Incentive Analysis of Bidirectional Threat Filtering in the Internet

MHR. Khouzani[1], Soumya Sen[2], and Ness B. Shroff[3]

[1] University of Southern California, rezaeikh@usc.edu
[2] Princeton University, soumyas@princeton.edu
[3] The Ohio State University, shroff@ece.osu.edu

Abstract. Continuous incidents of cybersecurity attacks have underscored the importance of adoption of intrusion detection and prevention systems. Internet Service Providers (ISPs) are specially under attention as they are strategically positioned to filter data flows entering and leaving access networks before they reach their targets. However, any policy prescription needs to take into account the incentives of ISPs as financial decision makers, as adoption of such security measures is not without cost. Further complicating the analysis is the cross effects of security decisions that can create 'free-riding' tendencies, as well as the 'shortsightedness' of the ISPs. In this work, we develop an analytic framework that incorporates technological factors such as asymmetries in the performance of bidirectional firewalls in the adoption dynamics of ISPs. Specifically, we demonstrate how the double-effect of egress filtering in improving blocking efficiency but increasing free-riding tendencies can make policy making nontrivial. We will investigate the optimum egress filtering for social welfare as well as the aggregate security of network, and explore the various effects of shortsightedness.

1 Introduction

The growing incidents of information security breaches, presumably by both state (e.g., Stuxnet) and non-state actors (e.g., hacktivists) on various Government, industrial and e-commerce networks have heightened the need for such institutions to adopt effective security measures. According to [1], "target firms suffer losses of 1%-5% in the days after an attack. For the average New York Stock Exchange corporation, price drops of these magnitudes translate into shareholder losses of between $50 million and $200 million". The authors in [2] put the estimate of losses for US businesses with more than 1000 employees at $266 billion or approximately 2.7% of the GDP [3]. For instance, cybersecurity has already been identified as *"one of the most serious economic and national security threats"* by the US Government in creating its Comprehensive National Cybersecurity Initiative (CNCI) [4].

Our work approaches the problem of cybersecurity investments in the context of adoption of *asymmetric, bidirectional* security measures by Internet Service Providers (ISPs) in the Internet. ISPs, as gateways in the Internet, are strategically positioned for installment of intrusion detection and prevention systems (IDPSs), as they can monitor both inbound (*ingress*) and outbound (*egress*) traffic from their subnets and block malicious flows. This combined ingress and egress filtering can potentially improve the efficacy of the security measures by decreasing the success rate of intrusion attempts. However, egress filtering may lead to under-investment (i.e., a lower level of equilibrium adoption) if it improves the option of non-adoption more than it improves the option of adoption. These 'opposing' effects of egress filtering makes the question of "what is the optimal level of egress filtering from a policy making standpoint" a non-trivial one. Insights on this question constitutes the main contribution of this work. Our other contributions are as follows:

- In general, the equilibrium state of adoption is not unique and depends on the initial seeding of the ISPs. However, for a given initial seeding, the equilibrium state is unique and stable.
- Higher efficiency of egress filtering *decreases* the overall adoption level among ISPs, hence creating a dilemma for policy making.
- The (centralized) *social* optimum solution for firewall adoption results in *each* ISP having an *individually* higher utility compared to the decentralized equilibrium. Note that in general, a socially pareto-optimal solution does not guarantee individual advantage for *all* agents, but we show it is the case in this problem.
- The optimum egress filtering in the *decentralized* scenario is (rather counter-intuitively) providing *no protection* at all (i.e., zero egress filtering). This is while the optimum egress filtering in the *centralized* scenario is providing maximum protection. This negative result suggests a need for regulation of *both* cybersecurity products and ISPs.
- Shortsightedness of the ISPs is undesirable for both social welfare of the ISPs and the overall security of the network as it leads to lower aggregate investment for security in the equilibrium. We characterize the inefficiency due to shortsightedness of the decision-takers by introducing the new (general) notion of *price of shortsightedness*.

Related Literature Game theory has been used to analyze investments in network security e.g. in [5–7], focusing on questions like self-protection versus self-insurance [5], the effect of uncertainties in risk estimations of the investors [6], and incentive alignment of agents in the face of positive network externalities [7]. Others have proposed cyberinsurance as a complementary security mechanism [8]. Epidemiologists have also studied free-riding's impact on vaccination strategies as some parents myopically choose not to vaccinate their children if the relative risks of contacting a disease given its current prevalence is less than the potential for damage from the vaccine [9]. Similar results were also reported using game-theoretic tools for modeling *vaccination games* [10–12].

Egress filtering by ISPs as an effective means for improving cybersecurity and the issues with incentives and externalities have been pointed out in previous literature: e.g. [13–20]. Bauer *et al.* in [13] stipulate that externalities in security investments lead to "decentralized individual decisions" that are not socially optimal, and result in internet services that are "less secure than is socially desirable". Hofmeyr *et al.* in [14] (presented at WEIS 2011) develop an agent-based simulation model called ASIM that helps study the effect of different filtering policies adopted by ISPs, specifically ingress, egress, transit traffic, and their combinations. The authors use their model to evaluates global effectiveness of each of these policies. For instance, they show that intervention by the top 0.2% of ASes is more effective than intervention by 30% of ASes chosen at random. The paper also identifies that egress filtering creates positive externalities “with non-intervening ASes experiencing similar reductions in wicked trafc rates as intervening ASe”, and the authors measure its magnitude through simulations. In [15] (presented at WEIS 2010), Clayton argues for involvement of government to subsidize ISP interventions as a means of overcoming their socially inefficient investments. In [19], van Eeten *et al.* (an OECD paper) build a model of ISP incentives to mitigate malware using extensive empirical data. In [3] (another OECD paper) section V, authors provide a literature review on the different origins and forms of externalities in computer networked environment emphasizing cybersecurity.

Our analytical work extends these contributions by (a) rigorously focusing on the question of ISP-wide adoption of security measures as opposed to individual user-level patching or cyberinsurance provisioning, (b) incorporating practical considerations such as differentiating between (one-time) installation fee and the usage cost of the security measures as well as allowing subsequent disabling/re-enabling the security measures as part of the security decisions of the ISPs, dynamics over time, discounting future events, dynamics of intrusion and clean-up, etc., (c) modeling and analyzing the performance of IDPSs with *asymmetric, bidirectional* traffic monitoring capabilities. The latter is particularly relevant from a policy-making perspective of optimum choice of egress filtering, especially in view of the US Government’s CNCI [4] policy of conducting *“real-time inspection and threat-based decision-making on network traffic entering or leaving executive branch networks,”* and (d) explicitly deriving regulation constraints for managing the free-riding and shortsightedness of the ISPs.

From a technical point of view, our model starts with a *population game* (also known as *nonatomic* or *mean-field* game) by assuming a continuum of ISPs as decision-making agents, and investigate its *stationary* (steady-state) equilibria. The link between the stationary equilibria of a nonatomic game and the game with (large but) finite number of players is made concrete in papers like [21,22]. Specially, they in part show that the stationary equilibria of nonatomic game approximate the (perfect) Nash Equilibria of their finite-player game counterpart. Also, our bidirectional security adoption game belongs to the larger class of *stable games* (sometimes also referred to as *submodular* games, or games with *strategic substitudes*), which are studied in e.g. [23,24]. Notably, they show that for such games, the set of Nash Equilibria *coincide* with the set of (globally) stable states, i.e., not only the stable stationary equilibria constitute Nash Equilibria, but also there is no other Nash Equilibria for such games. Moreover, a Mypoic Learning (ML) dynamics (which we use in our work too) is guaranteed to converge to the equilibria in such games.

2 System Model

An ISP provides a gateway that connects its subnet to the Internet. Intrusion Detection and Prevention Systems (IDPS) installed by ISPs can monitor the inbound (ingress) and outbound (egress) traffic for blocking security breaches. Adoption of such security measures has a stand-alone benefit for an ISP. Moreover, it can slow down the rate of attacks and provide positive externality to the rest of the ISPs (adopters and non-adopters) by improving the overall security of the network. Namely, the nodes in other

ISPs will be less likely to be targeted by an attack originating from the subnet of the protected ISPs. However, adoption of security measures is not without cost: there can be a one-time purchase and installation fee, as well as recurrent usage costs. These recurrent costs can represent routine maintenance, as well as the losses for degradation of the quality of service due to privacy concerns or slowdown of the communications by latencies introduced in traffic monitoring. Moreover, a security measure can have a *false positive* rate, that is, it occasionally leads to obstruction of legitimate traffic. In what follows we provide an abstract model that captures key attributes of adoption of security measures at the ISP or Autonomous System (AS) level. Note that our goal is a *qualitative analysis* of adoption patterns and the policies that can influence it. Hence, we make some technical assumptions along the way to keep the model analytically tractable. A list of our main notations is provided in Table 1.

We consider a continuous-time model with a continuum of interconnected ISPs. Once an ISP purchases the security measure, it may be able to un-adopt it by *disabling* it. Subsequent adoptions are realized by enabling the security measure, and in particular, do not entail paying the one-time purchase and installation fee of the security measure. Hence, we need a model that distinguishes between the first adoption and subsequent re-adoptions. To do this, we stratify the ISPs into three types: (1) ISPs that have *purchased* and *enabled* the security measure; (2) ISPs that have *not purchased* it; and (3) ISPs that have *purchased* the security measure but have *disabled* it. Let the *fraction* of ISPs of each of the above types at time t be $x(t)$, $y(t)$ and $1-x(t)-y(t)$, respectively. The pair $(y(t),x(t))$, where $x(t),y(t) \geq 0$, $x(t)+y(t) \leq 1$, represents the adoption state of the network at time t.

Each ISP independently revises its decision regarding the adoption of the security measure at independent random epochs that occur according to a Poisson processes with rate γ. These are the epochs at which an ISP updates its measure of the intrusion rates on its subnet and accordingly re-evaluates its contingent utilities. Specifically, the decisions of the ISPs are assumed to be their best response to the *current* state, that is, assuming the current rate of intrusions is not going to change.[4] Let $v_{ij}(x) : [0,1] \to [0,1]$ be the probability with which a decision-making ISP switches from state i to j given the current state x, where $i, j \in \{n,e,d\}$ represent the state of the ISP with respect to adoption. We set the convention that n indicates n̲ot purchased, e represents purchased and e̲nabled, and d denotes purchased but d̲isabled. Note that v_{ij} is a function of $x(t)$ only (and not $y(t)$), since the ISPs that have not obtained the security measure and those that have disabled it are functionally similar regarding the provided protection against threats. For a large number of ISPs, the dynamics of $(y(t),x(t))$ is path-wise close to the solution of the following ODE:

$$\begin{cases} \dot{y}(t) = -\gamma y(t) v_{ne}(t) \\ \dot{x}(t) = \gamma y(t) v_{ne}(t) + \gamma(1-x(t)-y(t)) v_{de}(t) - \gamma x(t) v_{ed}(t) \end{cases} \tag{1}$$

The decision of the ISPs is determined by comparing the expected utilities given each decision, where the expectation is taken with respect to future incidents of intrusions, i.e., their potential epochs of occurrence and durations before clean-up. Ac-

[4] This is an instance of *Myopic Learning (ML)* dynamics [24].

cordingly, we define $G_{ij}(x)$ for $i, j \in \{n, e, d\}$ to be the expected utility of an ISP if it changes its adoption state from i to j given that the current fraction of ISPs with enabled security measure is x. Simply put, $G_{ne}(x)$ is the expected utility of the ISP at its decision-taking epoch if it purchases and enables the security measure, $G_{nn}(x)$ is its expected utility if it stays yet-to-purchase, and so on. Introduction of these contingent utilities leads to the following decision rules in (1): If $G_{ed}(x(t)) \neq G_{ee}(x(t))$, then $v_{ed}(t) = \mathbf{1}_{G_{ed}(x(t))>G_{ee}(x(t))}$. If $G_{ed}(x(t)) = G_{ee}(x(t))$, then the ISP with the enabled security measure is indifferent between keeping it enabled and disabling it, and hence $v_{ed}(t) \in [0,1]$. Similar comments apply to $v_{ne}(t)$ and $v_{de}(t)$. To characterize the dynamics of the adoption, therefore, we need to evaluate these contingent utilities.

2.1 Evaluating Contingent Utilities

The utility of an ISP is a decreasing function of costs and losses. For ease of calculations, we assume risk-neutral ISPs, and hence, directly take the negative of the costs to be the ISPs' utility. These costs are composed of two main parts: one associated with the expected cost of intrusions, and the other is related to the adoption and usage costs (if any). Let C_0 denote the one-time purchase and installation fee of the security measure. Also, let c be the cost per unit time of using the security measure incurred by an adopter ISP due to maintenance, communication latencies, false positives, *etc.* Note that c differs from C_0, i.e., the single-time purchase fee to obtain the security measure. The usage cost of the security measure seen at a decision-making epoch is $\int_0^\infty e^{-rt} c\, dt = c/r$. The expected cost of intrusions itself can be broken down into two components: one related to already successful infiltrations to the subnet of the ISP, and the other related to future incidents of intrusions. To better understand the intrusion costs, we next introduce four auxiliary variables. Let $C_{\mathrm{p}}^{\mathrm{on}}(x)$ represent the expected cost due to *ongoing* incidents of successful intrusions, if the decision-making ISP is protected, i.e., is currently in state e. Similarly, define $C_{\mathrm{p}}^{\mathrm{fu}}(x)$ to be the expected cost of *future* successful intrusions if the decision-making ISP will be protected, i.e., will be in state e. In a similar manner, define $C_{\mathrm{u}}^{\mathrm{on}}(x)$ ($C_{\mathrm{u}}^{\mathrm{fu}}(x)$) to represent the expected costs associated with the ongoing (future) successful intrusions, if the decision making ISP is currently (will be) unprotected, i.e., is (will be) in state d or n. Introduction of these auxiliary variables helps to delineate the contingent utilities, as follows (recall that the negative sign is to convert the cost to reward):

$$\begin{cases} G_{nn}(x) = -C_{\mathrm{u}}^{\mathrm{on}}(x) - C_{\mathrm{u}}^{\mathrm{fu}}(x), & G_{ne}(x) = -C_{\mathrm{u}}^{\mathrm{on}}(x) - C_{\mathrm{p}}^{\mathrm{fu}}(x) - C_0 - c/r \\ G_{dd}(x) = -C_{\mathrm{u}}^{\mathrm{on}}(x) - C_{\mathrm{u}}^{\mathrm{fu}}(x), & G_{de}(x) = -C_{\mathrm{u}}^{\mathrm{on}}(x) - C_{\mathrm{p}}^{\mathrm{fu}}(x) - c/r \\ G_{ee}(x) = -C_{\mathrm{p}}^{\mathrm{on}}(x) - C_{\mathrm{p}}^{\mathrm{fu}}(x) - c/r, & G_{ed}(x) = -C_{\mathrm{p}}^{\mathrm{on}}(x) - C_{\mathrm{u}}^{\mathrm{fu}}(x) \end{cases} \tag{2}$$

Note that $G_{ne}(x)$ and $G_{de}(x)$ differ only in the purchase fee of the security measure; specifically, $G_{de}(x) = G_{ne}(x) + C_0$. Moreover, $G_{nn}(x) = G_{dd}(x)$, as both utilities are only associated with the expected cost of the intrusions, which is the same for an ISP with a disabled security measure and one that is yet to install it. Next, we complete the characterization of the contingent utilities by computing the auxiliary variables.

For simplicity of exposition, *we consider security breaches that do not propagate in the network*, that is, we *will not* consider attacks involving self-replicating malicious

Table 1: Main notations in the model

parameter	definition
$x(t)$	fraction of the ISPs at time t that have adopted and enabled the security measure
$y(t)$	fraction of the ISPs at time t that are yet to purchase
γ	rate at which each ISP updates its adoption decision
G_{ij}	expected utility of an ISP if it chooses adoption status j provided that its current adoption status is i
v_{ij}	the probability with which a decision-making ISP switches from adoption status i to j
Λ	rate of intrusion attempts on a subnet of an ISP
μ	rate at which a successful intrusion is detected and blocked
C_0	one-time purchase and installation fee of the security measure
c	per unit time (recurrent) usage cost of the security measure
K_0	instantaneous cost upon a successful intrusion
k	cost (loss/damage) per unit time of intrusion

codes (known as *worms*) in the current article. Hacking is a typical example of a non-replicating type of attack. We will refer to such attacks by the umbrella term of *intrusion* attempts. When a host in a subnet of an ISP is compromised, the ISP incurs an instantaneous cost of K_0 and a per unit cost of k that persist as long as the host is infiltrated. The instantaneous cost may reflect the losses due to exposure of private information or manipulation of data, while the per unit time cost can represent eavesdropping the network traffic, accessing the network at the cost of the victim, slowdown of the victim's machine or the ISP's service, *etc.*[5] We assume that the time it takes to detect and remove the infection is according to an exponential random variable with rate μ. This is an approximation that allows technical tractability: empirical data on the clean-up times suggest a more skewed distribution, for instance, a lognormal distribution as investigated in [25]. Machines are again susceptible to future attacks, since subsequent attacks are likely to exploit new techniques.

The success probability of an intrusion attempt depends in part on the status of the ISPs of the origin of the attack as well as the ISP of the target (destination) with regard to the adoption of the security measure.[6] Specifically, the highest chance of intrusion success is when *neither* of the ISPs have an enabled security measure, while the lowest likelihood is when *both* ISPs have (obtained and) enabled it. Based on the four different conditions for the adoption status of the ISPs of the origin and target of an intrusion, we define intrusion *success probabilities* π_0, π_1, Π_0 and Π_1 according to Table 2. Namely, π_1 is the success probability of intrusion if both ISPs have enabled security measures in

[5] These costs may be directly incurred the ISP (e.g., bandwidth leakage), or be internalized by the ISP given the terms of liabilities, or be due to loss of willingness to pay of customers.

[6] Note that intermediate routers do not monitor for threats and the only traffic monitoring for threats are at border (edge) ISPs.

place, π_1 is the success probability of an intrusion if only the target's ISP has adopted the security measure, and so forth.

Table 2: Success probabilities of an intrusion attempt

		ISP of the target	
		Protected	Not Protected
ISP of the attacher	Protected	π_1	Π_1
	Not Protected	π_0	Π_0

Without loss of generality, we let $\Pi_0 = 1$ and only consider the attempts that are successful in the absence of the security measure in the network. However, we continue to use the *notation* Π_0 in our formulation for presentation purposes. In general, the following ordering holds for the intrusion success probabilities:

$$0 \leq \pi_1 \leq \min\{\pi_0, \Pi_1\} \leq \max\{\pi_0, \Pi_1\} \leq \Pi_0 = 1. \tag{3}$$

These inequalities just state that the success probability of an intrusion that has to bypass the security measures of both the ISP of its own subnet and that of the target node is the smallest (π_1). The next probability in order, is the smaller between π_0, Π_1, depending on which protection is stronger: ingress (inbound) or egress (outbound), respectively. The highest probability of success (Π_0) pertains to the case in which the intrusion is not confronted with any security measure in either of the ISPs.

A successful intrusion has to bypass the security measure of its own ISP, *and* the security measure of the ISP of the target machine, when both ISPs are adopters. For a security measure whose mechanism of intrusion detection and prevention is only signature-based, rule-based, or blacklisting, if both ISPs have access to the same signature, rules or list databases then $\pi_1 = \min\{\pi_0, \Pi_1\}$, that is, if an intrusion can successfully bypass one of the security measures, it will be able to bypass the other one as well. We will refer to this case as the *mutually inclusive* scenario. However, it could be that they have access to different databases, hence it is likely that $\pi_1 < \pi_0$. Also, anomaly detection mechanisms are in essence probabilistic and they have a *false negative* chance. The past traffic history of the two ISPs differ, hence the blocking events of the two security measures may not be exactly mutually inclusive. In case the intrusion prevention outcomes of the security measures are *mutually independent*, for $\Pi_0 = 1$, we have $\pi_1 = \pi_0 \Pi_1$. A unifying model that captures both of the above scenarios at the two ends of a spectrum and as special cases is the following:

$$\pi_1 = \pi_0 \Pi_1 + \alpha(\min\{\pi_0, \Pi_1\} - \pi_0 \Pi_1), \text{ for an } \alpha \in [0,1] \tag{4}$$

Note that the *mutually inclusive* and *mutually independent* cases are retrieved for $\alpha =$ 1 and 0 respectively. We call the security measures that follow the structural equation of (4) "*non-cooperative*".[7] In the rest of the paper, we only consider such "non-cooperative" security measures.

Let Λ represent the rate of intrusion attempts on the subnet of an ISP in the absence of any security measure in the network. The rate of *successful* intrusion attempts on an ISP that *does not have* an enabled security measure is $\Lambda(x\Pi_1 + (1-x)\Pi_0)$. This is because x fraction of the intrusion attempts have to successfully bypass the security measure of their own ISP, hence their success probability is Π_1, and the rest of the intrusion attempts, i.e., $(1-x)$ fraction of them, are confronted with no security measure and hence, have success probability of Π_0. Similarly, the rate of *successful* intrusion attempts on an ISP that *has* an enabled security measure is $\Lambda(x\pi_1 + (1-x)\pi_0)$. These are the two rates that each ISP can readily measure, then calculate its contingent utilities and accordingly make an adoption decision. Note specifically that an ISP need not know or observe the values of x or Λ directly.[8]

We are now ready to compute the components of contingent utilities as in (2). Let σ represent the state of the decision-taking ISP with respect to the intrusion, specifically, $\sigma \in \{0,1,2,\ldots\}$ indicates the number of successful ongoing intrusions in an ISP's subnet at the time of the ISP's decision-taking. Without loss of generality (by shifting the time coordinate) we can take a decision taking epoch to be at $t = 0$. We also consider discount factor r in calculation of the utilities, that is, costs incurred at time t in future are discounted at e^{-rt} when evaluated at present time. A larger r designate more shortsighted ISPs. A successful intrusion that occurs at time $t = 0$ incurs the following expected cost on the network:

$$\chi = K_0 + \int_0^\infty \left(\int_0^t e^{-r\tau} k \, d\tau \right) e^{-\mu t} \mu \, dt = K_0 + \frac{k}{\mu + r} \tag{5}$$

From Wald's equation, $C_{\mathrm{u}}^{\mathrm{on}}(x)$ is $\mathbb{E}(\sigma|\text{unprotected}) \times (\chi - K_0)$, since $\mathbb{E}(\sigma|\text{unprotected})$ is the expected number of successful ongoing intrusions in an unprotected ISP (non-adopter or with disabled security), and $(\chi - K_0)$ is the expected cost of each of them.[9]

Let $\eta(x) := \Lambda(x\Pi_1 + (1-x)\Pi_0)$, i.e., the rate of successful intrusion attempts on an unprotected ISP. The expected cost of the future intrusions for an unprotected ISP, $C_{\mathrm{u}}^{\mathrm{fu}}(x)$, can be computed by conditioning on the first epoch at which a new successful

[7] For cooperative schemes it is theoretically possible for π_1 to be less than $\pi_0\Pi_1$.

[8] An implicit assumption here is that intrusion attempts are not *rare* events, in the sense that, each ISP at its decision revision epoch can measure the rate of intrusion attempts on its subnet. In terms of the parameters in our model, this translates to requiring $\Lambda \gg \gamma$. Otherwise, uncertainty and "learning" becomes a playing factor, which is in interesting direction of research in itself. As we will see, assuming $\Lambda \gg \gamma$ makes the value of γ irrelevant as far as the equilibria are concerned.

[9] This is based on the assumption that the costs of different intrusions add up, e.g., two concurrent successful intrusions from two independent intruders incur the ISP $2k$ cost per unit time. If this assumption is relaxed, that is, if multiple concurrent successful intrusions only account for k costs per unit time, then the expressions for contingent will be slightly different. However, *all of the results in this paper identically holds for that case too*.

intrusion attempt is made:

$$C^{\text{fu}}_{\text{u}}(x) = \int_0^\infty \left[e^{-rt}(\chi + C^{\text{fu}}_{\text{u}}(x))\right] e^{-\eta(x)t}\eta(x)dt = \left[\chi + C^{\text{fu}}_{\text{u}}(x)\right]\frac{\eta(x)}{\eta(x)+r} \Rightarrow C^{\text{fu}}_{\text{u}}(x) = \frac{\eta(x)}{r}\chi$$

The number of ongoing intrusions in the subnet of an unprotected ISP can be seen as the number of "requests" in an $M/M/\infty$ system with arrival (birth) rate $\eta(x)$ and service (death) rate μ. Hence $\mathbb{E}(\sigma|\text{unprotected})$ is simply $\eta(x)/\mu$. Combining these with (5) leads to the following expression for $G_{nn}(x)$ in (2):

$$\begin{aligned} G_{nn}(x) = -C^{\text{on}}_{\text{u}}(x) - C^{\text{fu}}_{\text{u}}(x) = -(\chi - K_0)\frac{\eta(x)}{\mu} - \chi\frac{\eta(x)}{r} &= -\frac{\eta(x)}{\mu r}(K_0\mu + k) \\ &= -\frac{K_0 + k/\mu}{r}\Lambda\left(\Pi_0 - x(\Pi_0 - \Pi_1)\right) \qquad (6) \end{aligned}$$

Derivation of the other utilities in (2) is now straightforward: First, note that $C^{\text{on}}_{\text{p}}(x)$ and $C^{\text{fu}}_{\text{p}}(x)$ are computed in the same way as $C^{\text{fu}}_{\text{u}}(x)$ and $C^{\text{fu}}_{\text{u}}(x)$, respectively. In particular, the attained expressions are correspondingly identical after replacing the unprotected intrusion success probabilities Π_0, Π_1, with the protected intrusion success probabilities π_0, π_1. For instance, $C^{\text{fu}}_{\text{p}}(x) = \chi\eta'(x)/r$ where $\eta'(x) := \Lambda(x\pi_1 + (1-x)\pi_0)$, that is the rate of successful intrusion attempts on a protected ISP. Hence:

$$\begin{aligned} G_{ne}(x) &= -C^{\text{on}}_{\text{u}}(x) - C^{\text{fu}}_{\text{p}}(x) - C_0 - \frac{c}{r} = -(\chi - K_0)\frac{\eta(x)}{\mu} - \chi\frac{\eta'(x)}{r} - C_0 - \frac{c}{r} \\ &= -\frac{\Lambda}{\mu}\left(\frac{k}{k+r}\right)(\Pi_0 - x(\Pi_0 - \Pi_1)) - \frac{\Lambda}{r}\left(K_0 + \frac{k}{k+r}\right)(\pi_0 - x(\pi_0 - \pi_1)) - C_0 - \frac{c}{r} \qquad (7) \end{aligned}$$

Recall that $G_{dd}(x)$ is equal to $G_{nn}(x)$, and $G_{de}(x)$ is just $G_{ne}(x) + C_0$. Finally, $G_{ee}(x)$ and $G_{ed}(x)$ are obtained in similar manners as in (6) and (7) to be:

$$G_{ee}(x) = -\frac{\Lambda}{r}(K_0 + k/\mu)(\pi_0 - x(\pi_0 - \Pi_1)) - \frac{c}{r} \qquad (8)$$

$$G_{ed}(x) = -\frac{\Lambda}{\mu}\left(\frac{k}{k+r}\right)(\pi_0 - x(\pi_0 - \pi_1)) - \frac{\Lambda}{r}\left(K_0 + \frac{k}{k+r}\right)(\Pi_0 - x(\Pi_0 - \Pi_1)) \qquad (9)$$

Note that all of the expected contingent utilities G_{ij} turn out to be linear in x. A straightforward yet important property of the contingent expected utilities is that they are (all) increasing in the value of x:

Lemma 1. *For any $x \in [0,1]$ we have: $\frac{\partial G_{ij}(x)}{\partial x} \geq 0$ for all $i, j \in \{n, e, d\}$. The equality holds only if $\Pi_1 = \Pi_0$.*

Hence, *positive externalities exist for both adopters and non-adopters*. In other words, both options of adoption and non-adoption improves as more ISPs adopt the security measure. Moreover, the positive externalities vanish only when there is no protection against the outgoing threats, i.e., zero egress filtering.

2.2 Equilibrium Points, Uniqueness and Stability

We open this section with another straightforward lemma that we will use later.

Lemma 2. *For $\Pi_1 < \Pi_0$, i.e., nonzero egress filtering, we have* $\frac{\partial}{\partial x}(G_{ne}(x) - G_{nn}(x)) < 0$ *at any* $x \in [0,1]$. *The same inequality holds for* $\frac{\partial}{\partial x}(G_{de}(x) - G_{dd}(x))$ *and* $\frac{\partial}{\partial x}(G_{ee}(x) - G_{ed}(x))$.

In words, *even though both adopters and non-adopters experience positive externalities of the security measure adopted by others, the non-adopters relatively benefit more*. Put other way, the option of non-adoption more rapidly becomes more appealing as adoption grows. This is the core cause of the problem of free-riding.[10]

The proof of the lemma reduces to verifying the following inequality:

$$(\Pi_0 - \Pi_1) - (\pi_0 - \pi_1) > 0. \tag{10}$$

For $\Pi_1 < \Pi_0$, the above inequality follows from (4).

Lemmas 1 and 2 have an important corollary: $G_{ne}(x)$ crosses $G_{nn}(x)$ (at most) once (over the interval of $(0,1)$). Let ζ designate this crossing point. Similarly, $G_{de}(x)$ crosses $G_{dd}(x)$ only once, which we will denote by ζ'. It follows from (2) that ζ' is also the crossing point of $G_{de}(x)$ and $G_{ee}(x)$. Hence, recalling that for $C_0 > 0$, $G_{ne}(x) < G_{de}(x)$, in the most general case $\exists \zeta, \zeta'$, $0 < \zeta < \zeta' < 1$, such that (Fig. 1(a)):

$$\begin{cases} G_{nn}(x) < G_{ne}(x),\ G_{dd}(x) < G_{de}(x),\ G_{ed}(x) < G_{ee}(x) \text{ for } x \in [0,\zeta) \\ G_{nn}(x) > G_{ne}(x),\ G_{dd}(x) < G_{de}(x),\ G_{ed}(x) < G_{ee}(x) \text{ for } x \in (\zeta,\zeta') \\ G_{nn}(x) > G_{ne}(x),\ G_{dd}(x) > G_{de}(x),\ G_{ed}(x) > G_{ee}(x) \text{ for } x \in (\zeta',1] \end{cases}$$

The extension to other special cases is straightforward.[11] This yields the phase portrait depicted in Fig. 1(b). Equilibrium points, as stationary states of the adoption process, are derived by equating both $\dot{x}$ and $\dot{y}$ in (1) to zero. From the phase diagram (or direct computation), the equilibrium points are derived to be $\{(y^*,x^*)|x^* \in (\zeta,\zeta'), y^* = 1 - x^*\} \cup \{(y^*,x^*)|x^* = \zeta', y^* \in [0, 1-\zeta']\}$. Note in particular that the equilibrium is *not* unique. Which equilibrium point is eventually achieved depends on the initial condition that the system starts with. For instance, if $x(0) = x_0 \in (\zeta,\zeta')$, say by seeding the ISPs with free copies of the security measure, then those ISPs that do not have the security measure have no incentive to obtain it. Hence, in the equilibrium, $y^* = y_0$. Now, if $y_0 > 1 - \zeta'$, then the ISPs that have the security measure will enable it until there is no ISP with a disabled security measure, and hence, $x^* = 1 - y_0$. On the other hand,

[10] Technically, this property makes our game one with *strategic substitutes*, or equivalently, a *stable* or *submodular* game. The reader is encourage to refer to [23, 24] for results on the stability and convergence properties of this class of games.

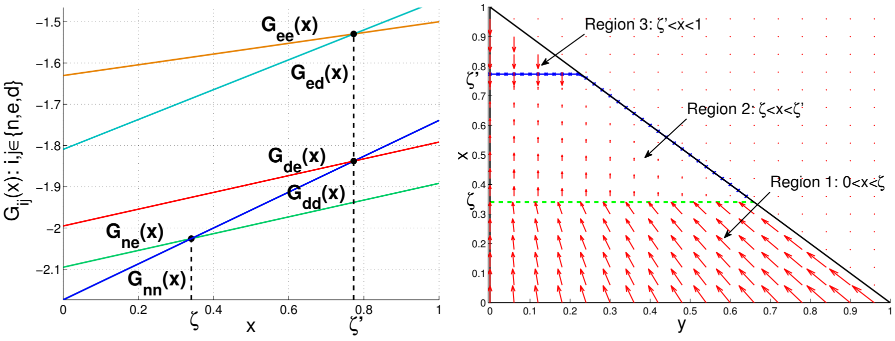
[11] For instance, in the trivial cases such as when $G_{ne}(x) > G_{nn}(x)$ for all $x \in (0,1)$ (sufficient to verify $G_{ne}(1) > G_{nn}(1)$), then the unique and stable equilibrium of the system is the $(y^*,x^*) = (0,1)$, i.e., full adoption. As another example, if $G_{ee}(x) < G_{nn}(x)$ for all $x \in (0,1)$ (suffices to check if $G_{ee}(0) < G_{nn}(0)$) then the equilibrium point is trivially $(y^*,x^*) = (1 - x_0 - y_0, 0)$.

if $y_0 < 1-\zeta'$, then ISPs with the security measure will progressively enable it until $x^* = \zeta'$ fraction of the ISPs have enabled their security measure. Subsequent ISPs with the security measure have no incentive to enable theirs. A practically interesting case is when the system starts with no seeding, that is $(y_0, x_0) = (1, 0)$. In this case $x(t)$ progressively rises to slightly above ζ and then the system is *locked* into the point $(\zeta^+, 1-\zeta^+)$. For practical purposes, as long as the value of the equilibrium level is concerned, we can (and henceforth will) take $(y^*, x^*) = (1-\zeta, \zeta)$ as the equilibrium point for the case of $(y_0, x_0) = (1, 0)$. We explored the issue of seeding in more detail in our previous work [26]. In part, we showed that for an initial seeding of S fraction of the ISPs with a free installation of the security measure, the equilibrium is given by:

$$y^* = (1-S)(1-\zeta),\; x^* = \begin{cases} \zeta' & \frac{\zeta'-\zeta}{1-\zeta} \le S \le 1 \\ S+\zeta-\zeta S & S < \frac{\zeta'-\zeta}{1-\zeta} \end{cases}$$

which can be further used for a comparative statics of seeding. In summary, we have the following proposition:

Proposition 1. *The equilibrium is not unique and depends on the initial value. However, for any given initial seeding of the ISPs, the equilibrium is unique and stable.*



(a) (b)

Fig. 1: (a) contingent utilities as functions of the level of adoption. (b) the phase portrait.

An implication of the proposition is that (for non-cooperative security measures), we do *not* have the phenomenon of *tipping point* (a.k.a. *critical mass*), as they are associated with unstable equilibrium points. For the rest of the paper, we assume that the system starts from $(y_0, x_0) = (1, 0)$. Moreover, without loss of generality, we assume that the equilibrium point is non-trivial. A corollary of this assumption and the assumption of starting from $(y_0, x_0) = (1, 0)$ is that the equilibrium point (y^*, x^*) is $(1-\zeta, \zeta)$), and at which we have $G_{ne}(x^*) - G_{nn}(x^*) = 0$.

2.3 The Effect of Outbound Protection and the Phenomenon of Free-Riding

Here, we show a rather intuitive result, that *less* protection against the outgoing threats by the security measure (hence a larger Π_1), leads to a *higher* equilibrium level of adoption. Equivalently, the more the security measure provides egress filtering, the less ISPs will be willing to adopt the security measure in the equilibrium. Intuitively, with lower ability of the security measure to block the outbound threats, ISPs feel the need to block the inbound threats instead of free-riding on others' investments.[12] This is expressed formally in the following simple proposition.

Proposition 2. $\frac{dx^*}{d\Pi_1} > 0$*, i.e., more outbound protection leads to lower equilibrium level of adoption.*

Proof. From our discussion in §2.2, for $(y_0, x_0) = (0, 1)$, x^* is the solution of $G_{nn}(x) = G_{ne}(x)$. Equating (6) and (7) yields:

$$-\chi\frac{\eta(x^*)}{r} = -\chi\frac{\eta'(x^*)}{r} - C_0 - \frac{c}{r} \Rightarrow \frac{dx^*}{d\Pi_1} = -\frac{\frac{\partial\eta(x^*)}{\partial\Pi_1} - \frac{\partial\eta'(x^*)}{\partial\Pi_1}}{\frac{\partial\eta(x^*)}{\partial x^*} - \frac{\partial\eta'(x^*)}{\partial x^*}}$$

$$\Rightarrow \frac{dx^*}{d\Pi_1} = \frac{(1 - \frac{d\pi_1}{d\Pi_1})x^*}{(\Pi_0 - \Pi_1) - (\pi_0 - \pi_1)}.$$

The proposition now follows from (4) and (10). □

2.4 Pareto-Optimal Level of Adoption

Suppose that a social planner imposes the adoption decisions of the ISPs. In one scenario, the social planner may attempt to maximize the aggregate expected utility. The problem of such central planner is the to maximize the *social* utility as defined below:

$$U(x) := xG_{ne}(x) + (1-x)G_{nn}(x). \tag{11}$$

We denote an optimum x by $\hat{x}$, i.e., $\hat{x} \in \arg\max U(x), 0 \leq x \leq 1$. If $0 < \hat{x} < 1$, it must satisfy the following first-order condition:

$$G_{ne}(\hat{x}) - G_{nn}(\hat{x}) + [x\frac{\partial G_{ne}(x)}{\partial x} + (1-x)\frac{\partial G_{nn}(x)}{\partial x}] \mid_{x=\hat{x}} = 0.$$

For $\Pi_1 < \Pi_0$, both $\frac{\partial G_{ne}(x)}{\partial x}$ and $\frac{\partial G_{nn}(x)}{\partial x}$ are strictly positive for all $0 \leq x \leq 1$ according to Lemma 1. Therefore, either $\hat{x} = 0$, or $\hat{x} = 1$, or $G_{ne}(\hat{x}) - G_{nn}(\hat{x}) < 0$. Recall that x^*, the equilibrium point if ISPs could decide themselves, satisfied the condition $G_{ne}(x^*) -$

[12] There is a cautionary pitfall in this intuitive argument: one could argue that as more ISPs adopt the security measure, its efficacy increases and hence it becomes a more attractive service, leading to higher level of adoption. To counter this argument, the lower bound on π_1 is necessary. That is, the improvement in the efficacy of the security measure is outweighed by the improvement in the efficacy of the security measure for non-adopters.

$G_{nn}(x^*) = 0$. Also recall that for $\Pi_1 < \Pi_0$, $G_{ne}(x) - G_{nn}(x)$ is a strictly decreasing function of x (by Lemma 2). Hence we have the following result:

Proposition 3. *The socially optimum (planner-imposed) level of adoption, $\hat{x}$, is always greater than the equilibrium level of adoption, x^*. Moreover, if $0 < x^* < 1$ and $\Pi_1 < \Pi_0$, then $\hat{x} > x^*$.*

This proposition has important corollaries: Each ISP enjoys a higher utility in the socially optimum fraction of adoption, and thus, the equilibrium level of adoption is not pareto-optimal. This is because following Lemma 1, both $G_{ne}(x)$ and $G_{nn}(x)$ are increasing in x and according to the proposition, $\hat{x} \geq x^*$. Then a question may arise as to what prevents the ISPs from reaching this pareto-optimum level of adoption in which everybody is better off. The answer is that some ISPs are *more* better off than the others. Specifically, those ISPs that do not adopt would enjoy a higher utility than those who adopt it in the socially optimum solution. If the ISPs could freely decide, then they stop adopting once opting out starts to yield more utility, even though continuing to adopt will also increase their utility (but less than opting out would increase).

3 Optimum Level of Egress Filtering

We hitherto assumed that the protection level on outbound traffic (and hence Π_1) is a given parameter of the security measure. For a firewall, for instance, Π_1 can be any value between Π_0 (no protection) and π_0 (the same protection on the outbound traffic as for the inbound traffic). Recall that lesser values of Π_1 correspond to more protection against the outbound threats. Theoretically, developers of the firewall can add protection on the outbound traffic at least as much as the protection on the inbound traffic at no additional cost of production. In Proposition 2 in §2.3, we showed that $dx^*/d\Pi_1 < 0$, therefore, the firewall developers have an incentive to remove any protection against outbound threats to maximize their sales.

The regulator can set a minimum requirement on the amount of protection against outbound threats relative to the protection provided against the inbound threats by the firewall. In the language of our model, there can be a maximum value imposed on Π_1 relative to π_0. A natural idea seems to put the bound on Π_1 to be equal to π_0 as it corresponds to the highest protection against outbound as well as inbound threats. However, as we showed in Proposition 2, this can lead to free-riding among the ISPs yielding a low equilibrium level of adoption, and hence potentially a less secure network. On the other hand, if the bar on Π_1 is Π_0, then the highest level of adoption is achieved but the amount of protection against outbound threats is the weakest (nonexistent). Hence, a choice of the "best" Π_1 is a *non-trivial* question and depends on the metric and the perspective used. In what follows, we consider some of these different metrics.

3.1 View 1: Decentralized Social Optimum

Let us define a *decentralized social optimum* Π_1 as follows:

$$\Pi_1 \in \arg \max_{\pi_0 \leq \Pi_1 \leq \Pi_0} U(x^*), \tag{12}$$

where $U(x^*) = x^* G_{ne}(x^*) + (1-x^*) G_{nn}(x^*)$, is the social utility of the ISPs. We use the phrase *decentralized* to emphasize that ISPs are allowed to freely choose their individual best decision of adoption, and the system is accordingly at the equilibrium value x^*. Assuming $0 < x^* < 1$, x^* satisfies $G_{nn}(x^*) = G_{ne}(x^*)$. Hence, the above optimization is transformed to:

$$\Pi_1 \in \arg \max_{\pi_0 \leq \Pi_1 \leq \Pi_0} G_{ne}(x^*),$$

or equivalently,

$$\Pi_1 \in \arg \max_{\pi_0 \leq \Pi_1 \leq \Pi_0} G_{ne}(x) \;\; s.t. \;\; G_{ne}(x) - G_{nn}(x) = 0. \tag{13}$$

The above optimization is non-convex (due to the nonlinear equality constraint). Nevertheless, we can observe the following proposition without solving the optimization.

Proposition 4. *For "non-cooperative" security measures, that is, if* (4) *holds, any* Π_1 *is decentralized-socially optimum (i.e., is a solution to* (12)*).*

In words, if the ISPs are allowed to individually take their adoption decisions, then the optimum value of the social utility is *not affected* by the imposed level of egress filtering. At first glance, this might come as a surprise; after all, the value of x^* *does* change with changing Π_1. The proof of the proposition reveals why this does not affect the value of the social utility. Intuitively, it is because as Π_1 is decreased, the protection provided by the security measure improves, however, a smaller fraction of the ISPs end up adopting the security measure. The net impact is that the overall utility in the network stays unaffected.

Proof. The proof for the *mutually inclusive* scenario, i.e., when $\pi_1 = \pi_0$, is trivial: in (13), for $\pi_0 = \pi_1$, $G_{ne}(x)$ does not depend on either x or Π_1. The proof for the more general case when equation (4) holds follows next. We need to show $\dfrac{dG_{ne}(x^*)}{d\Pi_1} = 0$. A change in Π_1 affects the utility at the equilibrium both directly (dependence of G_{nn}, G_{ne} on Π_1) and indirectly through changing the equilibrium level x^*. Hence:

$$\frac{dG_{ne}(x^*)}{d\Pi_1} = \frac{\partial G_{ne}(x^*)}{\partial \Pi_1} + \frac{\partial G_{ne}(x^*)}{\partial x^*} \frac{dx^*}{d\Pi_1}. \tag{14}$$

To calculate $\dfrac{dx^*}{d\Pi_1}$, we note that from $G_{ne}(x^*) = G_{nn}(x^*)$ we can deduce:

$$\frac{\partial G_{ne}(x^*)}{\partial \Pi_1} + \frac{\partial G_{ne}(x^*)}{\partial x^*} \times \frac{dx^*}{d\Pi_1} = \frac{\partial G_{nn}(x^*)}{\partial \Pi_1} + \frac{\partial G_{nn}(x^*)}{\partial x^*} \times \frac{dx^*}{d\Pi_1}$$
$$\Rightarrow \frac{dx^*}{d\Pi_1} = -\left(\frac{\partial G_{ne}(x^*)}{\partial \Pi_1} - \frac{\partial G_{nn}(x^*)}{\partial \Pi_1} \right) \Big/ \left(\frac{\partial G_{ne}(x^*)}{\partial x^*} - \frac{\partial G_{nn}(x^*)}{\partial x^*} \right).$$

Replacing the terms in (14), we obtain:

$$\frac{dG_{ne}(x^*)}{d\Pi_1} = \left(\frac{\partial G_{ne}(x^*)}{\partial x^*} \times \frac{\partial G_{nn}(x^*)}{\partial \Pi_1} - \frac{\partial G_{ne}(x^*)}{\partial \Pi_1} \times \frac{\partial G_{nn}(x^*)}{\partial x^*} \right) \Big/ \left(\frac{\partial G_{ne}(x^*)}{\partial x^*} - \frac{\partial G_{nn}(x^*)}{\partial x^*} \right).$$

After replacing from (6) and (7), the nominator simplifies to have the multiplicand:

$$\frac{d\pi_1}{d\Pi_1}(\Pi_0 - \Pi_1) - (\pi_0 - \pi_1).$$

For $\Pi_1 \geq \pi_0$, using (4) the above expression simplifies to zero. □

Proposition 4 establishes that if ISPs freely take their adoption decisions, then the social optimum utility is *not affected* by the amount of protection on the outgoing traffic. Hence, it is interesting to investigate other metrics of optimality for Π_1, as we do next.

3.2 View 2: Decentralized Individual Optimum

For individuals adopting the security measure, a decentralized optimum Π_1 is given by: $\Pi_1 \in \arg\max_{\pi_0 \leq \Pi_1 \leq \Pi_0} G_{ne}(x^*)$. For individuals who opt out, an optimum Π_1 is provided by: $\Pi_1 \in \arg\max_{\pi_0 \leq \Pi_1 \leq \Pi_0} G_{nn}(x^*)$. Since at equilibrium we have $G_{nn}(x^*) = G_{ne}(x^*)$, solutions of the above two optimizations are the same; and are further equal to the solution of the social optimum utility as was shown in (13). Hence, a decentralized social optimum value of Π_1, is also decentralized individual optimum for all ISPs. Moreover, Proposition 4 applies to the decentralized individual optimum Π_1 as well.

3.3 View 3: Decentralized Security Optimum

Since, both decentralized social and decentralized individual utilities are unaffected by the value of Π_1, we define a new metric of practical interest. One can consider only the cost of intrusions in the network to be the metric of optimality. Accordingly, we define the *security* utility $V(x)$ to be the negative of the expected aggregate damage incurred on the network as a result of the intrusions if the adoption decisions are taken in a decentralized manner by the ISPs. Note specifically that $V(x)$ does not include the costs of the adoption of the security measures. In our problem, $V(x)$ is as follows:

$$V(x) = x\left(G_{ne}(x) + \frac{c}{r} + C_0\right) + (1-x)G_{nn}(x) \tag{15}$$

It is easy to see that $V(x)$ is increasing in x. Hence, in the light of Proposition 3, $V(\hat{x}) \geq V(x^*)$. That is, the socially optimum level of adoption not only provides a better aggregate expected utility (by construction), it also provides better overall network security, compared to the equilibrium level of adoption. If we were to *centrally* maximize $V(x)$, then the optimum choice for x (denoted by $\tilde{x}$) would be $\tilde{x} = 1$. Note that the adopters in this case win a still higher utility than the adopters in the socially optimum scenario. However, this is achieved at the cost of lower utilities for the non-adopters in the socially optimum choice of x (hence its achieved aggregate expected utility is lower).

Now, as we did in the previous two subsections, we define a *decentralized security optimum* Π_1 to be a solution of the following optimization problem: $\max_{\pi_0 \leq \Pi_1 \leq \Pi_0} V(x^*)$. The decentralized security optimum Π_1 is provided by the following proposition:

Proposition 5. $\Pi_1 = \Pi_0$*, i.e., no egress filtering, is decentralized security optimum.*

Proposition 5 is in fact a negative result: the highest security utility in the decentralized case is achieved if the regulator requires no protection on the outgoing traffic. In the next subsection, we discuss how this reveals a significant inefficiency.

Proof. From Proposition 4, changing Π_1 does not affect the value of $U(x^*)$. From the definition of $V(x)$ in (15), we have: $V(x^*) = U(x^*) + x^*(c/r + C_0)$. Hence, a Π_1 that maximizes $V(x^*)$ must maximize $x^*(c/r + C_0)$. For any non-trivial security measure, we have $c/r + C_0 > 0$. From Proposition 2, higher x^* is achieved by increasing Π_1. Therefore, $\Pi_1 = \Pi_0$ maximizes x^*, and hence $V(x^*)$. □

3.4 Centralized Optimum

We can consider the optimum centralized version of each of the previous three viewpoints: social, individual and security. For a centralized optimizer, the choice of variables x^* and Π_1 are *decoupled*. For any $\Pi_1 < \Pi_0$, we have $\partial G_{nn}(x)/\partial x, \partial G_{ne}(x)/\partial x > 0$. Also, for any x, we have $\partial G_{nn}(x)/\partial \Pi_1 < 0$, $\partial G_{ne}(x)/\partial \Pi_1 \leq 0$. Hence, the centralized optimum value of Π_1 that simultaneously maximizes $U(x)$, $V(x)$, $G_{ne}(x)$, $G_{nn}(x)$ is $\Pi_{1,\min}$, i.e., maximum protection on the outbound traffic. Note that this is despite potentially different centrally optimum x for each of these optimizations. For instance, the pair (Π_1, x) that maximizes the social utility is $(\Pi_{1,\min}, \hat{x})$, while the pair (Π_1, x) that maximizes the security utility is $(\Pi_{1,\min}, 1)$, and it is possible to have $\hat{x} < 1$.

In the decentralized scenario, i.e., when ISPs are free to adopt or not, maximum security utility is achieved *at the cost of* eliminating the protection against the outbound threats, as the values of x^* and Π_1 were coupled. This suggests that a regulation on the security measures alone is inefficient, and an efficient regulation should be on both security measures and ISPs.

4 The Price of Shortsightedness

4.1 The Effect of the Discount Factor on the Equilibrium

We showed in §2.2 that when the market is not seeded, i.e., $(y_0, x_0) = (1, 0)$, the equilibrium fraction of the nodes that adopt and enable the security measure is (practically) equal to ζ, which is the solution of $G_{ne}(x) = G_{nn}(x)$ from (6) and (7). Taking the derivative of both sides of the equation $G_{ne}(x) = G_{nn}(x)$ with respect to r yields:

$$\chi\frac{\eta(x^*)}{r^2} + \frac{k}{(\mu+r)^2}\frac{\eta(x^*)}{r} - \chi\frac{\frac{\partial\eta(x^*)}{\partial x^*}\frac{dx^*}{dr}}{r} = \chi\frac{\eta'(x^*)}{r^2} + \frac{k}{(\mu+r)^2}\frac{\eta'(x^*)}{r} - \chi\frac{\frac{\partial\eta'(x^*)}{\partial x^*}\frac{dx^*}{dr}}{r} + \frac{c}{r^2}$$

$$\Rightarrow \frac{dx^*}{dr} = \frac{(\eta(x^*) - \eta'(x^*))(\chi + kr/(\mu+r)^2) - c}{-\chi r\left[(\Pi_0 - \Pi_1) - (\pi_0 - \pi_1)\right]}$$

Noting that x^* satisfies $\eta(x^*) - \eta'(x^*) = (C_0 r + c)/\chi$, and the fact that following (10) the denominator in the last expression is negative, we obtain:

$$\operatorname{sgn}(\frac{dx^*}{dr}) = \operatorname{sgn}\left[-(C_0 r + c)kr - C_0 r\chi(\mu+r)^2\right]$$

Hence, the following result:

Proposition 6. $\dfrac{dx^*}{dr} < 0$*, i.e., the equilibrium level of adoption decreases with more shortsightedness of the ISPs.*

This result shows that shortsightedness of the ISPs leads to lower aggregate investment for protection.

4.2 Measuring the Price of Shortsightedness

We introduce a measure of inefficiency with respect to the measure of shortsightedness of the ISPs in their decision taking, that is, the effect of discounting future events and hence having a bias in favor of near future outcomes.[13] To make this formal, we first discuss the notion of a *reference* discount factor r_0. The expected utility associated with an achieved equilibrium can be measured using a potentially different discount factor from the one used by ISPs in their computation of individual utilities. We have used r to denote the discount factor used by the ISPs. We will use $x^*(r)$ to represent the achieved equilibrium level of adoption to emphasize the dependence of x^* on the discount factor of the ISPs. Let r_0 be the discount factor used by a referee. For instance, $G_{ne}(x^*(r), r_0)$ is the expected r_0-discounted utility of adopters in an equilibrium level x^* where the ISPs' discount factor is r.

We can now formally define the Price of (temporal) Shortsightedness. The Social Price of Shortsightedness (*SoPoSh*) is defined as follows:

$$SoPoSh(r) := \lim_{r_0 \to 0} \frac{U(\lim_{\rho \to 0} x^*(\rho), r_0)}{U(x^*(r), r_0)}$$

Note the dependence of *SPoSh* on r. Also, since we showed $x^*(r)$ is unique, there is no ambiguity between the choices of equilibria. Similarly, we can define the Security Price of Shortsightedness $SePoSh(r)$ as $\frac{V(x^*(\rho \to 0), r_0 \to 0)}{V(x^*(r), r_0 \to 0)}$.

Following its definition, higher prices of shortsightedness means that from the referee's viewpoint, shortsightedness of the decision-takers leads to inefficiency. As r approaches 0, these prices of shortsightedness approaches one. However, unlike Price of Anarchy and Stability, there is no general rule that the prices of shortsightedness is always less (or greater) than unity. That is, depending on the parameters of the problem, it might be "beneficial" to have less or more shortsighted decision-takers. This has interesting policy-making implications. Nevertheless, for our specific problem, we have the following proposition:

Proposition 7. *For all $r > 0$, both $SoPosh(r)$ and $SePoSh(r)$ are greater than one, i.e., more shortsightedness of the ISPs is undesirable from both viewpoints of the social welfare of ISPs and the aggregate security of the network.*

This result has the unequivocal policy implication that shortsightedness is undesirable for security-related investment decisions.

[13] This discounting can either reflect the behavioral (subjective) traits of the decision-takers or the objective depreciation of capital, or a mixture of both.

Proof. We present the proof for $SoPoSh(r)$; the proof for $SePoSh(r)$ follows similarly. If $dU(x^*(r),r_0)/dr < 0$, then $SoPoSh(r) > 1$, and so on. We have: $dU(x^*(r),r_0)/dr = \partial U(x^*(r),r_0)/\partial x^*(r) \times dx^*(r)/dr$. Recall that $U(x^*(r),r_0) = x^*(r)G_{ne}(x^*(r),r_0) + (1 - x^*(r))G_{nn}(x^*(r),r_0)$, hence:

$$\frac{\partial U(x^*(r),r_0)}{\partial x^*(r)} = [G_{ne}(x^*(r),r_0) - G_{nn}(x^*(r),r_0)] \\ + x^*(r)\frac{\partial G_{ne}(x^*(r),r_0)}{\partial x^*(r)} + (1 - x^*(r))\frac{\partial G_{nn}(x^*(r),r_0)}{\partial x^*(r)}.$$

The first term is positive as a consequence of Lemma 2. Note that in the proof of Lemma 2, the specific value of r was irrelevant. The second and third terms are also positive (at least one of them strictly positive), as for any r_0, we have $\partial G_{nn}(x,r_0)/\partial x$, $\partial G_{ne}(x,r_0)/\partial x \geq 0$. The claim now follows from Proposition 6. □

5 Conclusion

We present a new analytical model for the adoption of security measures by autonomous systems, particularly in the context of asymmetric, bidirectional security measures. Our analysis provides insights into equilibrium and stability of adoption levels as well as policy perspectives on socially optimal levels of adoption, optimal level of egress filtering with respect to different measures of interest and planner's jurisdiction, and price of shortsightedness. We highlight several interesting results, including the dependence of the equilibrium adoption on the initial seeding, how performance improvements in blocking outbound threats can decrease the overall adoption of the security measure due to free-riding, that socially optimal solution increases *every* ISP's utility, that imposing minimum levels of egress filtering as a stand-alone means of policy is futile, and that shortsightedness of ISPs is detrimental to security.

In another work of ours [27], we investigated the effectiveness of introducing a honeypot-based monitoring and penalizing mechanism on the outbound threat activities of the ISPs. We also showed the undermining effects of a locally restricted authority of a planner in the face of the free-riding tendencies of the unregulated ISPs. We also touched on the impact of the heterogeneity of the ISPs (their shortsightedness, costs and subnet sizes) in their adoption decisions and the policy-making implications. In the future, we intend to extend this analytical model to consider other sources of externalities besides egress filtering, e.g., indirect costs, epidemic propagation, etc. We will also consider non-atomic scenarios of interactions and the role of information for a more accurate analysis. We also would like to incorporate the competition and migration of customers of the ISPs based on the quality of security service they receive. Last but not least, we intend to estimate the parameters of our models based on real-world data and cross-check our qualitative findings with practice.

References

1. Cashell, B., of Congress. Congressional Research Service, L.: The economic impact of cyber-attacks, Congressional Research Service, Library of Congress (2004)

2. Cavusoglu, H., Mishra, B., Raghunathan, S.: A model for evaluating it security investments. Communications of the ACM **47**(7) (2004) 87–92
3. Van Eeten, M., Bauer, J., Groenewegen, J., Lemstra, W.: The economics of malware, TPRC (2007)
4. Council, N.S.: The Comprehensive National Cybersecurity Initiative (2010) http://www.whitehouse.gov/sites/default/files/cybersecurity.pdf.
5. Grossklags, J., Christin, N., Chuang, J.: Secure or insure?: a game-theoretic analysis of information security games. In: Proc. of the 17th conference on World Wide Web, ACM (2008) 209–218
6. Johnson, B., Grossklags, J., Christin, N., Chuang, J.: Uncertainty in interdependent security games. Decision and Game Theory for Security (2010) 234–244
7. Lelarge, M.: Coordination in network security games. In: IEEE INFOCOM. (2012)
8. Böhme, R., Kataria, G.: Models and measures for correlation in cyber-insurance. In: Economics of Information Security. (2006)
9. d'Onofrio, A., Manfredi, P., Salinelli, E.: Vaccinating behaviour, information, and the dynamics of SIR vaccine preventable diseases. Theoretical population biology **71**(3) (2007) 301–317
10. Bauch, C., Earn, D.: Vaccination and the theory of games. Proceedings of the National Academy of Sciences of the United States of America **101**(36) (2004) 13391
11. Heal, G., Kunreuther, H.: The vaccination game. Center for Risk Management and Decision Process Working Paper (05-10) (2005)
12. Reluga, T., Galvani, A.: A general approach for population games with application to vaccination. Mathematical Biosciences (2011)
13. Bauer, J.M., van Eeten, M.J.: Cybersecurity: Stakeholder incentives, externalities, and policy options. Telecommunications Policy **33**(10) (2009) 706–719
14. Hofmeyr, S., Moore, T., Forrest, S., Edwards, B., Stelle, G.: Modeling internet-scale policies for cleaning up malware. In: Economics of Information Security and Privacy III. Springer (2013) 149–170
15. Clayton, R.: Might governments clean-up malware? Communication and Strategies (81) (2011) 87–104
16. van Eeten, M.J., Bauer, J.M.: Economics of malware: Security decisions, incentives and externalities. Technical report, OECD Publishing (2008)
17. Moore, T., Clayton, R.: The impact of incentives on notice and take-down. In: Managing Information Risk and the Economics of Security. Springer (2009) 199–223
18. Moore, T., Clayton, R.: The consequence of non-cooperation in the fight against phishing. In: eCrime Researchers Summit, 2008, IEEE (2008) 1–14
19. Van Eeten, M., Bauer, J., Asgharia, H., Tabatabaie, S., Rand, D.: The role of internet service providers in botnet mitigation an empirical analysis based on spam data, TPRC (2010)
20. Lichtman, D., Posner, E.: Holding internet service providers accountable. Supreme Court Economic Review (2006) 221–259
21. Bodoh-Creed, A.: Approximation of large games with applications to uniform price auctions. In: Auctions, Market Mechanisms, and Their Applications. Springer (2012) 54–54
22. Adlakha, S., Johari, R., Weintraub, G.: Equilibria of dynamic games with many players: Existence, approximation, and market structure. Approximation, and Market Structure (November 2011) (2011)
23. Hofbauer, J., Sandholm, W.H.: Stable games and their dynamics. Journal of Economic Theory **144**(4) (2009) 1665–1693
24. Sandholm, W.H.: Population games and evolutionary dynamics. MIT press (2011)
25. Moore, T., Clayton, R.: Examining the impact of website take-down on phishing. In: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, ACM (2007) 1–13

26. Khouzani, M., Sen, S., Shroff, N.B.: Managing the Adoption of Asymmetric Bidirectional Firewalls: Seeding and Mandating. In: IEEE GLOBECOM, Anaheim, December 3-7. (2012)
27. Khouzani, M., Sen, S., Shroff, N.B.: An Economic Analysis of Regulating Security Investments in the Internet. In: IEEE INFOCOM, Turin, Italy, April 14-19. (2013)