An Economic Analysis of Regulating Security Investments in the Internet

MHR. Khouzani
The Ohio State University
Email: khouzani@ece.osu.edu

Soumya Sen
Princeton University
Email: soumyas@princeton.edu

Ness B. Shroff
The Ohio State University
Email: shroff@ece.osu.edu

***Abstract*—Regulating the ISPs to adopt more security measures to protect their subnet has been proposed as an effective method in mitigating the threat of attacks in the Internet. However, selfish economic incentives of the ISPs may lead to under-investment in the adoption of such measures. We study the potential gains in a network's social utility when a regulator implements a monitoring and penalizing mechanism on the outbound threat activities of autonomous systems (ASes). We then show how free-riding can render regulations futile if the subset of ASes under the authority of the regulator is smaller than a threshold. Finally, we show how heterogeneity of the ASes affect the responses of the ISPs to regulations and discuss how the regulator can leverage such information to improve the overall security of the network.**

I. Introduction

Regular security breaches and intrusions into Governmental, financial, and personal computer systems by organized criminals, nation-states, and hackers, continue to plague the Internet. [1] reports that "According to the The Chief Security Officer of AT&T, cybercrime yielded $1 trillion in annual revenue. This would put cybercrime on course to be close to 2 percent of the global economy, larger than the entire pharmaceutical industry." Clearly, each ISP needs to invest more to improve network security. In fact, as the authors in [2] show, adoption of combined egress and ingress filtering by only the top 0.2% ISPs (in terms of their size) may be able to decrease the "total wicked traffic rate" by more than 40%.

Although technological solutions, such as firewalls, deep-packet inspection (DPI), quarantine measures, secure protocol and encryption, are available for deployment today, economic challenges due to free-riding hurt the adoption of security measures. Adoption of security measures entails an initial investment as well as recurring maintenance costs. It can also cause performance degradation as a result of false positives and latencies introduced by traffic monitoring. Consequently, ASes have perverse incentives to free-ride on the externalities generated from security investments by others. Regulators can implement mechanisms such as spam ranking [3], [4] and monetary penalties to promote adoption of security measures, in particular, monitoring of outbound threat activities in egress traffic from ASes. This work provides an analytical framework that quantifies the benefits of such regulatory mechanisms in improving the overall network security, and also identifies scenarios when such regulations may be rendered ineffective.

The key contributions of this paper are as follows:

- In Section III, we propose using monitoring mechanisms (e.g., honeypots) and penalizing (monetarily or reputation-wise) the ASes based on their outgoing threats. We specifically show that this scheme can improve the social utility as well as the overall security of the network. Further, if the penalties are monetary and the collected fines are redistributed among ASes, this scheme can improve even the utilities of the individual ASes.
- In Section IV, we investigate an important problem in implementing security related regulations: that any regulator has jurisdiction over only a subset of ASes in the Internet. We show that if the jurisdiction domain of the regulator is smaller than a threshold, then the global as well as regional objectives of the regulations is compromised due to free-riding by ASes in the unregulated domain.
- Finally, in Section V, we consider the role of heterogeneity in ASes' cost on the regulation policies. Specifically, we show how the regulator can leverage information on heterogeneity to improve the effectiveness of the regulations.

Related Literature: A considerable amount of previous research work has focused on the adoption of security measures with game-theoretic models of vaccination games [5], [6] and security investments [7], epidemic diffusion [8], and adoption of technologies with network externalities [9], and cyber-insurance [10]. The primary modeling considerations have been the role of externalities and economic incentives in the adoption of security measures for *ingress* filtering by autonomous systems (AS) to protect their network customers.

But a paradigm shift is now underway with both researchers and regulators realizing the value of preempting security threats (e.g, botnets, hacking, spam) through *egress* filtering [11], reputation systems, and the use of *bi-directional* firewalls [12]. But selfish behavior on the part of ASes, coupled with a lack of general understanding of the externality benefits of such egress filtering, threaten to undermine any such effort. Regulators will therefore need mechanisms to monitor and penalize ASes based on their outbound threat activities. Accounting for these aspects in economic models for security adoption and regulatory effectiveness is ever more important in view of the recent US Government's CNCI [13] policy of conducting *"real-time inspection and threat-based decision-making on network traffic entering or leaving executive branch networks,"* developing analytical frameworks is ever more important. To address this issue, our work specifically accounts for this new paradigm in (a) creating an analytical framework

that models the benefits that such regulatory mechanisms can have in improving overall network security, while also (b) identifying scenarios when such regulations are rendered ineffective (i.e., the overall network security is shown to improve only if the fraction of ASes over which a regulator has control is above a certain threshold).

II. Modeling

In this section, we provide an overview of the model that we developed in [12]. Our aim is to develop a qualitative investigation and identify phenomena that can shape the adoption of security measures in the internet at the level of Autonomous Systems (ASes), and understand the policies that can influence it. Accordingly, we make some technical assumptions to keep the model analytically tractable. This model captures key attributes of the adoption process of the security measures, and at the same time, is simple enough to facilitate analytical investigations.

In our context, the term *security measure* is general and can refer to installing (hardware/software) firewalls for *egress* (outbound) and *ingress* (inbound) filtering, using Access Control Lists (ACLs), blacklisting the attackers, quarantining the bots, using encryption, disallowing certain types of traffic (e.g. peer-to-peer), adopting a stricter terms of service that may deter potential attackers, etc. Adopting a security measure reduces the rate of inbound and outbound intrusion attempts (potentially differently). For now, we consider a single monolithic security measure accessible to all ASes.

We consider a continuous-time model with a network of N inter-connected ASes. Once an AS purchases the security measure, it may be able to un-adopt it by disabling it in order to avoid the recurrent costs the security measure (maintenance, false positives, network slowdown, privacy conflicts, etc.). Subsequent adoptions are performed by enabling the security measures, and in particular, do not entail paying its one-time purchase fee. Hence, for an AS that has obtained the security measure, the cost of a subsequent adoption only includes the recurrent usage cost. Hence, we need a model that distinguishes between the first adoption and subsequent re-adoptions. To do this, we introduce three different types of ASes: (1) ASes that have *obtained* and *enabled* the security measure; (2) ASes that have *not obtained* it; and (3) ASes that have *obtained* the security measure but have *disabled* it. Note that *obtaining* the security measure can be through purchasing it, or by being *seeded* for free by either regulators or vendors in an attempt to influence the equilibrium [12]. We will denote the *fraction* of ASes of each type at time t by $x(t)$, $y(t)$ and $1-x(t)-y(t)$, respectively. The adoption state of the network at time t is hence represented by the pair $(x(t),y(t))$. Table-I contains a list of the important notations of the model.

Each AS independently re-evaluates the rate of intrusion attempts on its subnet and accordingly updates its decision regarding the adoption of the security measures. These re-evaluations occur epoch of i.i.d. Poisson processes of rate γ. We assume that the decisions of each AS is its best response to the *current* measure of the intrusion rates, that is, assuming the current measure is not going to change. The decision of the ASes is determined by comparing the expected utilities given each decision. Accordingly, we define three utilities: $G_0(x)$, $G_1(x)$ and $G_2(x)$: Given the current level of adoption x, $G_0(x)$ is the expected utility of an AS that does not have the security measure and it decides to stay unadopted, hence it only includes the expected costs of future intrusions. The expectation is taken over the random intrusions over time while assuming the level of adoption is fixed at x. $G_1(x)$ is the expected utility of an AS that does not have the security measure if it decides to *purchase* and enable it, hence, $G_1(x)$ includes the purchase fee and the recurrent costs of security measure as well as the expected costs of future intrusions that can bypass the security measure. Finally, $G_2(x)$ is the expected utility of an AS that already has the security measure if it decides to enable it. Note that $G_1(x)$ and $G_2(x)$ differ only in the purchase fee of the security measure. Specifically, $G_2(x) = G_1(x) + C_0$, where C_0 denotes the (one-time) purchase fee of the security measure (also c.f. Table-I).

As we mentioned before, security measures may incur recurrent usage costs: they need to be routinely maintained and updated; they can slow down the connection through latencies introduced by traffic monitoring, hence compromising the quality of service provided by the AS; and last but not least,

TABLE I
Main notations in the model

parameter	definition
$x(t)$	Fraction of the ASes at time t that have obtained and enabled the security measure.
$y(t)$	Fraction of the ASes at time t that are yet to obtain it.
γ	Rate at which each AS updates its adoption decision.
$G_0(x)$	Expected utility of a non-adopter AS. $G_0(x)$ includes intrusion costs only.
$G_1(x)$	Expected utility of an AS that purchases and enables the security measure. $G_1(x)$ includes (reduced) intrusion costs along with the purchase fee and recurrent costs of the security measure.
$G_2(x)$	Expected utility of an AS that just enables its security measure. $G_2(x)$ does not include the purchase fee.
Λ	Rate of intrusion attempts on an AS in the absence of the security measure in the network.
μ	Recovery rate after a successful intrusion.
C_0	One-time purchase fee of the security measure.
c	Per unit time usage cost of the security measure.
K_0	Instantaneous cost upon a successful intrusion.
k	Cost (loss/damage) per unit time of intrusion.
r	Discount factor of the ASes.
L	Constant defined as $L := \frac{\Lambda}{\mu r}(K_0(\mu+r)+k)$
ξ	The penalty imposed on an AS by the regulator for each detected intrusion originated from its subnet (§III)
$U(x)$	Social utility: summation of the utilities of all ASes.
$S(x)$	Aggregate security utility: negative of summation of the intrusion costs over all ASes.

security measures have a rate of *false positives*, that is, they occasionally block legitimate traffic. Let c be the cost per unit time of using the security measure incurred by an adopter AS. For simplicity of exposition, we consider security breaches that do not propagate in the network. For example, we will *not* consider attacks involving self-replicating malicious codes (known as *worms*) in this 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 AS is compromised, the AS incurs an instantaneous cost of K_0 and a per unit cost of k that persists as long as the host is infiltrated by that specific hacker. We assume that the costs of different intrusions add up, e.g., two concurrent successful intrusions from two independent intruders incur the AS $2k$ cost per unit time. The instantaneous cost may reflect the losses due to exposure of private information such as credentials (e.g., fingerprints, voice recognition, passwords), credit card information, or manipulation of data. On the other hand, the per unit time cost can represent the accumulation of eavesdropped data such as keystroke logs, accessing the network at the cost of the victim, slowdown of the victim's machine or the AS's service, etc. The time it takes to remove an intrusion is according to an exponential random variable with rate μ. This is the time it takes for the CSIRT (Computer Security Incident Response Team) of an AS to detect and block the intrusion. We assume that the machines are again susceptible to future attacks, since new attacks are likely to exploit new techniques.

New security breaches can originate from the subnet of any of the ASes. For now, we assume that ISPs are homogeneous, that is, they assign the same parameters for costs and have similar subnet sizes, and that a target of an intrusion is chosen uniformly randomly from the space of IP addresses. In §V, we discuss the heterogeneous cases.

Security measures can affect the success of both outbound (egress) and inbound (ingress) threats. Accordingly, the success probability of an intrusion attempt depends in part on the status of the AS of the attacker as well as the AS of the target with regard to the adoption of the security measure.[1] Specifically, the *highest* chance of intrusion success pertains to the case in which *neither* of the (origin and the target) ASes have an active security measure (i.e., both are *exposed*), while the *lowest* likelihood is when *both* (the origin and the target) ASes have (obtained and) enabled the security measure (i.e., both are *protected*). Based on the *four* different conditions for the adoption status of the ASes of an attacker and its target, we define intrusion *success probabilities* π_0, π_1, Π_0 and Π_1 according to Table-II. Namely, π_1 is the success probability of an intrusion if the AS of the intruder's origin as well as the AS of the target user have active security measures in place, π_0 is the success probability of an intrusion if only the target user's AS has adopted the security measure, and so forth.

Without loss of generality, we take $\Pi_0 = 1$, considering only the attempts that are successful in the absence of the security measure. However, we continue to use the *notation* Π_0 in our formulation for presentation purposes. Based on the sensible meaning of a security measure, as a mechanism to deter successful intrusions, we have the following natural inequalities:

TABLE II
SUCCESS PROBABILITIES OF AN INTRUSION ATTEMPT

		Target (destination) AS	
		Protected	Not Protected
Originating AS	Protected	π_1	Π_1
	Not Protected	π_0	Π_0

$$0 \leq \pi_1 \leq \min\{\pi_0, \Pi_1\} \leq \max\{\pi_0, \Pi_1\} \leq \Pi_0 = 1. \quad (1)$$

The inequality just states that the success probability of an intrusion that has to bypass the security measures of both the AS of its own subnet and that of the target node is the smallest (π_1). The next probability in order, is the smaller of π_0, Π_1, depending on which protection is stronger: ingress or egress, respectively. The highest probability of success (Π_0) is pertinent to the case in which the intrusion is not confronted with any security measure in either of the ASes.

A successful intrusion has to bypass the security measure of its own AS, *and* the security measure of the AS of the target machine, when both ASes are adoptees. For a security measure whose mechanism of intrusion detection and prevention is only signature-based, rule-based, or blacklisting, if both ASes 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 ASes 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:

$$\pi_1 = \pi_0\Pi_1 + \alpha(\min\{\pi_0, \Pi_1\} - \pi_0\Pi_1), \text{ for an } \alpha \in [0,1] \quad (2)$$

Note that the *mutually inclusive* and *mutually independent* cases are retrieved for $\alpha = 1$ and 0 respectively.

Definition. We call the security measures that follow the structural equation of (2) "*non-cooperative*".[2]

For the rest of the paper, we consider "non-cooperative" security measures as defined above.

Let Λ represent the rate of intrusion attempts on an AS in the absence of any security measure in the network. The

[1]Note that we assume that the routed traffic is not monitored for threats and the only traffic monitoring for threats occur at border (edge) ASes.

[2]For cooperative schemes, as exampled in [14], it is possible for π_1 to be less than $\pi_0\Pi_1$.

rate of *successful* intrusion attempts on an AS 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 AS, 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, their success probability is Π_0. Similarly, the rate of *successful* intrusion attempts on an AS that *has* an enabled security measure is $\Lambda(x\pi_1+(1-x)\pi_0)$. These are the two rates that each AS can readily measure, then calculate its conditional utilities and accordingly make an adoption decision. Note specifically that the ASes need not observe the value of x or Λ directly.

The utility of an AS is a decreasing function of the costs due to future intrusions to its subnet (as well as the costs of security investments). For ease of calculation, we assume the ASes are risk-neutral [15]. Hence, we can directly take the negative of the costs incurred by an AS to be the its utility.

Based on the model described above, as is detailed in [12], given the current adoption level x, the expected utility of a non-adopting AS (i.e., $G_0(x)$) and the (net) expected utility of the ASes that purchases and enables the security measure (i.e., $G_1(x)$) are analytically derived as follows:

$$\begin{aligned} G_0(x) &= -L\left(\Pi_0 - x(\Pi_0-\Pi_1)\right) \\ G_1(x) &= -C_0 - \frac{c}{r} - L\left(\pi_0 - x(\pi_0-\pi_1)\right), \end{aligned} \tag{3}$$

where $L := \frac{\Lambda}{\mu r}\left(K_0(\mu+r)+k\right)$ (also in Table I).

A straightforward yet important property of the expected utilities is that all of them are increasing in the level of adoption. Hence, positive externality exists for adopters and non-adopters alike:

Lemma 1. *(A) For any* $x \in [0,1]$ *we have:* $\frac{\partial G_0(x)}{\partial x}$, $\frac{\partial G_1(x)}{\partial x}$, $\frac{\partial G_2(x)}{\partial x} \geq 0$. *The equality holds only if* $\Pi_1 = \Pi_0$.
(B) When $\Pi_1 < \Pi_0$, $\frac{\partial}{\partial x}(G_1(x)-G_0(x)) < 0$ *at any* $x \in [0,1]$.

Part (B) of the lemma states that *even though both adopter and non-adopters experience positive externalities of the security measure adopted by others, the non-adopters benefit more*. This is what creates the problem of free-riding.

III. Regulation through Monitoring

Despite the huge potential of abating outbound threats in order to improve the general security of the Internet [2], [11], ASes are generally reluctant to make such investments. One mechanism to provide the necessary incentives for ASes to invest in such protections is through a monitory scheme: the regulator can set up traps to trace the threats and penalize the ASes where the attacks originate from. These penalties can either be monetary, and/or as [3], [4] suggest, the reputation damage of the ASes as a result of public announcement of its pollution ranking and the loss of business as a result of that.

Tracing malicious activities can be done using *honeypots*. A honeypot is a trap for unauthorized or malicious access: it consists of a network site that appears to be part of a network, but is actually isolated, and can log and trace the intrusions (ref. e.g. [16]). Honeypots have been used in research to investigate the attacks in the Internet, and on a smaller scale, by organizations in their internal networks as a means to elevate their overall state of security [17].

Let the penalty for each trapped intrusion originating from the subnet of an AS be denoted by ξ. Further, let Λ_0 be the rate of intrusion attempts to the honeypot from an unprotected AS. Λ_0 is related to the intrusion attempts on an AS as follows: $\Lambda_0 = \Lambda\beta/N$ where Λ was the rate of intrusion attempts on each AS from *all* ASes (hence the division by N) if they were all unprotected, and β is the relative size of the honeypot compared to a subnet of an AS. Then an AS that does (does not) adopt the security measure is charged a rate of $\Lambda_0\xi\Pi_1$ ($\Lambda\xi\Pi_0$) over time by the regulator. This is because the honeypot is (intentionally) unprotected, hence, the rate of successful intrusions to the honeypot is discounted by Π_1 and Π_0 for AS with and without the security measure, respectively (ref. Table-II). The contingent utilities will therefore change from (3) to the following:

$$\begin{aligned} G_0(x,\xi) &= -L\left(\Pi_0 - x(\Pi_0-\Pi_1)\right) - \Lambda_0\xi\Pi_0/r \\ G_1(x,\xi) &= -C_0 - \frac{c}{r} - L\left(\pi_0 - x(\pi_0-\pi_1)\right) - \Lambda_0\xi\Pi_1/r. \end{aligned}$$

By definition, an *equilibrium* is the (asymptotic) level of adoption that will stay unchanged over time. To indicate the dependence on ξ, we will designate the equilibrium as $x^*(\xi)$. Following this notation, the equilibrium point in the absence of the honeypot is written as $x^*(0)$. For brevity, we will represent $G_i(x^*(\xi),\xi)$ by $G_i(x^*(\xi))$ for $i=0,1$. The new equilibrium point $x^*(\xi)$ satisfies: $G_0(x,\xi)=G_1(x,\xi)$. Noting $G_0(x,\xi) = G_0(x,0)-\Lambda_0\xi\Pi_0/r$ and $G_1(x,\xi)=G_1(x,0)-\Lambda_0\xi\Pi_1/r$, we have: $G_0(x^*(\xi),0)-G_1(x^*(\xi),0)=\Lambda_0\xi(\Pi_0-\Pi_1)/r$. This, along with Lemma 1-A, yield the following:

Proposition 2. *For* $\Pi_1<\Pi_0$ *and* $x^*<1$, $dx^*(\xi)/d\xi > 0$.[3]

In words, the introduction of the monitory-based regulation increases the fraction of ASes that adopt the security measure, and raising the penalty increases the level of adoption. In what follows, we investigate less straightforward questions: how does the introduction of the monitoring and penalties affect the *social utility* and the individual utilities of the ASes? The social utility is defined as the summation of the utilities of all of the ASes and denoted by U. Since the utility of both adopters and non-adopters are increasing in the level of adoption, the answer is non-trivial. It is not difficult to show that if the collected penalties are not included in the social utility, then the introduction of the honeypots decreases the social utility. If, in contrast, the penalties are monetary and the collected fines are considered in the social utility, we have:

$$\begin{aligned} U/N =& x^*(G_1(x^*,0)-\Lambda_0\xi\Pi_1/r)+(1-x^*)(G_0(x^*,0)-\Lambda_0\xi\Pi_0/r) \\ &+x^*\Lambda_0\xi\Pi_1/r+(1-x^*)\Lambda_0\xi\Pi_0/r \\ =& x^*G_1(x^*,0)+(1-x^*)G_0(x^*,0) \end{aligned}$$

where $x^*(\xi)$ is replaced with x^* for brevity. Now, note that for $\Pi_1<\Pi_0$, following Proposition 3 and Lemma 1-A, we have: $G_1(x^*(\xi),0), G_0(x^*(\xi),0) > G_0(x^*(0)) = G_1(x^*(0)) =$

[3]Note that when $x^*(\xi)=1$ for any ξ, then increasing ξ further does not affect x^* and yields no benefit.

$U(x^*(0))/N$. This leads to the following:

Proposition 3. *For $\Pi_1 < \Pi_0$ and $x^* < 1$, $dU(x^*(\xi))/d\xi > 0$.*

The proposition testifies that the introduction of the monitoring scheme not only increases the level of adoption of the security measure (and hence improves the security of the network), but also increases the social utility of the ASes if the collected fines are considered part of the social utility. The inclusion of the (monetary) penalties in the social utility can be implemented by investment in infrastructure that equally benefits all ASes, or just directly redistributing the funds among the ASes. In such cases, a question can be whether the utility of the individual ASes increases as well. Denoting $x^*(\xi)$ by x^*, the utility of an adopter is computed as:

$$\begin{aligned} G_1(x^*) =& G_1(x^*,0) - \Lambda_0\xi\Pi_1/r + \Lambda_0\xi(x^*\Pi_1 + (1-x^*)\Pi_0)/r \\ =& G_1(x^*,0) + \Lambda_0\xi(1-x^*)(\Pi_0 - \Pi_1)/r. \end{aligned}$$

The second term in the last expression is nonnegative. Also from Proposition 2 and Lemma 1-B, if $\Pi_1 < \Pi_0$ and $x^*(0) < 1$, then we have $G_1(x^*(\xi),0) > G_1(x^*(0))$, and hence, $G_1(x^*(\xi)) > G_1(x^*(0))$ for $\xi > 0$. Note that following the definition of $x^*(\xi)$, we have: $G_0(x^*(\xi)) = G_1(x^*(\xi))$ (including for $\xi = 0$). Therefore, the same conclusion applies to the non-adopters, and we have the following result:

Proposition 4. *If $\Pi_1 < \Pi_0$ and $x^*(0) < 1$, then for $\xi > 0$ we have $G_0(x^*(\xi)) > G_0(x^*(0))$, and $G_1(x^*(\xi)) > G_1(x^*(0))$.*

In words, adopter and non-adopter ASes of the security measure are both individually better-off (i.e., yield higher individual utilities) with the introduction of the monitoring scheme and the redistribution of the penalties.

IV. Regionally Restricted Jurisdiction

A problem with regulating the ASes is that no entity has full jurisdiction over the entire Internet. In this section, we investigate the impact of partial regulation, i.e., what happens when the authority domain of a regulator is restricted to only a subset of the ASes in the network. Specifically, we show how this can lead to "free-riding" of the ASes in the unregulated region, which in turn compromises the efficacy of the regulation and, in some cases, even undercuts the objective of regulation.

Suppose that the regulator has jurisdiction over a fraction f of the ASes. We refer to the subset of the ASes under the regulator's authority as Region A, and the rest of the ASes as Region B. Region A can be one country or a confederation of countries. Let x represent the fraction of the ASes that (belong to Region B and) choose to adopt the security measure based on their selfish preference. Figure 1 depicts a schematic representation of the setting in this section. In general, the regulator can also enforce a different protection on outgoing traffic compared to the ASes in Region B. To model this, we use these four new notations: Π_{1A}, Π_{1B}, π_{1A} and π_{1B}. Similar to the model in §III, Π_{1A} is the success rate of intrusions that originate from a protected AS in Region A and the destination's AS is unprotected, and π_{1A} is the success rate of intrusions that originate from a protected AS in Region A and the target's AS is protected. Π_{1B} and π_{1B} are defined in identical manner by replacing Region A with Region B. More protection against outgoing threats translates to lower values of Π_{1A} and Π_{1B}. Similar inequalities as in (1) hold:

$$0 \le \pi_{1A} \le \min\{\pi_0, \Pi_{1A}\}, \quad 0 \le \pi_{1B} \le \min\{\pi_0, \Pi_{1B}\}. \quad (4)$$

Also, as before, we consider "non-cooperating" security measures, that is, we have:

$$\pi_{1A} = \pi_0\Pi_{1A} + \alpha(\min\{\pi_0, \Pi_{1A}\} - \pi_0\Pi_{1A}), \text{ for an } \alpha \in [0,1], \quad (5)$$

and likewise for π_{1B}. This in turn implies:

$$(\Pi_0 - \Pi_{1A}) - (\pi_0 - \pi_{1A}) \ge 0 \quad (\Pi_0 - \Pi_{1B}) - (\pi_0 - \pi_{1B}) \ge 0. \quad (6)$$

We first investigate the case in which the regulator mandates that a z fraction of the total ASes adopt the security measure. Next, we consider the scenario in which the regulator imposes the level of egress filtering for Region A ASes. Finally, we explore the monitoring scheme from §III in which only ASes in Region A are subject to the penalties of the regulator.

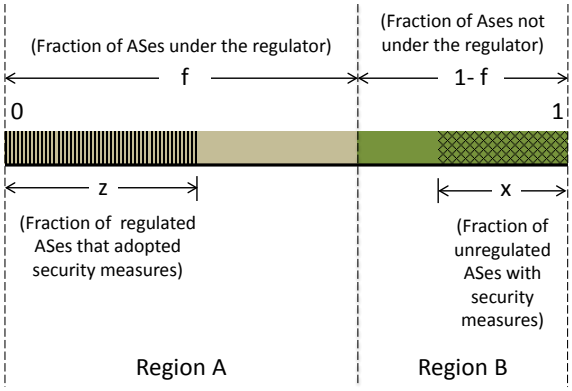

Fig. 1. The regulator has jurisdiction only over ASes in Region A (re: §IV).

A. Mandating ASes to Adopt

We denote the utility of an AS that does not and does adopt the security measure by $G_0(x,z)$ and $G_1(x,z)$, respectively. Note the dependence on both x and z. Also note that this utility is the same irrespective of whether the AS belongs to Region A or B. $G_0(x,z)$ and $G_1(x,z)$ are computed as follows:

$$\begin{aligned} G_0(x,z) =& -L\,(z\Pi_{1A} + x\Pi_{1B} + (1-x-z)\Pi_0) \\ G_1(x,z) =& -L\,(z\pi_{1A} + x\pi_{1B} + (1-x-z)\Pi_0) - C_0 - c/r \end{aligned} \quad (7)$$

Let $\bar{z}$ represent the minimum of 1 and the solution of $G_0(0,z) = G_1(0,z)$. Intuitively, $\bar{z}$ is the adoption level of ASes under regulation (Region A) for which *none* of the unregulated ASes (Region B) will adopt the security measure and will all free-ride. Also, define $\underline{z}$ as the maximum of zero and the solution of $G_0(1-f,z) = G_1(1-f,z)$. The interpretation of $\underline{z}$ is that for adoption levels of regulated ASes below $\bar{z}$, *all* of the unregulated ASes adopt the security measure and none will free-ride. The equilibrium fraction of ASes that choose to adopt the security measure, i.e., x^*, satisfies

$G_0(x^*,z) = G_1(x^*,z)$ as long as $\underline{z} < z < \bar{z}$. Note that if $z > \bar{z}$, then $x^* = 0$ and we can have $G_0(x^*,z) > G_1(x^*,z)$, and for $z < \underline{z}$, then $x^* = (1-f)$ and we can have $G_0(x^*,z) < G_1(x^*,z)$. We will use $x^*(z)$ to refer to the equilibrium level of adoption of Region B ASes to indicate its dependence on z. The first evident result is the "free-riding" of the Region B ASes:

Proposition 5. *If* $\Pi_{1A} < \Pi_0$*, then* $\dfrac{dx^*}{dz} < 0$ *as long as* $\underline{z} < z < \bar{z}$.

In words, if more ASes are mandated to adopt the security measure in Region A, then less ASes in Region B will end up adopting it. The proof follows.

Proof: For $\underline{z} < z < \bar{z}$, we have: $G_0(x^*,z) = G_1(x^*,z)$. Taking the derivative of this equation with respect to z yields:

$$\frac{\partial G_0(x^*,z)}{\partial x^*} \times \frac{dx^*}{dz} + \frac{\partial G_0(x^*,z)}{\partial z} = \frac{\partial G_1(x^*,z)}{\partial x^*} \times \frac{dx^*}{dz} + \frac{\partial G_1(x^*,z)}{\partial z}$$
$$\Leftrightarrow \frac{dx^*}{dz} = -\frac{\frac{\partial}{\partial z}G_0(x^*,z) - \frac{\partial}{\partial z}G_1(x^*,z)}{\frac{\partial}{\partial x^*}G_0(x^*,z) - \frac{\partial}{\partial x^*}G_1(x^*,z)}.$$

Replacing from (7) yields:

$$\frac{dx^*}{dz} = -\frac{(\Pi_0 - \Pi_{1A}) - (\pi_0 - \pi_{1A})}{(\Pi_0 - \Pi_{1B}) - (\pi_0 - \pi_{1B})}. \quad (8)$$

The proposition now follows from inequalities (6). When $z > \bar{z}$,[4] x^* remains at zero. Likewise when $z < \underline{z}$, x^* stays at $1-f$. ■

The above proposition suggests that when $z < \bar{z}$, the "beneficial" effects of increasing the fraction of the ASes under the regulator's jurisdiction may be compromised by the free-riding of the ASes in the unregulated region. In what follows, we formally investigate this effect taking into account different metrics of assessment.

We will refer to the sum of the utilities of all of the ASes (in both regions) as the *total social utility* and we will denote it by U_t. If the total number of ASes is N, then U_t is computed as $U_t = N(x^*+z)G_1 + N(1-x^*-z)G_0$. When $\underline{z} < z < \bar{z}$, we have $G_0(x^*,z) = G_1(x^*,z)$, hence:

$$U_t/N = (x^*+z)G_0 + (1-x^*-z)G_0 = G_0 = G_1.$$

We can now investigate the effect of changing z on U_t:

$$\frac{d}{dz}U_t/N = \frac{d}{dz}G_0 = \frac{d}{dz}G_1 = \frac{\partial G_0(x^*,z)}{\partial z} + \frac{\partial G_0(x^*,z)}{\partial x^*} \times \frac{\partial x^*}{\partial z}$$
$$= L(\Pi_0 - \Pi_{1A}) - L(\Pi_0 - \Pi_{1B}) \times \frac{dx^*}{dz}$$

Replacing from (8) and simplifying, we obtain:

$$\frac{d}{dz}U_t = NL\frac{(\Pi_0 - \Pi_{1B})(\pi_0 - \pi_{1A}) - (\Pi_0 - \Pi_{1A})(\pi_0 - \pi_{1B})}{(\Pi_0 - \Pi_{1B}) - (\pi_0 - \pi_{1B})}. \quad (9)$$

It is straightforward to verify that the RHS of the above equation simplifies to zero after replacing from (5). Hence, we have the following:

Proposition 6. $\dfrac{dU_t}{dz} = \dfrac{dG_0}{dz} = \dfrac{dG_1}{dz} = 0$ *for* $\underline{z} < z < \bar{z}$.

An important consequence of this observation is that for mandating to be *effective* (in the sense of improving the social utility of the network), the regulator must have jurisdiction over a large enough subset of the ASes, specifically, we must have $f \geq \bar{z}$. When $z < \underline{z}$, U_t increases with z. When $z > \bar{z}$, then: $U_t = -NzL(\pi_{1A} + \pi_0(1-z)) - Nz(C_0 + c/r) - N(1-z)L(z\Pi_{1A} - (1-z)\Pi_0)$, that is maximized at some point $z_1 \geq \bar{z}$.

Another measure of interest can be the security of the overall network. For a measure of the overall security, we define S_t to be the aggregate costs of the ASes due to intrusions. (c.f. Table I). Note the difference between S_t and U_t: unlike U_t, S_t does not involve the costs of adopting the security measure. Specifically, we have: $S_t = N(x^*+z)(G_1 + C_0 + c/r) + N(1-x^*-z)(G_0)$. When $\underline{z} < z < \bar{z}$, we have $G_0(x^*,z) = G_1(x^*,z)$. Therefore, for $\underline{z} < z < \bar{z}$, $S_t = NG_0 + N(x^*+z)(C_0 + c/r)$. When $\underline{z} < z < \bar{z}$, from Proposition 6, we have $U_t/N = G_0 = G_1$ and $dU_t/dz = 0$. Hence, for such cases, we have: $dS_t/dz = N(C_0 + c/r)(dx^*/dz + 1)$. Replacing from (8) and simplifying, we obtain:

$$\frac{dS_t}{dz} = N(C_0 + c/r)\frac{(\Pi_{1A} - \Pi_{1B}) - (\pi_{1A} - \pi_{1B})}{(\Pi_0 - \Pi_{1B}) - (\pi_0 - \pi_{1B})}$$

Replacing from (5), the resulting expression for dS_t/dz is simplified to $N(C_0 + c/r)\dfrac{\Pi_{1A} - \Pi_{1B}}{\Pi_0 - \Pi_{1B}}$. Hence, we have the following proposition:

Proposition 7. $\operatorname{sgn}(\dfrac{dS_t}{dz}) = \operatorname{sgn}(\Pi_{1A} - \Pi_{1B})$ *for* $\underline{z} < z < \bar{z}$.

A peculiar corollary of the above proposition is that when $\Pi_{1A} < \Pi_{1B}$, that is when the security measure in Region A provides more protection on outgoing traffic than the security measure in Region B, the overall security of the network goes down by mandating more ASes in Region A to adopt, unless the mandated fraction of ASes in above $\bar{z}$.

The regulator may only be interested in the social utility or the security of the regulated region, i.e., Region A. Next, we investigate the effect of regulation on these two metrics. Similar to U_t, we define the *regional social utility* to be the sum of the utilities of the ASes in one region. Hence, the social utility of Region A, denoted by U_A, is computed as $U_A = N(f-z)G_0(x^*,z) + NzG_1(x^*,z)$. For $\underline{z} < z < \bar{z}$, we have $G_0(x^*,z) = G_1(x^*z)$. Therefore, for $\underline{z} < z < \bar{z}$, $U_A = NfG_0(x^*,z)$. The following proposition is hence a direct consequence of Proposition 6:

Proposition 8. $dU_A/dz = 0$ *for* $\underline{z} < z < \bar{z}$.

In words, increasing the adopters in Region A, does *not* improve the *regional* utility either as long as there are ASes in Region B to free-ride. We can also define the regional security utility to be the aggregate costs incurred by the ASes of a region as a result of successful intrusions. The social utility of Region A, denoted by S_A, is hence computed as $S_A = N(f-z)G_0(x^*,z) + Nz(G_1(x^*,z) + C_0 + c/r)$, which for $\underline{z} < z < \bar{z}$ is equal to $NfG_0(x^*,z) + Nz(C_0 + c/r)$. Hence, following Proposition 6, $\dfrac{dS_A}{dz} = N(C_0 + c/r) > 0$ (this is in fact true for any z). That is, despite the potential free-riding of the unregulated ASes, the *regional* security can be improved by mandating more ASes to adopt security measures. This is one

[4] Note that $z > \bar{z}$ implies requiring $f > \bar{z}$ as well.

silver lining for regulation among the plethora of the negative results thus far.

B. Protection Against Outgoing Threats

Another way in which the regulator can try to influence the security of the network is to determine how much protection should be provided against outgoing threats by each AS (the amount of egress filtering). Here, we will investigate the effect of having local jurisdiction with this regulation.

Increasing the protection on outgoing traffic of the ASes in Region A translates to reducing the value of Π_{1A}. An immediate result is the free-riding of the ASes in Region B, i.e., the reduction of adoption level in the unregulated region:

Proposition 9. *$dx^*/d\Pi_{1A} > 0$ for $\underline{z} < z < \bar{z}$.*

The proof is straightforward and omitted for brevity.

How does this free-riding of the unregulated ASes affect the social utility and the total security of the network? Recall that for $\underline{z} < z < \bar{z}$, we have: $G_0(x^*,z) = G_1(x^*,z)$, and hence: $U_t/N = G_0(x^*,z) = G_1(x^*,z)$. Therefore:

$$\frac{1}{N}\frac{dU_t}{d\Pi_{1A}} = \frac{\partial G_0}{\partial \Pi_{1A}} + \frac{\partial G_0}{\partial x^*}\cdot\frac{\partial x^*}{\partial \Pi_{1A}} = \frac{\frac{\partial G_0}{\partial x^*}\frac{\partial G_1}{\partial \Pi_{1A}} - \frac{\partial G_1}{\partial x^*}\frac{\partial G_0}{\partial \Pi_{1A}}}{\frac{\partial G_0}{\partial x*} - \frac{\partial G_1}{\partial x^*}}$$
$$= z\frac{(\Pi_0 - \Pi_{1B})\frac{\partial \pi_{1A}}{\partial \Pi_{1A}} - (\pi_0 - \pi_{1B})}{(\Pi_0 - \Pi_{1B}) - (\pi_0 - \pi_{1B})}$$

For "non-cooperating" security measures, the RHS simplifies to zero. Hence, for such cases and for $\underline{z} < z < \bar{z}$, we have $\frac{dU_t}{d\Pi_{1A}} = 0$. It is now curious to see the impact on the overall security. For $\underline{z} < z < \bar{z}$:

$$\frac{1}{N}\frac{dS_t}{d\Pi_{1A}} = \frac{dG_0}{d\Pi_{1A}} + (C_0 + c/r)\frac{d}{d\Pi_{1A}}(z + x^*) = \frac{dx^*}{d\Pi_{1A}}$$

Hence from Proposition 9, we obtain the interesting result of $\frac{dS_t}{d\Pi_{1A}} > 0$ for $\underline{z} < z < \bar{z}$. Note that increasing Π_{1A} translates to reducing the protection on outgoing traffic. This means that regulating the ASes in Region A to increase protection on outgoing traffic does not improve the social utility, and in fact *hurts* the overall security of the network, as a result of the "free-riding" of the ASes in Region B. As in the previous subsection, one may argue that the metrics of interest for a regulator may only concern Region A. For the regional social utility of Region A, for $\underline{z} < z < \bar{z}$, we have:

$$\frac{1}{N}\frac{dU_A}{d\Pi_{1A}} = f\frac{dG_0}{d\Pi_{1A}} = 0$$

and for the regional security utility of Region A:

$$\frac{1}{N}\frac{dS_A}{d\Pi_{1A}} = f\frac{dG_0}{d\Pi_{1A}} + (C_0 + c/r)\frac{dz}{d\Pi_{1A}} = 0$$

Hence, when the regulator has authority only on a limited subset of the ASes (less than $\bar{z}$ of them), then due to the free-riding of the rest of the ASes, neither the social utility nor the security of even the ASes in its authority region can be improved by enforcing more outbound protection.

C. Regionally Restricted Regulation Through Monitoring

In the previous section, we proposed using honeypots and fining ASes as a means of regulation. However, as with the previous two regulatory mechanisms, it is more likely the case that only a restricted subset of ASes fall under the jurisdiction of the regulator. Here, we investigate this restriction on the efficacy of this policy and make interesting observations.

Here, ASes of both regions are free in their adoption decision. It is just that the ASes in Region A, unlike the rest of the ASes, are subject to charges (monetary or otherwise). Hence, the contingent utilities of the ASes in the two regions will be different. We will use superscripts A and B to indicate the regions. Suppose that initially the system is at equilibrium and the monitoring scheme is introduced post-equilibrium. Let the equilibrium pair before the introduction of the monitoring scheme be $(x^*(0), z^*(0))$ and consider the cases of $\underline{z} < z^*(0) < \bar{z}$. At this point we have: $G_0^A(x^*(0),z^*(0)) = G_0^B(x^*(0),z^*(0)) = G_1^A(x^*(0),z^*(0)) = G_1^A(x^*(0),z^*(0))$. We investigate what happens as the penalty fee is increased from zero. The contingent utilities of the two regions are related as follows: $G_0^A(x,z) = G_0^B(x,z) - \Lambda_0\xi\Pi_0/r$ and $G_1^A(x,z) = G_1^B(x,z) - \Lambda_0\xi\Pi_{1A}/r$. Since G_0^A is now less than G_1^A, more ASes in Region A start to obtain and activate the security measure. This creates a free-riding opportunity for the ASes in Region B. However, note that the ASes in Region B that have already obtained the security measure will disable it only if the utility of not having the security measure enabled, i.e., G_0^B, grows larger than the utility of keeping the (already obtained) security measure enabled, i.e., G_2^B. Therefore, increasing the penalty fee keeps increasing $z^*(\xi)$ without changing $x^*(\xi)$ until the penalty fee is raised to ξ_0 for which $G_0^B(x^*(0),z^*(\xi_0)) = G_2^B(x^*(0),z^*(\xi_0))$. We can compute ξ_0 as follows:

$$G_0^B(x^*(0),z^*(\xi_0)) = G_2^B(x^*(0),z^*(\xi_0)) \Leftrightarrow$$
$$G_0^A(x^*(0),z^*(\xi_0)) + \Lambda_0\xi_0\Pi_0/r = G_1^A(x^*(0),z^*(\xi_0)) + C + \Lambda_0\xi_0\Pi_{1A}/r$$
$$\Rightarrow \xi_0 = \frac{C}{\Lambda_0(\Pi_0 - \Pi_{1A})/r}$$

What happens if the penalty is increased above ξ_0? An equilibrium (z^*, x^*), if exists, needs to jointly satisfy the following:

$$G_0^B(x^*,z^*) = G_2^B(x^*,z^*) \quad G_0^A(x^*,z^*) = G_1^B(x^*,z^*)$$
$$\Leftrightarrow G_0^A(x^*,z^*) + \Lambda_0\xi\Pi_0/r = G_1^A(x^*,z^*) + C + \Lambda_0\xi\Pi_{1A}/r$$
$$\Lambda_0\xi\Pi_0/r = C + \Lambda_0\xi\Pi_{1A}/r$$

However, for any $\xi > \xi_0$, the last equality is a contradiction. Hence, there is no equilibrium. What happens is that once the penalty fee is raised above ξ_0, *all* of the ASes in Region B will end up disabling their security measures. Hence ξ_0 can be thought of as a *watershed* threshold: before ξ_0 there is no free-riding and after ξ_0 all ASes of Region B will free-ride.

V. Heterogeneous Autonomous Systems

A property that we have been assuming so far is that ASes are homogeneous in their characteristics, i.e., they share the same parameters such as: the perceived costs per intrusion, size of the subnet, discount factor (their shortsightedness), recovery rate and the fraction of attackers in their subnets. In what follows, we show how our model can be generalized to incorporate the heterogeneity in such parameters. To avoid

undue clutter and attain basic insights, we look at each parameter separately, i.e., we successively assume that except for the parameter under scrutiny, the rest of the parameters are similar among ASes.

In order to relate the heterogeneity of the ASes to their decisions, as is the convention, we define $\theta \in \mathbb{R}^+$ to be the *type* of an AS. The type of an AS, θ, determines the parameter value of the AS (e.g., its perceived costs of intrusion), which in turn influences its adoption decisions. Let $F(\theta)$ represent the fraction of ASes that have types less than or equal to θ. Also, let $F^c(\theta)$ represent the tail distribution (exceedance) of θ, i.e., $F^c(\theta) = 1 - F(\theta)$. We assume no specific distribution for θ. Note that here, by "distribution", we refer to the empirical (i.e., sample) distribution of the types. For simplicity of analysis, we consider $F(\cdot)$ to be a continuous function of θ. Without loss of generality, we assume that the realized value of the heterogeneous parameter is linearly related to the type. The generality is preserved because a new type can be defined that has a linear relationship and its distribution can be computed from the distribution of the original type. We define $x(\theta) : [\theta_{\min}, \theta_{\max}] \rightarrow [0,1]$ such that $x(\theta)\frac{dF(\theta)}{d\theta}$ is the *density* function of the ASes that posses and enable the security measure. That is, $\int_a^b x_\theta \, dF(\theta)$ is the *fraction* of the total ASes whose type is between a and b and that have adopted and enabled the security measure. Similarly, we let $y(\theta) : [\theta_{\min}, \theta_{\max}] \rightarrow [0,1]$ be such that $y(\theta)dF(\theta)/d\theta$ is the density function of the ASes that do not have the security measure. Let X represent the total fraction of the ASes that have adopted and enabled the security measure. Following the definition of $x(\theta)$, we have:

$$X = \int_{\theta_{\min}}^{\theta_{\max}} x(\theta)\, dF(\theta). \tag{10}$$

The first point to note is that by replacing x in the homogeneous case by X as given above, the results for the homogeneous case can be generalized to the heterogeneous cases as well. In what follows, we present the model that incorporates heterogeneity in the assigned costs per intrusion. Treatment of heterogeneity in sizes of the subnets, discount factors, recovery rates, and the "pollution" level of the ASes follow in a similar manner, and are relegated to our technical report [18] due to lack of space.

A. Heterogeneity in the (Perceived) Costs of Intrusion (K_0, k):

Not all ASes "care" similarly about intrusions. For instance, ASes serving military, financial or other business customers may be far more concerned about intrusions than ASes that serve residential customers. Hence, heterogeneity in the cost of intrusions is a better reflection of reality.

For simplicity, we assume that both K and k_0 depend similarly on the type: $K_0 = K\theta$ and $k = \kappa\theta$, where K and κ are two constants, and $0 \leq \theta_{\min} \leq \theta \leq \theta_{\max}$. We examine the two cases of $C_0 = 0$ and C_0 separately, as the analysis of the latter turns out to be more involved.

1) Case of No Purchase Fee ($C_0 = 0$): In this case, there is no difference between enabling and buying as all of the ASes have access to a "free" copy of the security measure ($C_0 = 0$). Hence, the only decision they (independently, and at independent epochs) make is to whether enable or disable the security measure. With X given by (10), we have:

$$G_0(\theta, x(\cdot)) = -L\theta\,(\Pi_0 - X(\Pi_0 - \Pi_1))$$
$$G_1(\theta, x(\cdot)) = G_2(\theta, x(\cdot)) = -\frac{c}{r} - L\theta\,(\pi_0 - X(\pi_0 - \pi_1))$$

where L was $\frac{\Lambda}{\mu r}(K(\mu + r) + \kappa)$. A point to notice here is that $\partial G_0/\partial\theta < \partial G_1/\partial\theta < 0$, and the explicit relation to θ is linear. Hence, for a given value of X, there exists a unique $\hat{\theta} \in [\theta_{\min}, \theta_{\max}]$ such that $G_1(\theta, X) > G_0(\theta, X)$ for $\theta \in (\theta_{\min}, \hat{\theta})$, and $G_1(\theta, X) > G_0(\theta, X)$ for $\theta \in (\hat{\theta}, \theta_{\max})$. Let X^* denote an equilibrium value of X. By definition, at an equilibrium, no ASes of *any* type has a strictly preferable option to switch to. Hence, at an equilibrium, we have:

$$\begin{cases} x^*(\theta) = 1 & \theta :\ G_1(\theta, X^*) > G_0(\theta, X^*) \\ x^*(\theta) = 0 & \theta :\ G_1(\theta, X^*) < G_0(\theta, X^*) \end{cases} \tag{11}$$

Let θ^* denote the *type* of the ASes that at equilibrium are *indifferent* with respect to enabling or disabling of the security measure. Combining (10) and (11) yields:

$$X^* = F^c(\theta^*) \tag{12}$$

From (11), (12), the value of θ^* (and hence the value of X^*) is computed by solving the following equation:

$$-L\theta^*(\Pi_0 - F^c(\theta^*)(\Pi_0 - \Pi_1)) = -\frac{c}{r} - L\theta^*(\pi_0 - F^c(\theta^*)(\pi_0 - \pi_1)) \tag{13}$$

In what follows we show that a valid θ^* is unique, i.e., there is at most one θ^* that satisfies the above equation for a given distribution $F(\theta)$. Note that $x^*(\cdot)$ is completely determined once θ^* is computed, and therefore, the uniqueness of θ^* implies the uniqueness of $x^*(\cdot)$ as well.

Let ψ be the LHS of (13) minus its RHS. Then:

$$\frac{1}{L}\frac{\partial\psi}{\partial\theta^*} = -\theta^* - (\Pi_0 - \pi_0) - F^c(\theta^*)\,((\Pi_0 - \Pi_1) - (\pi_0 - \pi_1))$$

Note that $\theta^* \geq 0$, and following (2), $((\Pi_0 - \Pi_1) - (\pi_0 - \pi_1)) \geq 0$. Hence, $\partial\psi/\partial\theta^* < 0$. This establishes that there is at most one θ^* for which $\psi = 0$, since ψ a differentiable function of θ^* can have at most one zero-crossing point. □

2) Case of Positive Purchase Fee ($C_0 > 0$): The utilities of an AS of type θ given the current X (as given by (10)) are:

$$G_0(\theta, x(\cdot)) = -L\theta\,(\Pi_0 - X(\Pi_0 - \Pi_1))$$
$$G_1(\theta, x(\cdot)) = -C_0 - \frac{c}{r} - L\theta\,(\pi_0 - X(\pi_0 - \pi_1))$$
$$G_2(\theta, x(\cdot)) = -\frac{c}{r} - L\theta\,(\pi_0 - X(\pi_0 - \pi_1))$$

Note that for any given X, all three utilities are linear and strictly decreasing in θ. Hence, there exist unique $\theta_e, \theta_b \in [\theta_{\min}, \theta_{\max}]$,[5] such that $G_0(\theta, X) > G_2(\theta, X) > G_1(\theta, X)$ for $\theta \in (\theta_{\min}, \theta_e)$, $G_2(\theta, X) > G_0(\theta, X) > G_1(\theta, X)$ for $\theta \in (\theta_e, \theta_b)$, and $G_2(\theta, X) > G_1(\theta, X) > G_0(\theta, X)$ for $\theta \in$

[5] e for *e*nabling and b for *b*uying.

$(\theta_b, \theta_{\max})$. θ_e is the indifferent type between enabling and disabling the security measure provided that the AS already has it, while θ_b is the indifferent type between buying+enabling and not buying. At an equilibrium, we have:

$$x^*(\theta) = 0 \text{ for } \theta < \theta_e^*, \quad \text{and} \quad x^*(\theta) = 1 \text{ for } \theta > \theta_b^* \quad (14)$$

The value of $y^*(\theta)$ for the above ranges is simply $1 - x^*(\theta)$, irrespective of the initial seeding. Values of $x^*(\theta)$ and $y^*(\theta)$ for $\theta \in (\theta_e^*, \theta_b^*)$, unlike the case of $C = 0$, depend on the dynamics of the system and the path history before reaching the equilibrium. Although we can continue with the assumption of each AS taking decisions at epochs of i.i.d. Poisson processes with rate γ, this is algebraically more challenging. Instead, in order to simplify the analysis, we make a rather technical assumption about the dynamics of the decision-taking process:

Assumption 1. We assume that the ASes who benefit the most from a decision, decide first.

This assumption allows us to eradicate the cases of partial densities of adoption, i.e., $x^*(\theta)$ or $y^*(\theta)$ will only assume 0/1 values. Under this assumption, given any initial seeding, the equilibrium is unique and at the equilibrium we have:

$$x^*(\theta) = s_{\theta,0}, \; y^*(\theta) = 1 - x_\theta^* \qquad \theta_e^* < \theta < \theta_b^* \qquad (15)$$

where $s_{\theta,0}$ is the density of the initial seeding across the types. From (14), (15) it follows that $X^* = F^c(\theta_b^*) + \int_{\theta_e^*}^{\theta_b^*} s_{\theta,0}\, dF(\theta)$. Then, at an equilibrium, we have:

$$G_0(\theta_e^*, X^*) = G_2(\theta_e^*, X^*), \qquad G_0(\theta_b^*, X^*) = G_1(\theta_b^*, X^*)$$

Uniqueness of the equilibrium given an initial seeding follows similar to the case of $C_0 = 0$.

3) Policy Implications: If the regulator can have an assessment of the types of the ASes, it can utilize such information to improve the efficiency of the regulation. For instance, if the regulator is to seed the ASes eith free copies of the security measure, it should seed only those ASes with types between θ_e^* and θ_b^*. This is because the ASes whose type is below θ_e^* will eventually not activate their security measures, and the ASes that are of type greater than θ_b^*, will purchase the security measure on their own anyway and enable it.

VI. Conclusion

In this paper we have developed an economic model to study the effectiveness of regulatory mechanisms that monitor outgoing AS-level threat activities and issue penalties based on threat origin to improve the security of the Internet. We showed that if the fees collected from penalties are reinvested or redistributed, then both social utility as well as individual utility of the ASes can improve, in addition to improving the security of the overall network. Next, we considered the fact that regulators have local jurisdiction and investigated the effect of regionally restricted authority. We showed that if the authority's region is smaller than a certain threshold, then free-riding of the unregulated ASes can undermine the objectives of the regulatory policy. This has policy implications on how Governmental or private corporations should approach the issue of grouping their ASes under the purview of different regulatory bodies. To the best of our knowledge this is the first analytical work that studies the above issues of regulatory implications in the context of egress filtering and bidirectional traffic monitoring.

In this simple qualitative analysis, we did not consider the fact that some ASes are subsets of other ASes. Also, the threats that we modeled were limited to intrusion attempts and not epidemic malware. Moreover, the investment decision of the ASes were simplified to the set of obtaining/activating/disabling of a monolithic product, as opposed to a continuum of investments and services. We aim to address these generalizations in our future research. This work can potentially stimulate further research in this area and attract regulator's and ISPs' attention to provide hard to obtain real data for further studies and guidelines.

References

[1] A. Friedman, "Economic and policy frameworks for cybersecurity risks," *Center for Technology Innovation at BROOKINGS*, 2011.

[2] S. Hofmeyr, T. Moore, S. Forrest, B. Edwards, and G. Stelle, "Modeling internet-scale policies for cleaning up malware," *Arxiv preprint arXiv:1202.4008*, 2012.

[3] J. Quarterman, S. Sayin, J. Reinikainen, E. Kumar, and A. Whinston, "Data, reputation, and certification against spam," *DDCSW: Collaborative Data-Driven Security for High Performance Networks*, 2010.

[4] J. Quarterman, S. Sayin, and A. Whinston, "Rustock botnet and asns," *TPRC*, September 2011.

[5] C. Bauch and D. Earn, "Vaccination and the theory of games," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 36, p. 13391, 2004.

[6] T. Reluga and A. Galvani, "A general approach for population games with application to vaccination," *Mathematical Biosciences*, 2011.

[7] J. Grossklags, N. Christin, and J. Chuang, "Secure or insure?: a game-theoretic analysis of information security games," in *Proc. of the 17th conference on World Wide Web*. ACM, 2008, pp. 209–218.

[8] A. d'Onofrio, P. Manfredi, and E. Salinelli, "Vaccinating behaviour, information, and the dynamics of SIR vaccine preventable diseases," *Theoretical population biology*, vol. 71, no. 3, pp. 301–317, 2007.

[9] S. Sen, Y. Jin, R. Guerin, and K. Hosanagar, "Modeling the dynamics of network technology adoption and the role of converters," *IEEE/ACM Trans. on Networking*, vol. 18, no. 6, pp. 1793–1805, 2010.

[10] R. Böhme and G. Kataria, "Models and measures for correlation in cyber-insurance," in *Economics of Information Security*, 2006.

[11] J. François, G. Moura, and A. Pras, "Cleaning your house first: Shifting the paradigm on how to secure networks," *Managing the Dynamics of Networks and Services*, pp. 1–12, 2011.

[12] M. Khouzani, S. Sen, and N. B. Shroff, "Managing the Adoption of Asymmetric Bidirectional Firewalls: Seeding and Mandating," in *To Appear In IEEE GLOBECOM, Anaheim, December 3-7*, 2012.

[13] N. S. Council, "The Comprehensive National Cybersecurity Initiative," 2010, http://www.whitehouse.gov/sites/default/files/cybersecurity.pdf.

[14] J. Thames and R. Abler, "Implementing distributed internet security using a firewall collaboration framework," in *SoutheastCon, 2007. Proceedings. IEEE.* IEEE, 2007, pp. 680–685.

[15] K. Tatsumi and M. Goto, "Optimal timing of information security investment: A real options approach," *Economics of Information Security and Privacy*, pp. 211–228, 2010.

[16] I. Mokube and M. Adams, "Honeypots: concepts, approaches, and challenges," in *Proceedings of the 45th annual southeast regional conference*. ACM, 2007, pp. 321–326.

[17] H. Artail, H. Safa, M. Sraj, I. Kuwatly, and Z. Al-Masri, "A hybrid honeypot framework for improving intrusion detection systems in protecting organizational networks," *Computers & Security*, vol. 25, no. 4, pp. 274–288, 2006.

[18] M. Khouzani, S. Sen, and N. B. Shroff, "An Economic Analysis of Regulating Security Investments in the Internet (detailed version)," in *Technical Report*, 2012, http://www2.ece.ohio-state.edu/*sim*khouzani/Preprints/I13techrep.pdf.