conduct ransomware, so that $a_2 = 1 - \bar{\theta}$ is the size of the attacker population conducting ransomware attacks, $a_1 = \bar{\theta} - \underline{\theta}$ is the size of the attacker population conducting standard non-ransomware attacks, and $a = a_1 + a_2$ is the total size of the attacker population.

Same as in the previous subsection, since the threshold consumer is not willing to pay any ransom charged in equilibrium, their expected net payoff would be $v - \kappa - q(a)\int_{\underline{\theta}}^{1} \alpha v \left(\frac{1}{a}\right) d\theta$. Since the threshold type $v_u$ deciding to enter the market has zero net surplus in equilibrium, this

means that $v_u$ solves $v - \kappa - q(a)\int_{\underline{\theta}}^{1} \alpha v \left(\frac{1}{a}\right) d\theta = 0$. Since $a = 1 - \underline{\theta}$ and $a = a_1 + a_2$, we have that $v_u = \frac{\kappa}{1-\alpha q(a_1+a_2)}$.

Next, consider the type indifferent between patching and remaining unpatched. Since the higher tier of consumers willing to remain unpatched would be willing to pay ransoms if hit, this means that this type (denoted $v_p$) would be indifferent between patching and remaining unpatched (paying ransom if hit). As in the previous subsection, this type is willing to pay any ransom charged by an attacker in equilibrium. What is different from the previous subsection is that now consumers have weights on the chance of ransomware attack vs. standard attacks.

Let $R(\theta)$ denote the ransoms charged as a function of $\theta$ for ransomware attackers. Since $v_p$ is indifferent between patching and remaining unpatched (facing the risk of both ransomware and standard attacks), $v_p$ solves: $v - \kappa - c_p = v - \kappa - q(a)\left(\int_{\bar{\theta}}^{1} (R(\theta) + \delta\alpha v)\left(\frac{1}{a}\right) d\theta + \int_{\underline{\theta}}^{\bar{\theta}} (\alpha v)\left(\frac{1}{a}\right) d\theta\right)$. Solving for $v$ and characterizing the expression in terms of $a_1$ and $a_2$ instead of $\underline{\theta}$ and $\bar{\theta}$, we have that

$$v_p = \frac{(a_1 + a_2)\left(c_p - q(a_1 + a_2)\left(\int_{1-a_2}^{1} \frac{R(\theta)}{a_1+a_2}\, d\theta\right)\right)}{\alpha(a_1 + a_2\delta)q(a_1 + a_2)}. \tag{C.9}$$

Given how $v_u$ and $v_p$ respond to attackers' strategies, we can derive the equilibrium $R^*(\theta)$. Consider an attacker of type $\theta$ entering the attacker market and deciding what $R$ to play. Assuming other attackers are playing according to some strategy profile $R^(\theta)$ and consumers are best responding to this, we can find this attacker's optimal $R$.

Same as in the previous section, for a type $\theta$ ransomware attacker choosing $R$, that attacker

$$\frac{q(a)}{a}\left(Rs(\theta)\left(v_p - \frac{R}{\alpha(1-\delta)}\right) + \tilde{\rho}\lambda(\theta)\left(\frac{R}{\alpha(1-\delta)} - v_u\right)\right) - \tau(1-\theta)(1+\omega). \tag{C.10}$$

The first-order condition gives $R^* = \frac{1}{2}\left(v_p\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)$. The second-order condition is satisfied since the second-derivative with respect to $R$ of the expected profit function above is $-\frac{2q(a)s(\theta)}{a\alpha(1-\delta)}$, which is positive. Substituting $R^* = \frac{1}{2}\left(v_p\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)$ into (C.10) to get the profit function at the optimal $R$, we get

$$\Pi(R^*) = \frac{q(a)\left(\tilde{\rho}\lambda(\theta)\left(\frac{\tilde{\rho}\lambda(\theta)}{\alpha s(\theta) - \alpha\delta s(\theta)} - 4v_u + 2v_p\right) - \alpha(\delta-1)v_p^2 s(\theta)\right)}{4a} - (1-\theta)\tau(1+\omega) \tag{C.11}$$

Consider what this attacker would get upon opting to go for standard attacks instead of ran-

somware. Denoting their profit upon entry as a standard attacker as $\Pi(E)$, they would get:

$$\Pi(E) = \frac{q(a)}{a}\left(\tilde{\rho}\lambda(\theta)\left(v_p - v_u\right)\right) - \tau(1-\theta). \tag{C.12}$$

Comparing (C.12) with (C.11), an attacker of type $\theta$ would prefer conducting ransomware over standard attacks if

$$4(1-\delta)\tau\omega\theta + \frac{q(a)(\tilde{\rho}\lambda(\theta) + \alpha(\delta-1)v_p s(\theta))^2}{a\alpha s(\theta)} \geq 4\tau\omega(1-\delta). \tag{C.13}$$

Taking the derivative of the left-hand side with respect to $\theta$, we get

$$\frac{q(a)\left(\alpha^2(1-\delta)^2 v_p^2 s(\theta)^2 - \tilde{\rho}^2\lambda(\theta)^2\right)}{a\alpha s(\theta)^2}s'(\theta) + \frac{2\tilde{\rho}q(a)(\tilde{\rho}\lambda(\theta) - \alpha(1-\delta)v_p s(\theta))}{a\alpha s(\theta)}\lambda'(\theta) + 4(1-\delta)\tau\omega. \tag{C.14}$$

This is positive as long as $s'(\theta)$ and $\lambda'(\theta)$ are sufficiently close to 0 for all $\theta$. Since $s(\theta)$ and $\lambda(\theta)$ are increasing functions of $\theta$ such that $s(1) = 1$ and $\lambda(1) = 1$, as long as $s(0)$ and $\lambda(0)$ are not too small (i.e., close enough to 1 rather than close to 0), then the left-hand side of (C.13) is monotonically increasing in $\theta$ so that higher types of attackers would prefer conducting ransomware over standard attacks. Under this assumption, we can derive the equilibrium $R^*(\theta)$ that arises.

Following the proof of the previous subsection, we substitute $v_p$ from (C.9) into the solution from the first-order condition, $R^* = \frac{1}{2}\left(v_p\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)$. In the same way we had done in the previous subsection, we can find $R^*(\theta)$ in terms of the attacker population segments $a_1$ and $a_2$:

$$R(\theta) = \frac{(1-\delta)\left(c_p(a_1^* + a_2^*) - a_2^* q(a_1^* + a_2^*)\left(\int_{1-a_2^*}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2a_2^* s(\theta)}\,d\theta\right)\right)}{(2a_1^* + a_2^*\delta + a_2^*)q(a_1^* + a_2^*)} + \frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)}. \tag{C.15}$$

From the earlier equation $R^* = \frac{1}{2}\left(v_p\alpha(1-\delta) + \frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right)$, since this holds for any type $\theta$, we can solve for $v_p$ in terms of $R^*(\theta)$. As in the previous subsection, this gives $v_p = \frac{2R^*(\theta)s(\theta) - \tilde{\rho}\lambda(\theta)}{\alpha s(\theta) - \alpha\delta s(\theta)}$. Substituting in (C.15) into the previous equation, we have

$$v_p = \frac{2c_p(a_1^* + a_2^*) - 2a_2^* q(a_1^* + a_2^*)\left(\int_{1-a_2^*}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2a_2^* s(\theta)}\,d\theta\right)}{\alpha(2a_1^* + a_2^*\delta + a_2^*)q(a_1^* + a_2^*)}. \tag{C.16}$$

Using the previous expression for $v_p$ along with $v_u = \frac{\kappa}{1-\alpha q(a_1+a_2)}$, we derive the equations that $a_1$ and $a_2$ have to solve. Using that type $\underline{\theta}$ is indifferent between entering and not entering, this means that this type is indifferent between conducting standard attacks and not being in the market at all

(obtaining zero net profit). On the other hand, type $\bar{\theta}$ is indifferent between conducting ransomware attacks and standard attacks. These two equations defining the threshold attacker types can be expressed in terms of the equations the equilibrium $a_1^*$ and $a_2^*$ solve below:

$$\frac{\tilde{\rho}\left(q(a_1+a_2)\left((2-2\alpha q(a_1+a_2))\left(\int_{1-a_2}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)}\,d\theta\right)+2a_1\alpha(\kappa+c_p)+a_2\alpha(\kappa\delta+\kappa+2c_p)\right)-2c_p(a_1+a_2)\right)}{\alpha(a_1+a_2)(2a_1+a_2\delta+a_2)(\alpha q(a_1+a_2)-1)}-$$
$$\tau(a_1+a_2)=0 \quad \text{(C.17)}$$

and

$$\frac{q(a_1+a2)}{a_1+a_2}\Bigg(-\frac{(\delta-1)s(1-a_2)\left(c_p(a_1+a_2)-q(a_1+a_2)\left(\int_{1-a_2}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)}\,d\theta\right)\right)^2}{\alpha(2a_1+a_2\delta+a_2)^2q(a_1+a_2)^2}+$$
$$\frac{\tilde{\rho}\lambda(1-\text{a2})\left(c_p(a_1+a_2)-q(a_1+a_2)\left(\int_{1-a_2}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)}\,d\theta\right)\right)}{\alpha(2a_1+a_2\delta+a_2)q(a_1+a_2)}+\frac{\kappa\tilde{\rho}\lambda(1-a_2)}{\alpha q(a_1+a_2)-1}-\frac{\tilde{\rho}^2\lambda(1-a_2)^2}{4\alpha(\delta-1)s(1-a_2)}\Bigg)-$$
$$\left(\frac{2\tilde{\rho}\left(c_p(a_1+a_2)-q(a_1+a_2)\left(\int_{1-a_2}^{1}\frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)}\,d\theta\right)\right)}{\alpha(a_1+a_2)(2a_1+a_2\delta+a_2)}+\frac{\kappa\tilde{\rho}q(a_1+a_2)}{(a_1+a_2)(\alpha q(a_1+a_2)-1)}+a_2\tau\omega\right)=0. \quad \text{(C.18)}$$

We use the above equations (along with work from the rest of the appendix) as part of the numeric exploration part of the paper. The core part of the paper focuses on when $\lambda(\theta)=1$ and $s(\theta)=1$ for all $\theta$, allowing for more tractable analyses. In the numeric explorations section, we explore how robust the results are as we relax the assumption that $\lambda(\theta)=1$ and $s(\theta)=1$ for all $\theta$.

## C.2 Pre-Crypto Benchmark Setting - Equilibrium Outcome with $\lambda(\theta)=1$ for all $\theta$

To start off the core section of the paper, we characterize the benchmark equilibrium outcome (prior to crypto-enabled ransomware). In a scenario without ransomware in the market, consumers who use the software have two options when a patchable vulnerability arises: 1) Patch or 2) Remain unpatched and bear the risk associated with remaining unpatched. Throughout the paper, we focus on the equilibrium outcome in this benchmark scenario in which some users patch and some do

not. Letting $\Phi = <\gamma, \sigma>$ denote an *attacker-consumer strategy profile*, the net utility function is:

$$U^c_{BM}(v, \Phi) \triangleq \begin{cases} v - \kappa - c_p & if \quad \sigma(v) = (A, P)\,; \\ v - \kappa - q(a(\Phi)) \times \alpha v & if \quad \sigma(v) = (A, NP)\,; \\ 0 & if \quad \sigma(v) = (NA, NP)\,. \end{cases} \tag{C.19}$$

Simplifying the notation by letting $a$ denote the size of the attacker population that arises, consumers who patch get a net payoff of $v - \kappa - c_p$ while consumers who remain unpatched get a net payoff of $v - \kappa - q(a) \times \alpha v$. Since patching protect from valuation-dependent losses, higher-valuation users would patch. Specifically, comparing those two options, $v \geq \frac{c_p}{\alpha q(a)}$ would prefer to patch over remaining unpatched. Define this threshold as $v_p$, the consumer type indifferent between patching and not patching. On the lower end of the market, the consumer type indifferent between remaining unpatched and not using the software solves $v - \kappa - q(a)\alpha v = 0$, so $v_u = \frac{\kappa}{1-\alpha q(a)}$ is the type indifferent between using the software and not using the software.

From the attacker's perspective in the benchmark case, they just decide whether or not to join the market (i.e., they do not set ransoms prior to crypto-enabled ransomware). An attacker of type $\theta$'s net profit upon entering the market to attack is given by

$$\Pi(E) = \frac{q(a)}{a} \rho \left(v_p - v_u\right) - \tau(1 - \theta). \tag{C.20}$$

The marginal type $\underline{\theta}$ entering the attacker market has zero profit, so that type is such that $\Pi(E) = 0$. Noting that $\underline{\theta} = 1 - a$, we have that the equilibrium attacker population size $a^*$ solves

$$a\tau = \frac{c_p \rho}{a\alpha} - \frac{\kappa \rho q(a)}{a - a\alpha q(a)}. \tag{C.21}$$

As a function of $a$, note that the right-hand side has a limit of $+\infty$ as $a \to 0$ (since $\frac{c_p\rho}{a\alpha} \to +\infty$ while $\frac{\kappa\rho q(a)}{a - a\alpha q(a)} \to \kappa\rho q'(0)$ by L'Hôpital's rule). The right-hand side goes to $-\infty$ as $a \to q^{-1}\left(\frac{1}{\alpha}\right)$. The derivative of the right-hand side of (C.21) is $-\frac{c_p\rho}{a^2\alpha} + \frac{\rho\left(-aq'(a) - \alpha q(a)^2 + q(a)\right)}{a^2(\alpha q(a) - 1)^2}\kappa$. For sufficiently small $\kappa$, this is negative so that the right-hand side is decreasing in $a$ from $+\infty$ as $a \to 0$ to $-\infty$ as $a \to q^{-1}\left(\frac{1}{\alpha}\right)$. On the other hand, the left-hand side is simply a linearly increasing function of $a$. Altogether, this implies that for sufficiently small $\kappa$, there is a unique solution in $a$ to (C.21) between 0 and $q^{-1}\left(\frac{1}{\alpha}\right)$. This defines $a^*$ in the benchmark case. Using this $a^*$, the equilibrium

consumer thresholds can be defined as

$$v_p^* = \frac{c_p}{\alpha q(a^*)} \tag{C.22}$$

and

$$v_u^* = \frac{\kappa}{1 - \alpha q(a^*)}. \tag{C.23}$$

In the next subsection, we characterize the equilibrium outcome in the post-crypto scenario (i.e., a world in which attackers can attack with ransomware).

## C.3 Post-Crypto Ransomware Setting - Equilibrium Outcome with $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$

In the focal region, two distinct attacker market structures arise: one in which some attackers go for standard attacks and some attackers develop and deploy ransomware, and another in which all attackers who join the attacker market attack with ransomware. In the next two subsections, we characterize both of these outcomes.

### C.3.1 Equilibrium Outcome with One Attacker Segment

This is a special case of the equilibrium characterized in Subsection C.1.1, with $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$.

From Subsection C.1.1 and using $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$, we have

$$R^*(\theta) = \frac{\delta\tilde{\rho}q(a) + c_p(1-\delta)}{(1+\delta)q(a^*)} \tag{C.24}$$

for all $\theta \geq 1 - a^*$, where $a^*$ solves

$$a\tau(1+\omega) = \frac{a(\delta\tilde{\rho}q(a) + c_p(1-\delta))^2}{\alpha(1-\delta)(a\delta+a)^2 q(a)} - \frac{\kappa\tilde{\rho}q(a)}{a - a\alpha q(a)}. \tag{C.25}$$

The derivative of the right-hand side of the equation defining $a^*$ with respect to $a$ has a zero-order Taylor series expansion around $\kappa = 0$ and $\delta = 0$ of $-\frac{c_p^2}{a^2\alpha q(a)} - \frac{c_p^2 q'(a)}{a\alpha q(a)^2} + O(\kappa + \delta)$, which is negative for $\kappa$ and $\delta$ close enough to 0. Taking the limit as $a \to 0$ from above of the right-hand side of (C.25), we have that the right-hand side goes to $+\infty$ as $a \to 0$ from above. As $a \to q^{-1}\left(\frac{1}{\alpha}\right)$ from below, the right-hand side has a limit of $-\infty$. Altogether, this means that for sufficiently small $c$

and $\delta$, there's a unique solution to the equation defining $a^*$ between 0 and $q^{-1}\left(\frac{1}{\alpha}\right)$.

Given this $a^*$, the equilibrium ransom charged is given in (C.24). From substituting $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$ into (C.7), we have

$$v_p = \frac{2c_p - \tilde{\rho}q(a^*)}{\alpha q(a^*)(1+\delta)}. \tag{C.26}$$

Also, we can characterize the marginal type indifferent between using the software and not using it in equilibrium. As was the case in Subsection C.1.1,

$$v_u = \frac{\kappa}{1 - \alpha q(a^*)}. \tag{C.27}$$

Lastly, we can define the consumer threshold type willing to pay some ransom in equilibrium. Since a consumer of type $v$ is willing to pay ransom if $R + \delta\alpha v \leq \alpha v$, this means that all $v \geq \frac{R}{\alpha(1-\delta)}$ is willing to pay ransom $R$ rather than not paying it. Substituting in (C.24) for $R$, we can define type

$$v_r = \frac{c_p(1-\delta) + \delta\tilde{\rho}q(a^*)}{\alpha q(a^*)(1-\delta^2)} \tag{C.28}$$

as the type indifferent between paying ransom and not paying ransom in equilibrium.

Before we get into characterizing the conditions under which this case can arise in equilibrium (for example, conditions to ensure $0 < v_u < v_r < v_p < 1$, among other things), we first characterize the thresholds in the equilibrium outcome with two segments of attackers.

### C.3.2 Equilibrium Outcome with Two Attacker Segments

This is a special case of the equilibrium characterized in Subsection C.1.2, with $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$.

From Subsection C.1.2 and using $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$, we have

$$R^*(\theta) = \frac{c_p(1-\delta)(a_1^* + a_2^*) + \tilde{\rho}(a_1^* + a_2^*\delta)q(a_1^* + a_2^*)}{(2a_1^* + a_2^*\delta + a_2^*)q(a_1^* + a_2^*)} \tag{C.29}$$

for all $\theta \geq 1 - a_2^*$, where $a_1^*$ and $a_2^*$ (the attacker population sizes for standard attacks and ransomware attackers, respectively) solve

$$\frac{\tilde{\rho}(2c_p(a_1+a_2) - a_2\tilde{\rho}q(a_1+a_2))}{\alpha(a_1+a_2)(2a_1+a_2\delta+a_2)} + \frac{\kappa\tilde{\rho}q(a_1+a_2)}{(a_1+a_2)(\alpha q(a_1+a_2)-1)} - \tau(a_1+a_2) = 0 \tag{C.30}$$

and

$$c_p^2(1-\delta)^2(a_1+a_2) - (1-\delta)q(a_1+a_2)\left(a_2\alpha\tau\omega(2a_1+a_2\delta+a_2)^2 + 2c_p\tilde{\rho}(a_1+a_2)\right) + \tilde{\rho}^2(a_1+a_2)q(a_1+a_2)^2 = 0 \quad \text{(C.31)}$$

Higher $\theta$ attackers prefer conducting ransomware over standard attacks from the argument in Subsection C.1.2, namely that the left-hand side of (C.13) is monotonically increasing in $\theta$ due to $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$. This means that types $\theta \geq \bar{\theta} = 1 - a_2^*$ conduct ransomware attacks and types $\theta \in [\underline{\theta}, \bar{\theta})$ conduct standard attacks (where $\underline{\theta} = 1 - a_1^* - a_2^*$).

Given $a_1^*$ and $a_2^*$, the equilibrium ransom charged is given in (C.29). From substituting $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$ into (C.16), we have

$$v_p = \frac{2c_p(a_1^*+a_2^*) - a_2^*\tilde{\rho}q(a_1^*+a_2^*)}{\alpha(2a_1^*+a_2^*\delta+a_2^*)q(a_1^*+a_2^*)}. \quad \text{(C.32)}$$

As was the case in Subsection C.1.2, the marginal type indifferent between using the software and not using it in equilibrium is given by:

$$v_u = \frac{\kappa}{1-\alpha q(a^*)} = \frac{\kappa}{1-\alpha q(a_1^*+a_2^*)}. \quad \text{(C.33)}$$

Lastly, we can define the consumer threshold type willing to pay some ransom in equilibrium. Since a consumer of type $v$ is willing to pay ransom if $R + \delta\alpha v \leq \alpha v$, this means that all $v \geq \frac{R}{\alpha(1-\delta)}$ is willing to pay ransom $R$ rather than not paying it. Substituting in (C.29) for $R$, we can define type

$$v_r = \frac{c_p(1-\delta)(a_1^*+a_2^*) + \tilde{\rho}(a_1^*+a_2^*\delta)q(a_1^*+a_2^*)}{\alpha(1-\delta)(2a_1^*+a_2^*\delta+a_2^*)q(a_1^*+a_2^*)}. \quad \text{(C.34)}$$

as the type indifferent between paying ransom and not paying ransom in equilibrium.

In the next subsection, we characterize the focal region of parameters that we focus on throughout the analysis. This focal region is the intersection of the set of conditions needed for $0 < v_u < v_p < 1$ to arise under the benchmark scenario, $0 < v_u < v_r < v_p < 1$ to arise under the ransomware scenario, and $a^* < 1$.

## C.4 Proofs of Lemmas and Propositions in Main Text

**Proof of Lemmas 1 and 2:** This follows from the work in the prior subsections. To summarize, we will repeat some of the arguments here.

To establish consumer preference rankings (and a threshold structure for the equilibrium), note that patching protects users from valuation-dependent losses while remaining unpatched exposes users to valuation-dependent losses. This implies higher-valuation users prefer to patch over remaining unpatched. Similarly, since paying ransom reduces valuation-dependent losses compared to not paying ransom, higher-valuation users prefer to pay ransoms than not pay. Specifically, comparing the net utility of the different options, $v \geq \frac{c_p}{\alpha q(a)}$ prefer to patch over remaining unpatched in the benchmark case (equivalently, these users prefer to patch over remain unpatched and not pay ransom in the ransomware scenario). Given a ransom $R$, users with valuations $v \geq \frac{R}{\alpha(1-\delta)}$ prefer to pay ransom over not paying ransom. Similarly, users with valuations $v \geq \frac{c_p - Rq(a)}{\alpha\delta q(a)}$ prefer to patch over remaining unpatched (facing chance of attack $q(a)$) and paying ransom $R$ if hit. This establishes the consumer preference ordering and the threshold equilibrium structure.

To see that higher $\theta$ attackers prefer to conduct ransomware attacks over standard attacks, see the argument following (C.13). That higher $\theta$ attackers prefer to enter the market at all over not entering follows from the cost function being linearly decreasing in $\theta$. For more details relating to Lemma 1 and the equilibrium characterization of the benchmark case, see Subsection C.2. For more details relating to Lemma 2 and the equilibrium characterization of the ransomware case, see Subsection C.3. ■

**Proof of Remark 1:** Under the benchmark scenario, the attacker population size is given in (C.21). Viewing the equilibrium attacker population size under the benchmark, $a_{BM}$, as a function of $\kappa$ and implicitly differentiating (C.21) with respect to $\kappa$, then $a'_{BM}(\kappa)$ is given by

$$a'_{BM}(\kappa) = \frac{\alpha\rho a(\kappa)q(a(\kappa))(\alpha q(a(\kappa)) - 1)}{(\alpha q(a(\kappa)) - 1)\left(\alpha\rho(\kappa + c_p)q(a(\kappa)) + \alpha\tau a(\kappa)^2(\alpha q(a(\kappa)) - 1) - c_p\rho\right) + \alpha\kappa\rho a(\kappa)q'(a(\kappa))}, \tag{C.35}$$

where the subscripts "BM" for $a_{BM}(\kappa)$ will be left off to keep the notation simpler. For sufficiently small $\kappa$ (i.e., taking a Taylor series expansion around $\kappa = 0$), this can be written as

$$a'(\kappa) = -\frac{a(0)\alpha\rho q(a(0))}{(1 - \alpha q(a(0)))\left(a(0)^2\alpha\tau + c_p\rho\right)} + O(\kappa), \tag{C.36}$$

where $a(0)$ is what $a^*$ (for the benchmark case) approaches as $\kappa \to 0$. Since $1-\alpha q(a) > 0$ holds for any $\kappa$ from $v_u > 0$ in equilibrium (recall $v_u = \frac{\kappa}{1-\alpha q(a)}$), it follows that $a'(\kappa) < 0$ for sufficiently small $\kappa$ for the benchmark scenario.

Under the ransomware scenario when $\tau > \hat{\tau}$ (so that there is only one attacker segment in equilibrium), the attacker population size is given in (C.25). Viewing the equilibrium $a$ as a function of $\kappa$ and implicitly differentiating (C.25) with respect to $\kappa$, then $a'(\kappa)$ is given by

$$a'(\kappa) = \Bigg(\alpha(\delta-1)(\delta+1)^2\tilde{\rho}a(\kappa)q(a(\kappa))^3(\alpha q(a(\kappa))-1)\Bigg)\Bigg(-q(a(\kappa))^3\Bigg(2\alpha a(\kappa)\Bigg(\alpha(\delta-1)(\delta+1)^2\tau(\omega+1)a(\kappa)+ \delta^2\tilde{\rho}^2q'(a(\kappa))\Bigg) + \alpha(\delta-1)\tilde{\rho}\Bigg(\kappa(\delta+1)^2+4c_p\delta\Bigg) + \alpha^2c_p^2(\delta-1)^2+\delta^2\tilde{\rho}^2\Bigg) + q(a(\kappa))^2\Bigg(a(\kappa)q'(a(\kappa))\Bigg(\tilde{\rho}\times \Bigg(\alpha\kappa(\delta-1)(\delta+1)^2+\delta^2\tilde{\rho}\Bigg)-\alpha^2c_p^2(\delta-1)^2\Bigg)+(\delta-1)\Bigg(\alpha(\delta+1)^2\tau(\omega+1)a(\kappa)^2+2c_p(\alpha c_p(\delta-1)+\delta\tilde{\rho})\Bigg)\Bigg)+ c_p^2(\delta-1)^2q(a(\kappa))\Bigg(2\alpha a(\kappa)q'(a(\kappa))-1\Bigg)-c_p^2(\delta-1)^2a(\kappa)q'(a(\kappa))+\alpha q(a(\kappa))^4\Bigg(\alpha a(\kappa)\Bigg(\alpha(\delta-1)(\delta+1)^2\tau\times (\omega+1)a(\kappa)+\delta^2\tilde{\rho}^2q'(a(\kappa))\Bigg)+\tilde{\rho}\Bigg(\alpha(\delta-1)\Bigg(\kappa(\delta+1)^2+2c_p\delta\Bigg)+2\delta^2\tilde{\rho}\Bigg)\Bigg)-\alpha^2\delta^2\tilde{\rho}^2q(a(\kappa))^5\Bigg)^{-1}, \tag{C.37}$$

where the $a'(\kappa)$ and $a(\kappa)$ above are interpreted to be for the ransomware case (not to be confused with the same notation used for the benchmark case). Taking the limit as $\kappa \to 0$ and $\delta \to 0$, we see that for sufficiently small $\kappa$ and $\delta$, this can be expressed as

$$a'(\kappa) \;=\; \frac{a(0)\alpha\tilde{\rho}q(a(0))^3}{(\alpha q(a(0))-1)\left(a(0)c_p^2q'(a(0))+q(a(0))\left(a(0)^2\alpha\tau(\omega+1)q(a(0))+c_p^2\right)\right)} \;+\; O(\kappa+\delta), \tag{C.38}$$

where $a(0)$ is what $a^*$ (for the ransomware case) approaches as $\kappa \to 0$. The numerator is positive. Since $1-\alpha q(a) > 0$ holds from $v_u > 0$, it follows that $\alpha q(a(0)) - 1 < 0$ so that the denominator is negative. Altogether, this means that $a'(\kappa) < 0$ for sufficiently small $\kappa$ and $\delta$ for the ransomware scenario when $\tau > \hat{\tau}$. ■

**Proof of Proposition 1:** i) For the benchmark scenario, the attacker population size is given by

(C.21). Viewing $a$ as a function of $\tau$ and implicitly differentiating with respect to $\tau$, we have that

$$a'(\tau) = -\left(\alpha a(\tau)^3(\alpha q(a(\tau))-1)^2\right)\Bigg((\alpha q(a(\tau))-1)\Big(\alpha\rho(\kappa+c_p)q(a(\tau))+\alpha\tau a(\tau)^2(\alpha q(a(\tau))-1)- c_p\rho\Big)+\alpha\kappa\rho a(\tau)q'(a(\tau))\Bigg)^{-1}. \quad \text{(C.39)}$$

Let $a_{0,BM}$ be the unique positive solution to (C.21) when $\kappa$ is 0. Specifically, $a_{0,BM}$ is the unique positive solution to $a^2\tau = \frac{c_p\rho}{\alpha a}$. This has a unique solution since the left-hand side is strictly increasing in $a$ while the right-hand side is strictly decreasing in $a$ (the existence of the solution comes from the focal region conditions).

Let $a_{0,BM}(\tau)$ denote $a_{0,BM}$ as a function of $\tau$. For sufficiently small $\kappa$, (C.39) can be written in the following way (i.e., taking a Taylor series approximation of (C.39) around $\kappa = 0$, we can write (C.39) as follows):

$$a'(\tau) = -\frac{\alpha a(\tau)^3}{c_p\rho+\alpha\tau a(\tau)^2} + O(\kappa). \quad \text{(C.40)}$$

This is negative for sufficiently small $\kappa$, so the attacker population size decreases under the benchmark scenario when $0 < v_u < v_p < 1$ arises.

ii) For the ransomware scenario, let $\hat{\tau}$ be the lower $\tau$ bound of the scenario with one attacker segment (i.e., one in which all attackers who enter the market conduct ransomware attacks). For $\tau \in (\hat{\tau}, \bar{\tau})$, the consumer equilibrium outcome is $0 < v_u < v_r < v_p < 1$ and all attackers who enter conduct ransomware attacks in equilibrium (characterized by Subsection C.3.1).

The equilibrium attacker population size is given in (C.25). Viewing $a$ as a function of $\tau$ and

implicitly differentiating with respect to $\tau$, we have that

$$a'(\tau) = -\Bigg((\delta+1)^3(\omega+1)a(\tau)^3(\alpha-\alpha\delta)q(a(\tau))^2(\alpha q(a(\tau))-1)^2\Bigg)$$
$$\Bigg((\delta+1)a(\tau)q'(a(\tau))\Bigg(q(a(\tau))\Bigg((q(a(\tau))\Bigg(\alpha\delta^2\tilde{\rho}^2q(a(\tau))(2-\alpha q(a(\tau)))+\tilde{\rho}\Bigg(\delta^2(-\tilde{\rho})-\alpha\kappa(\delta-1)(\delta+1)^2\Bigg)+$$
$$\alpha^2c_p^2(\delta-1)^2\Bigg)-2\alpha c_p^2(\delta-1)^2\Bigg)+c_p^2(\delta-1)^2\Bigg)+\alpha^2(-\kappa)(\delta-1)(\delta+1)^3\tilde{\rho}q(a(\tau))^4+\alpha\kappa(\delta-1)(\delta+1)^3\tilde{\rho}q(a(\tau))^3+$$
$$2\delta q(a(\tau))(\alpha q(a(\tau))-1)^2(\delta\tilde{\rho}q(a(\tau))+c_p(-\delta)+c_p)^2-(\delta+1)q(a(\tau))(\alpha q(a(\tau))-1)^2(\delta\tilde{\rho}q(a(\tau))+c_p(-\delta)+$$
$$c_p)^2+2q(a(\tau))(\alpha q(a(\tau))-1)^2(\delta\tilde{\rho}q(a(\tau))+c_p(-\delta)+c_p)^2-\alpha(\delta-1)(\delta+1)^3\tau(\omega+1)a(\tau)^2\times$$
$$q(a(\tau))^2(\alpha q(a(\tau))-1)^2\Bigg)^{-1}. \quad \text{(C.41)}$$

Let $a_0$ be the unique positive solution to (C.25) when $\kappa$ and $\delta$ are both 0. Specifically, $a_0$ is the unique positive solution to $a^2\tau(1+\omega) = \frac{c_p^2}{\alpha q(a)}$. This has a unique solution since the left-hand side is strictly increasing in $a$ while the right-hand side is strictly decreasing in $a$ (the existence of the solution comes from the focal region conditions).

Let $a_0(\tau)$ denote $a_0$ as a function of $\tau$. For sufficiently small $\kappa$ and $\delta$, (C.41) can be written in the following way (i.e., taking a Taylor series approximation of (C.41) around $\kappa = 0$ and $\delta = 0$, we can write (C.41) as follows):

$$a'(\tau) = -\frac{\alpha(\omega+1)a_0(\tau)^3q(a_0(\tau))^2}{c_p^2a_0(\tau)q'(a_0(\tau))+q(a_0(\tau))\left(\alpha\tau(\omega+1)a_0(\tau)^2q(a_0(\tau))+c_p^2\right)}+O(\kappa+\delta). \quad \text{(C.42)}$$

This is negative for sufficiently small $\kappa$ and $\delta$, so the attacker population size decreases in $\tau$ for $\tau \in (\hat{\tau}, \bar{\tau})$.

For the lower range of $\tau$, we show that $\frac{d}{d\tau}[a_1+a_2] < 0$ at the smaller $\tau$ boundary defining the case (i.e., $\underline{\tau}$) for sufficiently small $\kappa$ and $\delta$. Viewing $a_1$ and $a_2$ as functions of $\tau$, implicitly differentiating (C.30) and (C.31), solving for $\frac{d}{d\tau}[a_1]$ and $\frac{d}{d\tau}[a_2]$, evaluating this at $\tau = \underline{\tau}$ and noting that $a_1(\underline{\tau}) = q^{-1}\left(\frac{c_p(1-\delta)}{\tilde{\rho}}\right)$ and $a_2(\underline{\tau}) = 0$, we have $\frac{d}{d\tau}[a_1+a_2]|_{\tau=\underline{\tau}} = -\frac{\alpha \mathrm{S} q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^3}{2c_p\tilde{\rho}}+O(\kappa+\delta)$. By continuity, $\frac{d}{d\tau}[a_1+a_2] < 0$ also holds for some right-sided neighborhood around $\underline{\tau}$, so the attacker population size shrinks for low $\tau$ in this case.

On the other hand, to show that there is an intermediate range of $\tau$ for which $\frac{d}{d\tau}[a_1+a_2] > 0$, we show that $\frac{d}{d\tau}[a_1+a_2] > 0$ at the larger $\tau$ boundary defining the two attacker segment range of $\tau$ (i.e., $\hat{\tau}$) for sufficiently small $\kappa$ and $\delta$. Viewing $a_1$ and $a_2$ as functions of $\tau$, implicitly

differentiating (C.30) and (C.31), solving for $\frac{d}{d\tau}[a_1]$ and $\frac{d}{d\tau}[a_2]$, evaluating this at $\tau = \hat{\tau}$ and noting that $a_1(\hat{\tau}) = 0$ and $a_2(\hat{\tau}) = q^{-1}\left(\frac{c_p(1+\omega-\sqrt{\omega(\omega+1)})}{\tilde{\rho}\omega+\tilde{\rho}}\right) + O(\kappa+\delta)$, we have $\frac{d}{d\tau}[a_1+a_2]|_{\tau=\hat{\tau}} = \frac{\alpha\left(3\sqrt{\omega(\omega+1)}+\omega\left(-4\omega+4\sqrt{\omega(\omega+1)}-5\right)-1\right)q^{-1}\left(\frac{c_p\omega-c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)^3}{\tilde{\rho}\left(\tilde{\rho}\left(-3\omega+\sqrt{\omega(\omega+1)}-3\right)q^{-1}\left(\frac{c_p\omega-c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)q'\left(q^{-1}\left(\frac{c_p\omega-c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)+2c_p\left(2\omega-2\sqrt{\omega(\omega+1)}+1\right)\right)} + O(\kappa+\delta)$. Comparing this expression to 0, we want to show

$$q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)q'\left(q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right) > \frac{2c_p\left(-2\omega+2\sqrt{\omega(\omega+1)}-1\right)}{\tilde{\rho}\left(-3\omega+\sqrt{\omega(\omega+1)}-3\right)} \quad \text{(C.43)}$$

in order for $\frac{d}{d\tau}[a_1+a_2]|_{\tau=\hat{\tau}} > 0$. Letting $y = \frac{c_p(1+\omega-\sqrt{\omega(\omega+1)})}{\tilde{\rho}\omega+\tilde{\rho}}$, we want to find a condition so that $\frac{q^{-1}(y)q'(q^{-1}(y))}{y} > \frac{1}{\frac{3}{2}+\omega+\sqrt{\omega(1+\omega)}}$. Define the function $H(y) = \frac{q^{-1}(y)q'(q^{-1}(y))}{y}$. We will show that $H(y)$ is a decreasing function of $y$. Note that $H'(y) < 0$ is equivalent to $q''\left(q^{-1}(y)\right) < q'\left(q^{-1}(y)\right)\left(\frac{q'\left(q^{-1}(y)\right)}{y} - \frac{1}{q^{-1}(y)}\right)$. Make a substitution of $x = q^{-1}(y)$, so that $H(y)$ decreasing in $y$ is equivalent to showing $q''(x) < q'(x)\left(\frac{q'(x)}{q(x)} - \frac{1}{x}\right)$. This in turn is implied by the elasticity assumption on $q$, namely that $\frac{d}{dx}\left[\frac{q'(x)}{\frac{q(x)}{x}}\right] < 0$ for all $x$. So altogether, $H(y)$ is a decreasing function of $y$.

Then $\frac{q^{-1}(y)q'(q^{-1}(y))}{y} > \frac{1}{\frac{3}{2}+\omega+\sqrt{\omega(1+\omega)}}$ can be written as $H(y) > \frac{1}{\frac{3}{2}+\omega+\sqrt{\omega(1+\omega)}}$. Since $H(y)$ is a decreasing function, this is equivalent to $y < H^{-1}\left(\frac{1}{\frac{3}{2}+\omega+\sqrt{\omega(1+\omega)}}\right)$. Substituting back in $y = \frac{c_p(1+\omega-\sqrt{\omega(\omega+1)})}{\tilde{\rho}\omega+\tilde{\rho}}$, we want a condition so that $\frac{c_p(1+\omega-\sqrt{\omega(\omega+1)})}{\tilde{\rho}\omega+\tilde{\rho}} < H^{-1}\left(\frac{1}{\frac{3}{2}+\omega+\sqrt{\omega(1+\omega)}}\right)$. This can be written as $\frac{c_p}{\tilde{\rho}} < \left(1+\omega+\sqrt{\omega(\omega+1)}\right)H^{-1}\left(\frac{1}{\omega+\sqrt{\omega(\omega+1)}+\frac{3}{2}}\right)$. The derivative of the right-hand side with respect to $\omega$ is given by $\left(\frac{2\omega+1}{2\sqrt{\omega(\omega+1)}}+1\right)\left(H^{-1}\left(\frac{1}{\omega+\sqrt{\omega(\omega+1)}+\frac{3}{2}}\right) - \frac{\omega+\sqrt{\omega(\omega+1)}+1}{\left(\omega+\sqrt{\omega(\omega+1)}+\frac{3}{2}\right)^2 H'\left(H^{-1}\left(\frac{1}{\omega+\sqrt{\omega(\omega+1)}+\frac{3}{2}}\right)\right)}\right)$, which is altogether positive since $H$ being a decreasing function means that $H'\left(H^{-1}\left(\frac{1}{\omega+\sqrt{\omega(\omega+1)}+\frac{3}{2}}\right)\right) < 0$.

Define the function $G(\omega) = \left(1+\omega+\sqrt{\omega(\omega+1)}\right)H^{-1}\left(\frac{1}{\omega+\sqrt{\omega(\omega+1)}+\frac{3}{2}}\right)$. Since $G(\omega)$ is increasing in $\omega$, we can write the needed condition as $\omega > \hat{\omega}$ where $\hat{\omega} = G^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)$. With this condition, we have that $\frac{d}{d\tau}[a_1+a_2]|_{\tau=\hat{\tau}} > 0$. By continuity of the derivative, this condition $\omega > \hat{\omega}$ implies that the attacker population size is increasing in some left-neighborhood of $\hat{\tau}$ for sufficiently small $\kappa$ and $\delta$. ■

**Proof of Proposition 2:** Recall again that $\hat{\tau}$ is defined to be the lower $\tau$ bound of the scenario with one attacker segment (i.e., one in which all attackers who enter the market conduct ransomware attacks). For $\tau \in (\hat{\tau}, \bar{\tau})$, the consumer equilibrium outcome is $0 < v_u < v_r < v_p < 1$ and all attackers who enter conduct ransomware attacks in equilibrium (characterized by Subsection C.3.1). From Proposition 1, we have that the attacker population size shrinks in $\tau$ when this market outcome arises.

For the ransom-paying population size (i.e., the size of the population willing to pay ransom if hit), in this scenario, that is defined to be $r(\sigma^*) = v_p - v_r$. We have the equilibrium $v_p$ and $v_r$ thresholds in (C.26) and (C.28), respectively. Simplifying, $v_p - v_r = \frac{c_p(1-\delta)-\tilde{\rho}q(a)}{\alpha(1-\delta^2)q(a)}$. Viewing $a$ as a function of $\tau$ and differentiating this, we have $\frac{d}{d\tau}\left[v_p - v_r\right] = -\frac{c_p a'(\tau) q'(a(\tau))}{\alpha(1+\delta)q(a(\tau))^2}$. For sufficiently small $k$ and $\delta$, we showed earlier that $a'(\tau) < 0$. Since all the other factors are positive, it follows that $\frac{d}{d\tau}\left[v_p - v_r\right] > 0$ for sufficiently small $\kappa$ and $\delta$.

For the equilibrium ransom charged, that is (C.24). Viewing $a$ as a function of $\tau$ and differentiating (C.24) with respect to $\tau$, we have that $\frac{dR^*}{d\tau} = -\frac{c_p(1-\delta)a'(\tau)q'(a(\tau))}{(1+\delta)q(a(\tau))^2}$. For sufficiently small $k$ and $\delta$, we showed earlier that $a'(\tau) < 0$. Since all the other factors are positive, it follows that $\frac{dR^*}{d\tau} > 0$ for sufficiently small $\kappa$ and $\delta$.

Lastly, the expected total ransom paid is $T = q(a^*)R^*r(\sigma^*)$. Using $v_p - v_r = \frac{c_p(1-\delta)-\tilde{\rho}q(a)}{\alpha(1-\delta^2)q(a)}$ from earlier for $r(\sigma^*)$ and (C.24) for the equilibrium ransom charged, we can express $T$ as

$$T = \frac{(c_p(1-\delta) - \tilde{\rho}q(a))(\delta\tilde{\rho}q(a) + c_p(1-\delta))}{\alpha(1-\delta)(1+\delta)^2 q(a)}. \tag{C.44}$$

Viewing $a$ as a function of $\tau$ and differentiating this with respect to $\tau$, we have

$$\frac{dT}{d\tau} = -\frac{a'(\tau)q'(a(\tau))\left(\delta\tilde{\rho}^2 q(a(\tau))^2 + c_p^2(1-\delta)^2\right)}{\alpha(1-\delta)(1+\delta)^2 q(a(\tau))^2}. \tag{C.45}$$

For sufficiently small $k$ and $\delta$, we showed earlier that $a'(\tau) < 0$. Since all the other factors are positive, it follows that $\frac{dT}{d\tau} > 0$ for sufficiently small $\kappa$ and $\delta$. ■

**Proof of Proposition 3:** Under the benchmark scenario, the attacker's profit function is given in (C.20). Viewing $a$ as a function of $\tau$ and differentiating with respect to $\tau$, the derivative of the

profit function is given below:

$$\frac{d}{d\tau}\left[\Pi_{BM}\right] = \frac{\rho a'(\tau)\left(-((\alpha q(a(\tau))-1)(\alpha(\kappa+c_p)q(a(\tau))-c_p))-\alpha\kappa a(\tau)q'(a(\tau))\right)}{\alpha a(\tau)^2(\alpha q(a(\tau))-1)^2} - (1-\theta) \quad \text{(C.46)}$$

From (C.21), we can differentiate this with respect to $\tau$ to find $a'(\tau)$. For sufficiently small $k$ (i.e., $\kappa$ close enough to 0), this can be written as

$$a'(\tau) = -\frac{\alpha a(\tau)^3}{c_p\rho + \alpha\tau a(\tau)^2} + O(\kappa). \quad \text{(C.47)}$$

Substituting this back into (C.46), for sufficiently small $\kappa$, we have

$$\frac{d}{d\tau}\left[\Pi_{BM}\right] = \frac{c_p\rho a(\tau)}{c_p\rho + \alpha\tau a(\tau)^2} - (1-\theta) + O(\kappa). \quad \text{(C.48)}$$

For $\theta$ close to 1, this is positive for sufficiently small $\kappa$. On the other hand, for $\theta$ close to the marginal type $\underline{\theta} = 1 - a(\tau)$, (C.48) becomes $-\frac{\alpha\tau a(\tau)^3}{c_p\rho+\alpha\tau a(\tau)^2} + O(\kappa)$, which is negative for sufficiently small $\kappa$. Altogether, for the benchmark scenario, this shows that the higher skilled attacker have profits that increase in $\tau$, whereas the less-skilled attacker who enter the market have equilibrium profits that decrease in $\tau$.

The analysis above can be done for the ransomware scenario, substituting in (C.6) into (C.3) (using $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$) and taking the derivative with respect to $\tau$. The algebra is omitted for brevity. ■

**Proof of Proposition 4:** For the benchmark case, welfare is $\int_{v_u}^{v_p}(v-\kappa-q(a^*)(\alpha v))\,dv + \int_{v_p}^{1}(v-\kappa-c_p)\,dv$, where $a^*$ here is the attacker population for the benchmark case. Substituting in (C.22) for $v_p$ and (C.23) for $v_u$, the welfare can then be expressed as

$$W_{BM} = \frac{1}{2}\left(\frac{c_p^2}{\alpha q(a^*)} + \kappa\left(\frac{\kappa}{1-\alpha q(a^*)} - 2\right) - 2c_p + 1\right). \quad \text{(C.49)}$$

Viewing the attacker population size $a^*$ as a function of $\tau$ (recall again that $a^*$ solves (C.21)) and differentiating (C.49) with respect to $\tau$, we have that

$$\frac{d}{d\tau}\left[W_{BM}\right] = \frac{q'(a^*)\left(\frac{\alpha^2\kappa^2}{(\alpha q(a^*)-1)^2} - \frac{c_p^2}{q(a^*)^2}\right)}{2\alpha}. \quad \text{(C.50)}$$

For sufficiently small $\kappa$, this can be written as

$$\frac{d}{d\tau}\left[W_{BM}\right] = -\frac{c_p^2 q'(a^*)}{2\alpha\left(q(a^*)\right)^2} + O(\kappa). \tag{C.51}$$

Since the zero-order term from (C.51) is negative, it follows that $\frac{d}{d\tau}\left[W_{BM}\right] < 0$ for sufficiently small $\kappa$.

For the ransomware case, first consider $\tau > \hat{\tau}$, where $\hat{\tau}$ is the $\tau$ boundary when the attacker market segmentation changes between two segments of attackers and one segment of attackers. Whether there is one segment or two segments of attackers in equilibrium, the welfare expression looks the same:

$$W_{RW} = \int_{v_u}^{v_r} (v - \kappa - q(a^*)(\alpha v))\, dv + \int_{v_r}^{v_p} (v - \kappa - q(a^*)(R^* + \delta\alpha v))\, dv + \int_{v_p}^{1} (v - \kappa - c_p)\, dv, \tag{C.52}$$

where $a^*$ is interpreted as the total size of the attacker population under the ransomware case.

When one attacker segment arises in equilibrium (i.e., for $\tau > \hat{\tau}$), then we can substitute (C.24) for $R$, (C.26) for $v_p$, (C.28) for $v_r$, and (C.27) for $v_n$ into (C.52), resulting in

$$W_{RW} = \frac{1}{2}\left(\frac{c_p^2(\delta-1)(3\delta+1) - \delta\tilde{\rho}^2 q(a^*)^2}{\alpha(\delta-1)(\delta+1)^2 q(a^*)} + \kappa\left(\frac{\kappa}{1-\alpha q(a^*)} - 2\right) + c_p\left(-\frac{2\delta\tilde{\rho}}{\alpha(\delta+1)^2} - 2\right) + 1\right). \tag{C.53}$$

Viewing $a^*$ as a function of $\tau$, differentiating with respect to $\tau$, and looking at the asymptotic expansion around $\kappa = 0$ and $\delta = 0$, we have that

$$\frac{d}{\tau}\left[W_{RW}\right] = -\frac{c_p^2 a_0'(\tau) q'(a_0(\tau))}{2\alpha q(a_0(\tau))^2} + O(\kappa + \delta), \tag{C.54}$$

where $a_0(\tau)$ comes from the proof of Proposition 1. Since $a_0'(\tau) < 0$, it follows that $\frac{d}{d\tau}\left[W_{RW}\right] > 0$ for sufficiently small $\kappa$ and $\delta$ when $\tau > \hat{\tau}$.

On the other hand, consider $\tau < \hat{\tau}$. We will show that for low enough $\tau$, (i.e., $\tau$ close to $\underline{\tau}$), welfare increases in $\tau$, while when $\tau$ is close to $\hat{\tau}$, then welfare can actually decrease in $\tau$. In both of these cases, the expression for welfare when there are both segments of attackers is given by:

$$W_{RW} = \int_{v_p}^{1} (v - c_p - \kappa)\, dv + \int_{v_r}^{v_p} \left(v - q(a_1^* + a_2^*)\left(\frac{a_2^*}{a_1^* + a_2^*}(R^* + \alpha\delta v) + \frac{a_1^*}{a_1^* + a_2^*}(\alpha v)\right) - \kappa\right) dv + \int_{v_u}^{v_r} (v - q(a_1^* + a_2^*)(\alpha v) - \kappa)\, dv, \tag{C.55}$$

where $a_1^*$ and $a_2^*$ solve (C.30) and (C.31), $R^*$ comes from (C.29), and $v_p$, $v_r$, and $v_u$ come from (C.32), (C.34), and (C.33), respectively. Substituting these expressions in, we have that welfare in this case can be written as:

$$W_{RW} = \Bigg(q(a_1^*+a_2^*)\Bigg((\delta-1)\Bigg(4(a_1^*)^2\alpha\Bigg(c_p(c_p+2)-(\kappa-1)^2\Bigg)+a_1^*a_2^*\Bigg(\alpha\Bigg(c_p^2(3\delta+5)+8c_p(\delta+1)-4(\delta+1)(\kappa-1)^2\Bigg)+$$
$$2c_p\tilde{\rho}\Bigg)+(a_2^*)^2\Bigg(\alpha\Bigg(c_p^2(3\delta+1)+2c_p(\delta+1)^2-(\delta+1)^2(\kappa-1)^2\Bigg)+2c_p\delta\tilde{\rho}\Bigg)\Bigg)+q(a_1^*+a_2^*)\Bigg(-\alpha^2(\delta-1)$$
$$(2c_p+2\kappa-1)(2a_1^*+a_2^*\delta+a_2^*)^2-2a_2^*\alpha c_p(\delta-1)\tilde{\rho}(a_1^*+a_2^*\delta)+a_2^*\tilde{\rho}^2(a_1^*+a_2^*\delta)-a_2^*\alpha\tilde{\rho}^2(a_1^*+a_2^*\delta)q(a_1^*+a_2^*)\Bigg)\Bigg)-$$
$$c_p^2(\delta-1)(a_1^*+a_2^*)(4a_1^*+3a_2^*\delta+a_2^*)\Bigg)\Bigg(2\alpha(\delta-1)(2a_1^*+a_2^*\delta+a_2^*)^2q(a_1^*+a_2^*)(\alpha q(a_1^*+a_2^*)-1)\Bigg)^{-1}. \tag{C.56}$$

Viewing $a_1^*$ and $a_2^*$ as functions of $\tau$, differentiating, substituting in $\tau=\underline{\tau}$, $a_1(\underline{\tau})=q^{-1}\left(\frac{c_p(1-\delta)}{\tilde{\rho}}\right)$ and $a_2(\underline{\tau})=0$, and getting the asymptotic expansion of this expression around $\kappa=0$ and $\delta=0$, we have that

$$\frac{d}{d\tau}\left[W_{RW}\right]|_{\tau=\underline{\tau}} = -\frac{\tilde{\rho}^2q'\left(q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)\right)\left(a_1'\left(\frac{c_p\tilde{\rho}}{\alpha q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^2}\right)+a_2'\left(\frac{c_p\tilde{\rho}}{\alpha q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^2}\right)\right)}{2\alpha}+O(\kappa+\delta). \tag{C.57}$$

To find $a_1'\left(\frac{c_p\tilde{\rho}}{\alpha q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^2}\right)$ and $a_2'\left(\frac{c_p\tilde{\rho}}{\alpha q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^2}\right)$, note again that $a_1^*$ and $a_2^*$ together solve (C.30) and (C.31). Implicitly differentiating that system of equations, using $a_1(\underline{\tau})=q^{-1}\left(\frac{c_p(1-\delta)}{\tilde{\rho}}\right)$ and $a_2(\underline{\tau})=0$, $\tau=\underline{\tau}$, and taking the asymptotic expansion of the expressions around $\kappa=0$ and $\delta=0$, we have

$$a_1'\left(\frac{c_p\tilde{\rho}}{\alpha q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^2}\right) = -\frac{\alpha q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^3}{2c_p\tilde{\rho}}+O(\kappa+\delta) \tag{C.58}$$

and

$$a_2'\left(\frac{c_p\tilde{\rho}}{\alpha q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^2}\right) = O(\kappa+\delta). \tag{C.59}$$

Substituting (C.58) and (C.59) into (C.57), we have

$$\frac{d}{d\tau}\left[W_{RW}\right]|_{\tau=\underline{\tau}} = \frac{\tilde{\rho}q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)^3 q'\left(q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)\right)}{4c_p} + O(\kappa+\delta). \tag{C.60}$$

This is positive, so for sufficiently small $\kappa$ and $\delta$, $\frac{d}{d\tau}\left[W_{RW}\right]|_{\tau=\underline{\tau}} > 0$. By continuity, there is some right-neighborhood around $\underline{\tau}$ for which welfare increases in $\tau$.

Lastly, for $\tau$ close to $\hat{\tau} = \frac{c_p\tilde{\rho}}{\alpha\left(\omega-\sqrt{\omega(\omega+1)}+1\right)q^{-1}\left(\frac{c_p\omega-c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)^2} + O(\kappa+\delta)$, when both segments of attackers are in the market, we will show that welfare can be decreasing in $\tau$. Equations (C.55) and (C.56) still apply here, since we are still in the market structure with two attacker segments.

Viewing $a_1^*$ and $a_2^*$ as functions of $\tau$, differentiating, substituting in $\tau = \hat{\tau}$, $a_1(\hat{\tau}) = 0$ and $a_2(\hat{\tau}) = q^{-1}\left(\frac{c_p(1+\omega-\sqrt{\omega(1+\omega)})}{\tilde{\rho}(1+\omega)}\right)$, and getting the asymptotic expansion of this expression around $\kappa = 0$ and $\delta = 0$, we have that

$$\begin{aligned}\frac{d}{d\tau}\left[W_{RW}\right]|_{\tau=\hat{\tau}} = \Bigg(\tilde{\rho}\Bigg(&\frac{c_p\omega\left(\omega-\sqrt{\omega(\omega+1)}+1\right)a_1'(\hat{\tau})}{q^{-1}\left(\frac{c_p\omega-c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)} - \\ &\tilde{\rho}(\omega+1)^2 q'\left(q^{-1}\left(\frac{c_p\omega-c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)\Big(a_1'(\hat{\tau})+ \\ &a_2'(\hat{\tau})\Big)\Bigg)\Bigg)\left(2\alpha\left(\omega-\sqrt{\omega(\omega+1)}+1\right)^2\right)^{-1} + O(\kappa+\delta).\end{aligned} \tag{C.61}$$

To find $a_1'(\hat{\tau})$ and $a_2'(\hat{\tau})$, note again that $a_1^*$ and $a_2^*$ together solve (C.30) and (C.31). Implicitly differentiating that system of equations, using $a_1(\hat{\tau}) = 0$ and $a_2(\hat{\tau}) =$, $\tau = \hat{\tau}$, and taking the

asymptotic expansion of these derivative expressions around $\kappa = 0$ and $\delta = 0$, we have

$$a_1'(\hat{\tau}) = \left(2\alpha(\omega+1)^2\left(\sqrt{\omega(\omega+1)}-\omega\right)q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)^4 q'\left(q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)\right)\left(c_p\omega\left(\tilde{\rho}\left(-3\omega+\sqrt{\omega(\omega+1)}-3\right)q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)q'\left(q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)+ 2c_p\left(2\omega-2\sqrt{\omega(\omega+1)}+1\right)\right)\right)^{-1} + O(\kappa+\delta) \quad \text{(C.62)}$$

and

$$a_2'(\hat{\tau}) = \left(\alpha\left(-\omega+\sqrt{\omega(\omega+1)}-1\right)q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)^3 \left(2\tilde{\rho}\sqrt{\omega(\omega+1)^3}q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)q'\left(q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)+ c_p\omega\left(2\omega-2\sqrt{\omega(\omega+1)}+1\right)\right)\right)\left(c_p\tilde{\rho}\omega\left(\tilde{\rho}\left(-3\omega+\sqrt{\omega(\omega+1)}-3\right)q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right) q'\left(q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)+2c_p\left(2\omega-2\sqrt{\omega(\omega+1)}+1\right)\right)\right)^{-1} + O(\kappa+\delta). \quad \text{(C.63)}$$

Substituting (C.62) and (C.63) into (C.61) and simplifying, we have

$$\frac{d}{d\tau}\left[W_{RW}\right]|_{\tau=\hat{\tau}} = \left(\tilde{\rho}(\omega+1)^2\left(-\sqrt{\omega^3(\omega+1)}+\omega^2+\omega\right)q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)^3 \times q'\left(q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)\right)\left(2\omega\left(\omega-\sqrt{\omega(\omega+1)}+1\right)^2 \times \left(\tilde{\rho}\left(-3\omega+\sqrt{\omega(\omega+1)}-3\right)q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)q'\left(q^{-1}\left(\frac{c_p\omega - c_p\sqrt{\omega(\omega+1)}+c_p}{\tilde{\rho}\omega+\tilde{\rho}}\right)\right)+ 2c_p\left(2\omega-2\sqrt{\omega(\omega+1)}+1\right)\right)\right)^{-1} + O(\kappa+\delta). \quad \text{(C.64)}$$

The zero-order term of this asymptotic expression being negative is equivalent to (C.43). Following the same argument as in the end of the proof of Proposition 1, it follows that if $\omega > \hat{\omega}$, then $\frac{d}{d\tau}\left[W_{RW}\right]|_{\tau=\hat{\tau}} < 0$ for sufficiently small $\kappa$ and $\delta$. Consequently, for sufficiently small $\kappa$ and $\delta$, welfare decreases in a left-neighborhood of $\tau = \hat{\tau}$ by continuity of $\frac{d}{d\tau}\left[W_{RW}\right]$. ■

**Proof of Proposition 5:** The expression for welfare in the benchmark case is given in (C.49). The equation defining $a^*$ is given by (C.21). For ease of reference, these two expressions are provided below:

$$a\tau = \frac{c_p\rho}{a\alpha} - \frac{\kappa\rho q(a)}{a - a\alpha q(a)} \text{ and} \tag{C.65}$$

$$W_{BM} = \frac{1}{2}\left(\frac{c_p^2}{\alpha q(a^*)} + \kappa\left(\frac{\kappa}{1-\alpha q(a^*)} - 2\right) - 2c_p + 1\right). \tag{C.66}$$

For sufficiently small $\kappa$, by taking a Taylor series expansion around $\kappa = 0$ in (C.65) and solving for $a$, the asymptotic expression for $a^*_{BM}$ is given by

$$a^*_{BM} = \sqrt{\frac{c_p\rho}{\alpha\tau}} + O(\kappa). \tag{C.67}$$

Substituting this into (C.66) and taking a Taylor series expansion around $\kappa = 0$ again, the asymptotic expression for $W_{BM}$ is given by

$$W_{BM} = \frac{1}{2}\left(1 - 2c_p + \frac{c_p^2}{\alpha q\left(\sqrt{\frac{c_p\rho}{\alpha\tau}}\right)}\right) + O(\kappa). \tag{C.68}$$

On the other hand, for the ransomware case when $\tau < \hat{\tau}$ (i.e., when there are two segments of attackers), then the welfare expression is given in (C.56). For ease of reference, this is provided

below:

$$W_{RW} = \Bigg( q(a_1^*+a_2^*)\Bigg( (\delta-1)\Bigg( 4(a_1^*)^2\alpha\Bigg( c_p(c_p+2)-(\kappa-1)^2\Bigg)+a_1^*a_2^*\Bigg(\alpha\Bigg( c_p^2(3\delta+5)+8c_p(\delta+1)-4(\delta+1)(\kappa-1)^2\Bigg)+ 2c_p\tilde{\rho}\Bigg)+(a_2^*)^2\Bigg(\alpha\Bigg( c_p^2(3\delta+1)+2c_p(\delta+1)^2-(\delta+1)^2(\kappa-1)^2\Bigg)+2c_p\delta\tilde{\rho}\Bigg)\Bigg)+q(a_1^*+a_2^*)\Bigg( -\alpha^2(\delta-1) (2c_p+2\kappa-1)(2a_1^*+a_2^*\delta+a_2^*)^2-2a_2^*\alpha c_p(\delta-1)\tilde{\rho}(a_1^*+a_2^*\delta)+a_2^*\tilde{\rho}^2(a_1^*+a_2^*\delta)-a_2^*\alpha\tilde{\rho}^2(a_1^*+a_2^*\delta)q(a_1^*+a_2^*)\Bigg)\Bigg)- c_p^2(\delta-1)(a_1^*+a_2^*)(4a_1^*+3a_2^*\delta+a_2^*)\Bigg)\Bigg(2\alpha(\delta-1)(2a_1^*+a_2^*\delta+a_2^*)^2q(a_1^*+a_2^*)(\alpha q(a_1^*+a_2^*)-1)\Bigg)^{-1}. \tag{C.69}$$

Recalling at the lower $\tau$ boundary $\underline{\tau}$ that we have $a_2^* = 0$ and $a_1^* = q^{-1}\left(\frac{c_p(1-\delta)}{\rho}\right)$. Substituting these two into (C.69) and taking an asymptotic expansion of welfare for sufficiently small $\kappa$ and $\delta$, we have

$$W_{RW}|_{\tau=\underline{\tau}} = \frac{1}{2}\left( c_p\left(\frac{\tilde{\rho}}{\alpha}-2\right)+1\right). \tag{C.70}$$

Recall again that $\underline{\tau}$ is given as

$$\underline{\tau} = \frac{c_p\tilde{\rho}(\alpha(\delta-1)(c_p+\kappa)+\tilde{\rho})}{\alpha(\alpha c_p(\delta-1)+\tilde{\rho})q^{-1}\left(\frac{c_p-c_p\delta}{\tilde{\rho}}\right)^2} \tag{C.71}$$

Substituting this into (C.68) and taking an asymptotic expansion around $\kappa = 0$, we have

$$W_{BM}|_{\tau=\underline{\tau}} = \frac{1}{2}\left(1-2c_p+\frac{c_p^2}{\alpha q\left(\sqrt{\frac{\rho}{\tilde{\rho}}}q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)\right)}\right)+O(\kappa). \tag{C.72}$$

Comparing this to (C.70), we see that for sufficiently low $\kappa$ and $\delta$, $W_{RW}|_{\tau=\underline{\tau}} > W_{BM}|_{\tau=\underline{\tau}}$ if and only if $\tilde{\rho} < \rho$. By continuity of welfare in $\tau$, if $\tilde{\rho} < \rho$, then $W_{RW} > W_{BM}$ holds in some right-neighborhood of $\underline{\tau}$ for sufficiently small $\kappa$ and $\delta$.

On the other side of the $\tau$ range, namely when $\tau = \bar{\tau}$, the equilibrium outcome under ransomware has all attackers who enter the market conducting ransomware. In this case, the welfare expression

is given by (C.53), provided here for easy reference:

$$W_{RW} = \frac{1}{2}\left(\frac{c_p^2(\delta-1)(3\delta+1) - \delta\tilde{\rho}^2 q(a^*)^2}{\alpha(\delta-1)(\delta+1)^2 q(a^*)} + \kappa\left(\frac{\kappa}{1-\alpha q(a^*)} - 2\right) + c_p\left(-\frac{2\delta\tilde{\rho}}{\alpha(\delta+1)^2} - 2\right) + 1\right). \tag{C.73}$$

Taking asymptotics for sufficiently small $\kappa$ and $\delta$, we can write the asymptotic expansion as

$$W_{RW} = \frac{1}{2}\left(1 - 2c_p + \frac{c_p^2}{\alpha q(a)}\right) + O(\kappa+\delta). \tag{C.74}$$

Recall again that for the benchmark case, the asymptotic welfare expression is given by (C.68). To find the $\tau$ at which these welfare expressions cross, we can equate their zero-order terms:

$$\frac{1}{2}\left(1 - 2c_p + \frac{c_p^2}{\alpha q(a)}\right) = \frac{1}{2}\left(1 - 2c_p + \frac{c_p^2}{\alpha q\left(\sqrt{\frac{c_p\rho}{\alpha\tau}}\right)}\right). \tag{C.75}$$

This is equivalent to $a = \sqrt{\frac{c_p\rho}{\alpha\tau}}$, where $a$ is for attacker population size when all attackers are in the market are ransomware attackers, which is given in (C.25). For sufficiently small $\kappa$ and $\delta$, the solution of (C.25) converges to the solution of

$$a\tau(1+\omega) = \frac{c_p^2}{a\alpha q(a)}. \tag{C.76}$$

Then the $\tau$ for which $a = \sqrt{\frac{c_p\rho}{\alpha\tau}}$ holds can be found by substituting in $a = \sqrt{\frac{c_p\rho}{\alpha\tau}}$ into (C.76). This is equivalent to solving in $\tau$ the equation:

$$\tau\sqrt{\frac{c_p\rho}{\alpha\tau}}\left(-\frac{c_p}{\rho q\left(\sqrt{\frac{c_p\rho}{\alpha\tau}}\right)} + \omega + 1\right) = 0. \tag{C.77}$$

This has a unique solution in $\tau$ for $\tau > 0$ given by $\tau = \frac{c_p\rho}{\alpha\left(q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)\right)^2}$. Then the $\tau$ at which $W_{BM}$ and $W_{RW}$ cross is given by $\hat{\tau} = \frac{c_p\rho}{\alpha\left(q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)\right)^2} + O(\kappa+\delta)$.

To find a condition when $W_{BM} > W_{RW}$ by $\tau = \bar{\tau}$ (the upper bound of the focal range of $\tau$), we just need $\hat{\tau} < \bar{\tau}$. Recall again that $\bar{\tau} = \frac{c_p(\alpha+\tilde{\rho})}{2\alpha(\omega+1)q^{-1}\left(\frac{2c_p}{\alpha+\tilde{\rho}}\right)^2} + O(\kappa+\delta)$.

Comparing the zero-order terms, we want to find a condition so that:

$$\frac{c_p\rho}{\alpha q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)^2} < \frac{c_p(\alpha+\tilde{\rho})}{2\alpha(\omega+1)q^{-1}\left(\frac{2c_p}{\alpha+\tilde{\rho}}\right)^2}. \tag{C.78}$$

Rearranging, this becomes

$$\frac{\rho(\omega+1)}{c_p q^{-1}\left(\frac{c_p}{\rho\omega+\rho}\right)^2} < \frac{\alpha+\tilde{\rho}}{2c_p q^{-1}\left(\frac{2c_p}{\alpha+\tilde{\rho}}\right)^2}. \tag{C.79}$$

Consider the function $b(x) = \frac{\left(\frac{1}{x}\right)}{(q^{-1}(x))^2}$. The derivative of this function is $b'(x) = -\frac{q^{-1}(x)+\frac{2x}{q'(q^{-1}(x))}}{x^2 q^{-1}(x)^3}$, which is negative for all $x > 0$. So then for (C.79) to hold, a necessary and sufficient condition is $\frac{c_p}{\rho(1+\omega)} > \frac{2c_p}{\alpha+\tilde{\rho}}$. Rearranging, this means that for sufficiently small $\kappa$ and $\delta$, $W_{BM} > W_{RW}$ at $\tau = \bar{\tau}$ if and only if $\tilde{\rho} > -\alpha + 2\rho(1+\omega)$.

Comparing the two bounds on $\tilde{\rho}$, note that $\rho > -\alpha + 2\rho(1+\omega)$ if $\alpha > \rho$ and $\omega < \frac{\alpha-\rho}{2\rho}$.

Suppose that $\alpha > \rho$ and $\omega < \frac{\alpha-\rho}{2\rho}$ so that $\rho > -\alpha + 2\rho(1+\omega)$. Given that for sufficiently small $\kappa$ and $\delta$, $W_{BM} > W_{RW}$ at $\tau = \bar{\tau}$ if and only if $\tilde{\rho} > -\alpha + 2\rho(1+\omega)$ and $W_{RW} > W_{BM}$ holds in some right-neighborhood of $\underline{\tau}$ for sufficiently small $\kappa$ and $\delta$, this then splits into three sub-cases. If $\tilde{\rho} < -\alpha + 2\rho(1+\omega)$, then $W_{RW} > W_{BM}$ at both the upper and lower $\tau$ bounds. In an intermediate range of $\tilde{\rho}$ (i.e., $-\alpha + 2\rho(1+\omega) < \tilde{\rho} < \rho$), we have that $W_{RW} > W_{BM}$ at the lower $\tau$ bound, but $W_{RW} < W_{BM}$ at the upper $\tau$ bound. When $\tilde{\rho} > \rho$, then $W_{RW} < W_{BM}$ at both $\tau$ bounds.

On the other hand, suppose $\omega > \frac{\alpha-\rho}{2\rho}$. Then $\rho < -\alpha + 2\rho(1+\omega)$. Note that if $\omega > \frac{\alpha-\rho}{2\rho}$ holds under the focal region, then since $\omega < \frac{(\alpha-\rho)^2}{4\alpha\tilde{\rho}}$, it follows that $\frac{(\alpha-\rho)^2}{4\alpha\tilde{\rho}} > \frac{\alpha-\rho}{2\rho}$ holds. This is equivalent to $\rho > \frac{2\alpha^2\tilde{\rho}}{\alpha^2+\tilde{\rho}^2}$. This implies that $\tilde{\rho} < \rho$ since otherwise, $\tilde{\rho} > \frac{2\alpha^2\tilde{\rho}}{\alpha^2+\tilde{\rho}^2}$ would hold, which is equivalent to $\tilde{\rho} > \alpha$ (violating a focal region condition). Consequently, if $\omega > \frac{\alpha-\rho}{2\rho}$ holds in the focal region, then $\tilde{\rho} < \rho$ holds as well, in which case $W_{RW} > W_{BM}$ at both $\tau$ bounds of the focal region. ■

**Proof of Proposition 6:** The expression for welfare in the benchmark case is given in (C.49). The equation defining $a^*$ is given by (C.21). For ease of reference, these two expressions are provided below:

$$a\tau = \frac{c_p\rho}{a\alpha} - \frac{\kappa\rho q(a)}{a - a\alpha q(a)} \text{ and} \tag{C.80}$$

$$W_{BM} = \frac{1}{2}\left(\frac{c_p^2}{\alpha q(a^*)} + \kappa\left(\frac{\kappa}{1-\alpha q(a^*)} - 2\right) - 2c_p + 1\right). \tag{C.81}$$

Differentiating (C.81) with respect to $\alpha$, we have that

$$\frac{d}{d\alpha}\left[W_{BM}\right] = -\frac{(\alpha a'(\alpha)q'(a(\alpha)) + q(a(\alpha)))\left(\alpha q(a(\alpha))\left(\alpha(c_p-\kappa)(c_p+\kappa)q(a(\alpha)) - 2c_p^2\right) + c_p^2\right)}{2\alpha^2 q(a(\alpha))^2(\alpha q(a(\alpha)) - 1)^2}. \tag{C.82}$$

Taking a limit as $\kappa \to 0$, we have that the asymptotic expression for this for sufficiently small $\kappa$ is given by

$$\frac{d}{d\alpha}\left[W_{BM}\right] = \frac{c_p^2\left(\alpha a'(\alpha)|_{\kappa=0} q'(a(\alpha)|_{\kappa=0}) + q(a(\alpha)|_{\kappa=0})\right)}{2\alpha^2 q(a(\alpha)|_{\kappa=0})^2} + O(\kappa). \tag{C.83}$$

To simplify notation, we will drop the subscript with $\kappa = 0$, but note that $a(\alpha)$ and $a'(\alpha)$ in the next few paragraphs refer to what these measures look like when $\kappa = 0$. Showing that (C.83) is negative for sufficiently small $\kappa$ is equivalent to showing that

$$\alpha a'(\alpha)q'(a(\alpha)) + q(a(\alpha)) > 0. \tag{C.84}$$

Viewing $a$ as a function of $\kappa$ in (C.80), taking the derivative with respect to $\kappa$, and taking the limit as $\kappa \to 0$, we have that

$$a'(\alpha) = -\frac{c_p \rho a(\alpha)}{\alpha^2 \tau a(\alpha)^2 + \alpha c_p \rho} + O(\kappa). \tag{C.85}$$

Substituting this into (C.84), we want to show:

$$q'(a(\alpha)) < \frac{q(a(\alpha))}{a(\alpha)} + \frac{\alpha\tau a(\alpha) q(a(\alpha))}{c_p \rho}. \tag{C.86}$$

Note that the second term on the right-hand side is positive while $q'(a(\alpha)) < \frac{q(a(\alpha))}{a(\alpha)}$ can be written as $q'(a(\alpha)) < \frac{q(a(\alpha))-q(0)}{a(\alpha)-0}$. This inequality holds since $q(a)$ is increasing, concave with $q(0) = 0$. Altogether, $\frac{d}{d\alpha}\left[W_{BM}\right] < 0$ for sufficiently small $\kappa$.

Next, we show that for $\tau > \hat{\tau}$ in the focal region, welfare decreases in $\alpha$ as well. In this range of $\tau$, the equilibrium outcome under ransomware has all attackers who enter the market conducting ransomware. In this case, the welfare expression is given by (C.53), provided here for easy reference:

$$W_{RW} = \frac{1}{2}\left(\frac{c_p^2(\delta-1)(3\delta+1) - \delta\tilde{\rho}^2 q(a^*)^2}{\alpha(\delta-1)(\delta+1)^2 q(a^*)} + \kappa\left(\frac{\kappa}{1-\alpha q(a^*)} - 2\right) + c_p\left(-\frac{2\delta\tilde{\rho}}{\alpha(\delta+1)^2} - 2\right) + 1\right). \tag{C.87}$$

Taking asymptotics for sufficiently small $\kappa$ and $\delta$, we can write the asymptotic expansion as

$$W_{RW} = \frac{1}{2}\left(1 - 2c_p + \frac{c_p^2}{\alpha q(a)}\right) + O(\kappa + \delta). \tag{C.88}$$

Viewing $a$ as a function of $\alpha$ in (C.88), differentiating it with respect to $\alpha$, and taking asymptotics in $\kappa$ and $\delta$, we have that

$$\frac{d}{d\alpha}\left[W_{RW}\right] = -\frac{c_p^2\left(\alpha a'(\alpha)q'(a(\alpha)) + q(a(\alpha))\right)}{2\alpha^2 q(a(\alpha))^2} + O(\kappa + \delta). \tag{C.89}$$

To show that this is negative for sufficiently small $\kappa$ and $\delta$ is equivalent to showing that

$$q(a(\alpha)) + \alpha a'(\alpha)q'(a(\alpha)) > 0, \tag{C.90}$$

where $a(\alpha)$ and $a'(\alpha)$ in the next few paragraphs refer to what these measures look like when $\kappa = 0$ and $\delta = 0$.

Recall again that the attacker population size for ransomware in this case (i.e., when all attackers in the market are ransomware attackers), which is given in (C.25). For sufficiently small $\kappa$ and $\delta$, the solution of (C.25) converges to the solution of

$$a\tau(1+\omega) = \frac{c_p^2}{a\alpha q(a)}. \tag{C.91}$$

Implicitly differentiating $a$ with respect to $\alpha$ in (C.91) and taking asymptotics in $\kappa$ and $\delta$,

$$a'(\alpha) = -\frac{c_p^2 a(\alpha)q(a(\alpha))}{\alpha c_p^2 a(\alpha)q'(a(\alpha)) + \alpha q(a(\alpha))\left(\alpha\tau(\omega+1)a(\alpha)^2 q(a(\alpha)) + c_p^2\right)} + O(\kappa + \delta). \tag{C.92}$$

Substituting this into (C.90), we want to show that the following holds:

$$c_p^2 a(\alpha)q'(a(\alpha)) + q(a(\alpha))\left(\alpha\tau(\omega+1)a(\alpha)^2 q(a(\alpha)) + c_p^2\right) > 0. \tag{C.93}$$

This is true. Therefore, $\frac{d}{d\alpha}\left[W_{RW}\right] < 0$ for sufficiently small $\kappa$ and $\delta$.

Lastly, we will show that $\frac{d}{d\delta}\left[W_{RW}\right] > 0$ for sufficiently small $\kappa$ and $\delta$ when all attackers who enter the market opt to conduct ransomware. The asymptotic expression for welfare is given in

(C.88). Implicitly differentiating $a$ with respect to $\delta$ and taking asymptotics in $\kappa$ and $\delta$, we have

$$\frac{d}{d\delta}\left[W_{RW}\right] = -\frac{c_p^2 a'(\delta) q'(a(\delta))}{2\alpha q(a(\delta))^2} + O(\kappa+\delta). \tag{C.94}$$

Then to show that welfare is increasing in this asymptotic regime, it suffices to show that $a'(\delta) < 0$ for sufficiently small $\kappa$ and $\delta$. Recall again that the equation defining $a$ in this case is given by (C.25). For ease of reference, we repeat it here:

$$a\tau(1+\omega) = \frac{a(\delta\tilde{\rho}q(a) + c_p(1-\delta))^2}{\alpha(1-\delta)(a\delta+a)^2 q(a)} - \frac{\kappa\tilde{\rho}q(a)}{a - a\alpha q(a)}. \tag{C.95}$$

Implicitly differentiating this with respect to $\delta$ and solving for $a'(\delta)$, we have

$$\begin{aligned} a'(\delta) = &\left((\delta\tilde{\rho}q(a(\delta)) + c_p(-\delta) + c_p)(((\delta-1)\delta+2)\tilde{\rho}q(a(\delta)) - c_p(\delta-3)(\delta-1))\right)\times \\ &\left((\delta-1)^2(\delta+1)^3\left(\frac{q'(a(\delta))\left(\tilde{\rho}q(a(\delta))^2\left(\frac{\alpha\kappa}{(\alpha q(a(\delta))-1)^2} + \frac{\delta^2\tilde{\rho}}{(\delta-1)(\delta+1)^2}\right) - \frac{c_p^2(\delta-1)}{(\delta+1)^2}\right)}{q(a(\delta))} + \right.\right. \\ &\left(q(a(\delta))\left(\tilde{\rho}q(a(\delta))\left(-\alpha\delta^2\tilde{\rho}q(a(\delta)) + \alpha(\delta-1)\left(2c_p\delta + (\delta+1)^2\kappa\right) + \delta^2\tilde{\rho}\right) - c_p(\delta-1)(\alpha c_p(\delta-1) + 2\delta\tilde{\rho})\right) + \right. \\ &\left.\left.\left.c_p^2(\delta-1)^2\right)\left((\delta-1)(\delta+1)^2 a(\delta)(\alpha q(a(\delta)) - 1)\right)^{-1} + \alpha\tau(\omega+1)a(\delta)q(a(\delta))\right)\right)^{-1}. \end{aligned} \tag{C.96}$$

Taking the limit as $\kappa$ and $\delta$ approach 0, the asymptotic expression for this is

$$a'(\delta) = \frac{a(\delta)|_{\delta=0} c_p q(a(\delta)|_{\delta=0})(2\tilde{\rho}q(a(\delta)|_{\delta=0}) - 3c_p)}{a(\delta)|_{\delta=0} c_p^2 q'(a(\delta)|_{\delta=0}) + q(a(\delta)|_{\delta=0})\left(a(\delta)|_{\delta=0}^2 \alpha\tau(\omega+1)q(a(\delta)|_{\delta=0}) + c_p^2\right)} + O(\kappa+\delta) \tag{C.97}$$

Then to show that $a'(\delta) < 0$ for sufficiently small $\delta$, this is equivalent to showing that $2\tilde{\rho}q(a(0)) < 3c_p$. Since $q$ is invertible, this is equivalent to showing that $a(0) > q^{-1}\left(\frac{3c_p}{2\tilde{\rho}}\right)$. Given that $a$ solves (C.91) asymptotically and that the left-hand side of (C.91) is increasing in $a$ while the right-hand side is decreasing in $a$, we want to show that $a\tau(1+\omega) > \frac{c_p^2}{a\alpha q(a)}$ for $a = q^{-1}\left(\frac{3c_p}{2\tilde{\rho}}\right)$. This is equivalent to showing that $\tau > \frac{2c_p\tilde{\rho}}{3\alpha(1+\omega)q^{(-1)}\left(\frac{3c_p}{2\tilde{\rho}}\right)^2}$. To show that this holds under the focal region, we compare this bound to the lower bound on $\tau$, $\frac{c_p\tilde{\rho}}{\alpha\left(q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)\right)^2}$. That $\frac{c_p\tilde{\rho}}{\alpha\left(q^{-1}\left(\frac{c_p}{\tilde{\rho}}\right)\right)^2} > \frac{2c_p\tilde{\rho}}{3\alpha(1+\omega)q^{(-1)}\left(\frac{3c_p}{2\tilde{\rho}}\right)^2}$ is equivalent to showing that $\frac{\tilde{\rho}}{c_p q^{(-1)}\left(\frac{c_p}{\tilde{\rho}}\right)^2} > \frac{2\tilde{\rho}}{3c_p q^{(-1)}\left(\frac{3c_p}{2\tilde{\rho}}\right)^2}$. Note that $\frac{d}{dx}\left[\frac{\frac{1}{x}}{(q^{-1}(x))^2}\right] = -\frac{q^{-1}(x) + \frac{2x}{q'(q^{-1}(x))}}{x^2 q^{-1}(x)^3} < 0$. Then

showing $\frac{\tilde{\rho}}{c_p q^{(-1)}\left(\frac{c_p}{\tilde{\rho}}\right)^2} > \frac{2\tilde{\rho}}{3c_p q^{(-1)}\left(\frac{3c_p}{2\tilde{\rho}}\right)^2}$ holds is equivalent to showing $\frac{c_p}{\tilde{\rho}} < \frac{3c_p}{2\tilde{\rho}}$, which is true. Then $a'(\delta) < 0$ for sufficiently small $\kappa$ and $\delta$, which means $\frac{d}{d\delta}[W_{RW}] > 0$ for sufficiently small $\kappa$ and $\delta$ from (C.94). ■

# Appendix D:
# Robustness Check for Generalized Model

In this appendix, we demonstrate how our insights from the main model are robust to the generalization of attacker heterogeneity at both entry cost and revenue levels (as per the generalized model introduced in Appendix C.1).

When the attacker market structure consists exclusively of attackers entering with ransomware, we derived in Appendix C.1.1 the equilibrium type-dependent ransom $R^*(\theta)$ as

$$R^*(\theta) = \frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)} + \frac{(1-\delta)\left(c_p - q(a^*)\left(\int_{1-a^*}^{1} \frac{\tilde{\rho}\lambda(\xi)}{2a^* s(\xi)}\, d\xi\right)\right)}{(1+\delta)q(a^*)}. \tag{D.1}$$

where $a^*$ is the equilibrium attacker population size (suppressing $\tilde{\Phi}^*_{RW}$ for brevity). Similarly, when the attacker market structure consists of two segments in equilibrium, those who enter with ransomware and those who enter with standard attacks, we derived in Apprendix C.1.2 the equilibrium type-dependent $R^*(\theta)$ as

$$R^*(\theta) = \frac{\tilde{\rho}\lambda(\theta)}{2s(\theta)} + \frac{(1-\delta)\left(c_p a^* - a_2^* q(a^*)\left(\int_{1-a_2^*}^{1} \frac{\tilde{\rho}\lambda(\xi)}{2a_2^* s(\xi)}\, d\xi\right)\right)}{(2a_1^* + a_2^*\delta + a_2^*)q(a^*)}, \tag{D.2}$$

where $a_1^*$ is the equilibrium size of the attacker population entering with standard attacks, $a_2^*$ is the size of the attacker population entering with ransomware attacks, and $a^* = a_1^* + a_2^*$.

Examining the expression for $R^*(\theta)$ in both cases and noting that only the first term depends on $\theta$, it follows that:

**Proposition D.1** *There exist bounds $\bar{s}$, $\bar{\lambda} > 0$ such that if $s'(\theta) < \bar{s}$ and $\lambda'(\theta) < \bar{\lambda}$ for all $\theta \in \Theta$, then $R^*(\cdot)$ is increasing (decreasing) if and only if $\lambda(\cdot)/s(\cdot)$ is increasing (decreasing) on $\Theta$.*

**Proof of Proposition D.1:** Whether the equilibrium outcome has one attacker segment or two, $R(\theta)$ is $R(\theta) = \frac{1}{2}\left(\frac{\tilde{\rho}\lambda(\theta)}{s(\theta)}\right) + C$ for some constant $C$ (see (C.6) and (C.15)). Differentiating both sides with respect to $\theta$, it follows that $\frac{dR(\theta)}{d\theta} > 0$ if and only if $\frac{d}{d\theta}\left(\frac{\lambda(\theta)}{s(\theta)}\right) > 0$. ■

This establishes a monotonicity characterization of the equilibrium ransom strategy profile. For illustration purposes, we utilize linear functional forms $\lambda(\theta) = 1 - k_1(1-\theta)$ and $s(\theta) = 1 - k_2(1-\theta)$ with $0 \le k_1, k_2 \le 1$. With respect to these specific functional forms, Proposition D.1 establishes that $R^*(\cdot)$ is increasing when $k_1 > k_2$ and decreasing when $k_1 < k_2$, for $k_1$ and $k_2$ satisfying the bounds

in Proposition D.1.

The main goal of this section is to illustrate how our insights from Sections 4 and 5 extend into a world in which different attackers charge different ransoms. We first note that the analysis in those sections can be thought of as a special case of the model in this section, with $k_1 = k_2 = 0$ (more generally, with $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$). For example, examining equations (D.1) and (D.2), when $\lambda(\theta) = 1$ and $s(\theta) = 1$ for all $\theta$, $R^*(\theta)$ would no longer depend on $\theta$ anymore in either equation. A natural question is how would our main insights change when accounting for attacker heterogeneity at the revenue level which, in turn, leads to different ransoms being charged.

In analyzing this model extension with different ransoms from different attackers, we will need to understand what it means to be a consumer willing to pay ransom if hit. When different attackers charge different ransoms, some consumers may decide to pay some ransoms charged in equilibrium but not other ransoms that are too high for them. As discussed in Section 3.2, $R_{max}(v) = \alpha v(1-\delta)$ is the maximum ransom amount that a consumer of type $v$ would be willing to pay if hit with ransomware. Let $\tilde{v}_r$ now represent the consumer type who is only willing to pay the lowest ransom charged in equilibrium, denoted $\underline{R}$ (among all the different ransoms charged by attackers in equilibrium). In this case, $\tilde{v}_r = \frac{\underline{R}}{\alpha(1-\delta)}$, which characterizes the marginal type who is unpatched and indifferent between paying the minimum ransom charged and not paying any ransom in equilibrium.

Next, consider consumers with valuations slightly higher than $\tilde{v}_r$. When deciding to remain unpatched, they risk being hit by an attacker who charges a ransom higher than their willingness to pay (i.e., being hit with ransoms greater than $v\alpha(1-\delta)$ when $v$ is close to $\tilde{v}_r$). By (7), even though they may pay some ransoms that are low enough from their perspective, if the ransom is higher than their threshold, then they would opt to incur the full loss $\alpha v$ instead of paying ransom.

As $v$ increases further, consumers are more and more willing to pay the higher ransoms charged in equilibrium. At some point, there is a type willing to pay any ransom charged in equilibrium. Denoting this type $\tilde{v}_s$, then similarly $\tilde{v}_s = \frac{\bar{R}}{\alpha(1-\delta)}$, where $\bar{R}$ is the highest ransom charged by an attacker in equilibrium. Consumers of type $v \in (\tilde{v}_r, \tilde{v}_s)$ would only pay ransoms lower than their threshold, but consumers with $v \geq \tilde{v}_s$ would be willing to pay any ransom charged in equilibrium (said differently, $\bar{R} \leq R_{max}(\tilde{v}_s)$). Consequently, consumers of valuations in $v \in (\tilde{v}_s, \tilde{v}_p)$ would be willing to pay any ransom charged by an attacker in equilibrium. Lastly, those with valuations $v \geq \tilde{v}_p$ prefer to patch rather than be exposed to risk of a cyberattack. This patching threshold can be written as either

$$\tilde{v}_p = \frac{c_p - q(a^*)\left(\int_{1-a^*}^{1} \frac{R(\theta)}{a^*}\, d\theta\right)}{\alpha\delta q(a^*)} \tag{D.3}$$

or

$$\tilde{v}_p = \frac{a^*\left(c_p - q(a^*)\left(\int_{1-a_2^*}^{1} \frac{R(\theta)}{a^*}\, d\theta\right)\right)}{\alpha(a_1^* + a_2^*\delta)q(a^*)}, \tag{D.4}$$

when all attackers prefer to enter with ransomware, and when attackers segment between ransomware and standard attacks, respectively.

Going forward, we continue to assume that $\lambda(\cdot)/s(\cdot)$ is strictly monotone on $\Theta$ to ensure that $R^*(\theta)$ is invertible such that $R^{*-1}(v\alpha(1-\delta))$ is well-defined for any $v$. In the case of the specific linear functional form used above, this assumption is satisfied provided that either $k_1 > k_2$ or $k_2 > k_1$ are selected such that they satisfy $k_1 < \bar{\lambda}$ and $k_2 < \bar{s}$.

With this foundation, we can numerically explore the robustness of our results, examining whether they extend to this more complex setting with attackers charging different ransoms. We demonstrate robustness using $k_1 = 0.2$ and $k_2 = 0.15$ as an example of moderate attacker revenue heterogeneity. We reproduce Figures 2 and 3 from our primary analysis using this extended framework while retaining consistent parameter values in Figures D.1 and D.2, respectively. Because ransoms are now type dependent, in panel (e) of Figure D.1, we plot the expected ransom over all attacker types entering with ransomware, defined as $E_\theta[R^*(\theta;\tau)\,|\,\tau] = \frac{1}{a^*}\int_{1-a^*}^{1} R^*(\theta;\tau)d\theta$ (with $R^*(\theta)$ coming from (D.1) when all attackers enter the market with ransomware) and $E_\theta[R^*(\theta;\tau)\,|\,\tau] = \frac{1}{a_2^*}\int_{1-a_2^*}^{1} R^*(\theta;\tau)d\theta$ (with $R^*(\theta)$ coming from (D.2) when attackers segment in their attack mode). Also, in panel (a) of Figure D.1, the ransom-paying population reflects the consumers who are willing to pay some of the ransoms being demanded. As long as they will pay the minimum ransom demanded, they belong to this group. Overall, the added complexity of $R^*(\theta)$ allows us to capture a richer consumer market structure. Despite the added complexity stemming from different attackers charging different ransoms, in terms of the comparative analysis we do in the primary analysis, Figures D.1 and D.2 illustrate that the insights remain robust.

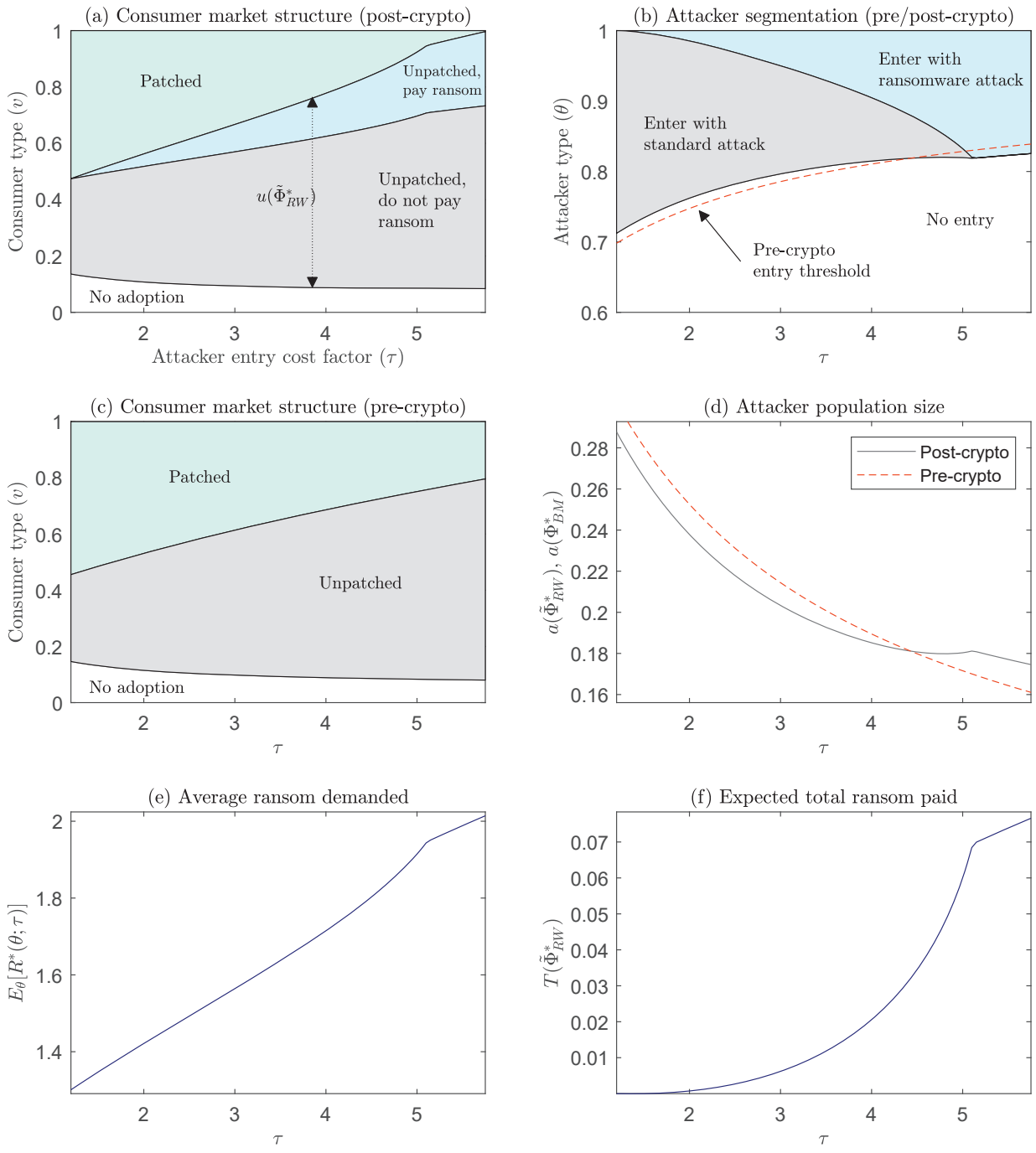


Figure D.1: The impact of attacker entry factor ($\tau$) on equilibrium market outcomes under pre-crypto benchmark and post-crypto ransomware settings with attacker heterogeneity in revenues. The parameter values are: $c_p = 0.3$, $\delta = 0.02$, $\kappa = 0.05$, $\alpha = 2.8$, $\rho = 1.6$, $\tilde{\rho} = 1.3$, $\omega = 0.1$, $q(a) = 0.9a - 0.4a^2$, $\lambda(\theta) = 1 - 0.2(1-\theta)$, and $s(\theta) = 1 - 0.15(1-\theta)$.

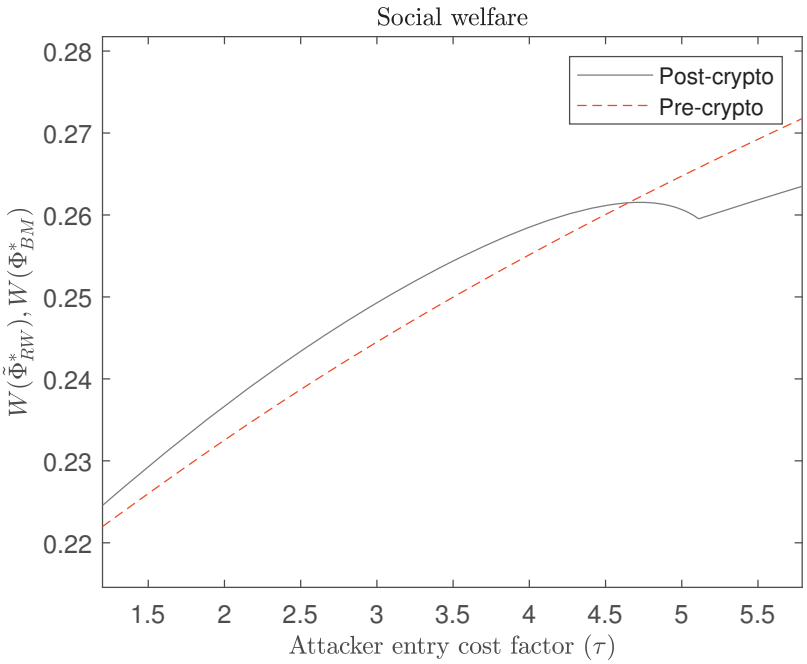


Figure D.2: The impact of attacker entry factor ($\tau$) on social welfare outcomes under pre-crypto benchmark and post-crypto ransomware settings with attacker heterogeneity in revenues. The parameter values are: $c_p = 0.3$, $\delta = 0.02$, $\kappa = 0.05$, $\alpha = 2.8$, $\rho = 1.6$, $\tilde{\rho} = 1.3$, $\omega = 0.1$, $q(a) = 0.9a - 0.4a^2$, $\lambda(\theta) = 1 - 0.2(1-\theta)$, and $s(\theta) = 1 - 0.15(1-\theta)$.