

SECURITY STANDARDIZATION IN THE PRESENCE OF UNVERIFIABLE CONTROL

Completed Research Paper

Chul H. Lee

The University of Texas at Dallas
Richardson, TX 75080
irontiger@utdallas.edu

Xianjun Geng

The University of Texas at Dallas
Richardson, TX 75080
geng@utdallas.edu

Srinivasan Raghunathan

The University of Texas at Dallas
Richardson, TX 75080
sraghu@utdallas.edu

Abstract

Increasingly, policy makers in both private and public sectors mandate information security standards upon organizations in order to protect both organizational and individual digital assets. One major issue in security standardization is that standards often cannot cover all possible security efforts by organizations because some efforts are unverifiable by nature. This paper analytically studies how a policy maker should design the standard on a verifiable security control when another related unverifiable one exists. We find that naively ignoring the existence of the unverifiable control will in general lead to a sub-optimal standard. Furthermore, optimal standard depends critically on how the two security controls work together -- which we refer to as security configurations -- to protect the firm's digital asset. Under parallel configuration, the existence of the unverifiable control induces the policy maker to set a higher standard; under serial configuration, a lower standard. Under best-shot configuration and if the verifiable control is more cost-efficient, the existence of the unverifiable control has no impact on the optimal standard. We also find that whether attackers are strategically targeting the weakest-link control has a significant impact on optimal security standard under the parallel configuration. Such strategic attacking behavior can severely handicap the policy maker's optimal standard and thus reduce its effectiveness.

Keywords: Information security, control, standard, unverifiability, strategic attacker

Introduction

In this networked economy, when an organization's digital asset or online service is compromised by attacks, damages often go beyond the organizational boundary. For example, in 2005 the information system of a credit card processor, CardSystems Solutions, was breached and subsequently 40 million credit card numbers were stolen (Rothke et al. 2009, Zetter 2009). In this example, breach happened to a single company yet millions of consumers were affected -- an event that caused hot debates on whether organizations alone have enough motivation to invest adequately in information security especially when others (consumers in this case) shoulder the consequences of a breach. Increasingly, policy makers in both private and public sectors mandate information security standards upon organizations not only to reduce the chance of and damage from direct security breaches upon these organizations, but also to protect the value of all stakeholders (such as an organization's supply chain partners and clients) whose private information are shared with these organizations. Two such prominent policy makers are PCI Security Standards Council in the private sector that mandates information security standards upon all merchants that use major payment cards (such as Visa and Master cards), and the National Institute of Standards and Technology (NIST) that mandates information security standards upon all US governmental agencies.

One major issue in security standardization is that standards cannot cover all possible security controls by organizations because some controls are unverifiable to a policy maker.¹ Unverifiability of security controls arise for a variety of reasons. First, some controls -- especially ones involve human diligence -- are difficult (if not entirely impossible) to measure. For example, the effectiveness of an Intrusion Detection System (IDS) depends critically on IT security staff's professional judgment and prioritization of security tickets generated, whereas such professionalism is hard to quantify and measure. Second, it may be cost-prohibitive for policy makers (or their delegates) to adequately monitor an organization's internal controls. Third, information security is a fast-evolving field where new security threats and accordingly new security controls constantly emerge. A security standard, at time of its inception, is not likely to cover security controls not yet invented. *Will the existence of unverifiable security controls affect a policy maker's decision on the standards of verifiable security controls? If yes, how? Furthermore, will the policy maker's decision on standards be affected by attacker strategy?* To our knowledge no existing research in the information security field has answered these highly relevant questions.

As a first attempt in analyzing the issue of security standardization in the presence of unverifiable security control, in this paper we consider a game-theoretical model that includes one verifiable security control and one unverifiable security control.² The purpose of these two controls is to *together* protect a digital asset. For example, to protect internal digital data, it is popular for a firm to deploy both a firewall and an IDS. To enforce a sound password policy, a firm can both use password management software to force employees to pick long passwords that include special characters and to send IT security personnel to regularly tour the offices to send warning to any employee that has password attached to her/his computer on a paper note. Depending on how these two security controls are connected to each other and to the digital asset to be protected -- which we refer to as *security configurations*, an attacker can cause damage to the digital asset either by breaching one or both controls. We consider three basic and fundamental security configurations: *parallel*, *serial* and *best-shot* configurations. We also consider two types of security attacks: nonstrategic attacks that take place regardless of a firm's defense, and strategic attacks that try to target the weakest link among the two security controls.

We find that, with one exception, the answer to our first research question is affirmative. Though a policy maker cannot direct mandate a firm's investment on the unverifiable security control, it turns out its standard on the verifiable control will *indirectly* affect firm incentive regarding whether and how to invest on the unverifiable control, as the firm tries to strike an optimal balance between the two controls in order to protect the digital asset efficiently (from firm perspective). Therefore, in general it is not optimal for the policy maker to naively ignore the existence of any unverifiable security control when designing security standards. The only exception is for best-shot configuration and when the verifiable control is more cost-efficient than the unverifiable control -- in this case, optimality calls for the firm to put all investment on the former control, thus the latter control becomes irrelevant.

¹ "Security control" is a commonly used term in information security domain that refers to safeguards or countermeasures for avoiding or minimizing security risks.

² Throughout this paper, by "unverifiable" we always mean "unverifiable from the policy maker's perspective."

We find that how the existence of an unverifiable control affects security standard depends critically on the specific security configuration that the two controls are embedded in. Under parallel configuration, firm investment on the unverifiable security control increases in the standard on the verifiable control. Therefore, the existence of the unverifiable control encourages the policy maker to set a higher standard (than it would in absence of this control). Under best-shot configuration and if the verifiable control is relatively cost-efficient, the unverifiable control has no impact on the standard; nevertheless, if the unverifiable control is much more cost-efficient, the policy maker should not impose any standard at all, so the firm can make cost-efficient investment on the unverifiable control. Under serial configuration, firm investment on the unverifiable control decreases in the standard on the verifiable control. Therefore, the existence of the unverifiable control encourages the policy maker to set a lower standard.

Our third major finding is that, under parallel configuration, whether security attacks are strategic (i.e. targeting the weakest link) or not has a significant impact on the optimal standard by the policy maker. When strategic attackers try to first find out which security control is the weakest link before pounding this weakest link, as a counter-measure it is beneficial for the firm to balance the defense on the two controls so neither one is the apparent weakest link. A too high security standard from the policy maker, however, will force the firm to invest excessively on the verifiable control to a level where the firm finds it no longer economically worthwhile to balance between the two controls. Consequently, a too high security standard reduces the information advantage of the firm over strategic attackers and actually helps the latter better identify and thus attack the weakest security control. As a result, the existence of strategic attacks significantly handicaps the policy maker's capability in setting a high security standard.

The rest of the paper is organized as follows. In Section 2 we review relevant literatures. We present our model in Section 3. We discuss three security configurations -- parallel, best-shot and serial -- in Section 4, and strategic attacks in Section 5. Section 6 concludes this paper.

Literature Review

Since security standards as a strategy to manage information security is a recent development, the extant research on this topic is limited. Much of the prior work on security standards has taken a descriptive approach to the standard setting problem and focused on principles that should govern information security standards (Keblawi and Sullivan 2007, Ross 2007, Morse and Raval 2008, Culnan and Williams 2009). Some of the recent work has empirically examined the impact of standards and laws related to breach disclosure and data encryption on security incidents. Romanosky et al. (2009) show that the adoption of data breach disclosure laws has marginal effect on the reduction in incidences of identity thefts. Miller and Tucker (2010) show that adoption of encryption software because of safe harbor provisions in breach notification regulations increases the incidence of publicized data losses because of carelessness with respect to other protection activities on the part of those that should protect the information asset.

While the extant literature on security standards is sparse, extensive work has been done on standards in other settings. Of particular relevance is the literature on financial auditing standards. Dye (1993) shows that the average quality of audits may decline as auditing standard becomes tougher. Willekens et al. (1996) argue that the increased difficulty of firing a compliant auditor with that follows standards can reduce rather than increase the quality of audit work supplied. Schwarts (1998) finds that the socially optimal commitment according to auditing standards is achievable if the auditor's legal liability regime is strict liability and is independent of the actual investment. While the research in the auditing standards literature model auditing as a single observable activity on which standards can be imposed, we consider a model in which multiple security controls exist and standards cannot be imposed on all of them.

One unique aspect of information security is the presence of strategic hackers who may use information about standards and change their attack strategy. Such strategic adversaries are not present in contexts such as auditing. The literature on information security economics has analyzed scenarios with strategic attackers. Cavusoglu et al. (2005) analyze the value of IDS and show that IDS offer a positive value only when they deter hackers. Cavusoglu et al. (2009) highlight the complex interactions between firewall and IDS technologies when they are used together in a security architecture, and, hence, the need for proper configuration to benefit from these technologies. They show that every technology has different optimal configuration level according to their performance and circumstances. Starting with Varian (2004), several papers have examined the economic incentive of agents which have interdependency on security (Grossklags et al. 2008, Narasimhan et al. 2010). Narasimhan et al.(2010) show that the success of cooperative security efforts depends on the nature of the attack and the attitude of the defenders. On the other hand, Schechter and Smith (2003) analyze how much security is required when attacker focus only one

attractive target or penetrate as many systems as possible. However, this stream of work does not consider security standards.

Our work is also related to the literature on incomplete contracts with unverifiable services (Bernheim and Whinston 1998, Battigalli and Maggi 2002). Bernheim and Whinston (1998) shows it is often optimal to specify an incomplete contract, when some aspects of performance are unverifiable. Battigalli and Maggi (2002) further propose optimal contracts with rigidity and discretion if writing contract is very costly. This research does not consider strategic adversaries.

Additionally, our paper is also related to Hendricks and McAfee (2006) and Crawford (2003) who consider a signaling model to analyze attacker-defender games. In our case, standards are set by the social planner and these signals could be used by attackers to compromise the defender's information asset.

The Model

The model consists of one *firm* that is in charge of protecting a digital asset or service using two security controls, a *representative attacker* that may assail the security controls in order to compromise the digital asset/service, and one *policy maker* that aims to optimize social welfare by setting security standards that the firm must follow.

The Firm

We are interested in the scenario where, if the digital asset or service is compromised by attacks, damages go beyond the firm boundary. A real-life example is the well-publicized 2005 incident in which a credit card processor, CardSystems Solutions, was breached and subsequently 40 million credit card numbers were stolen (Rothke et al. 2009, Zetter 2009). In this example, breach happened to a single company yet millions of consumers were affected.³ Broadly speaking, whenever a firm stores information for (or provides services to) their customers and supply-chain stakeholders, there is a possibility that customers or other stakeholders may be affected when this firm's information security is breached. Formally in this model, if the digital asset is compromised by attacks, let the damage to the firm be a constant D_F and the damage to social welfare be D_{SW} . $D_{SW} > D_F > 0$. Let damages include opportunity costs -- what the firm (and society) would have normally gained should the compromise not take place. We also assume that any contingent transfer payments upon a security incident (e.g. ones designated in a Service-Level Agreement (SLA)) are included in D_F .

Note that the firm's primary business can be (and in practice often is) different from security provision. For example, CardSystems Solutions provides security services yet its primary business function is to process credit card transactions. We focus on security issues in this paper and assume that, notwithstanding a security compromise, the firm earns a business profit of V_F and the society in total receives a benefit of V_{SW} . $V_{SW} > V_F > 0$. We further assume that V_F is large enough so that the firm will not exit the market due to information security problems.⁴

Security Controls

The firm protects the digital asset using security controls.⁵ As modern information systems are getting increasingly

³ Another implication of this incident is that a security breach can have a negative social impact far beyond a firm's immediate customers (which are merchants that use CardSystems Solutions in this case). In fact, CardSystems Solutions was not a household name for ordinary credit-card holders. Therefore, it is more appropriate to say "damages to social welfare" than "damages to customers/clients" in this research.

⁴ This assumption is realistic because only in rare occasions will a firm declare bankruptcy (or steer clear of the affected businesses) following a security incident. Furthermore, modeling individual rationality for the firm does not lead to any significantly new insights beyond what this paper currently offers.

⁵ "Security control" is a widely-adopted term for countermeasures to information security risks. For example, NIST defines security controls as "the management, operational, and technical safeguards or countermeasures prescribed for an information

complex and interconnected, organizations often find themselves having a plural of security weaknesses to address. Accordingly, a common practice is for organizations to deploy multiple security controls (*controls* in short) in a comprehensive protection plan, such as multiple firewalls to safeguard all entrances to a corporate network. In this model we consider a simple case in which, in order to protect the digital asset, the firm needs to invest in two security controls, V and N.⁶ Let b_i represent the probability that attackers successfully breach security control i , $i \in \{V, N\}$. We consider the following breach probability function:

$$b_i = \begin{cases} \exp(-\frac{m_i}{K_i t_i}) & \text{if } t_i > 0 \\ 0 & \text{if } t_i = 0 \end{cases} . \quad (1)$$

The breach probability of control i ($i \in \{V, N\}$) is a negative exponential function that decreases in the firm's investment, m_i , on control i . Investment can take diverse forms such as technological purchases, development and maintenance, and labor. We assume that all investment can be measured in total by a non-negative monetary variable m_i . The breach probability increases in the effort by the representative attacker⁷, t_i . Hereafter we refer to t_i as "attack intensity" for ease of exposition. The possible difference between constants K_V and K_N captures the heterogeneous cost structures in the two controls: for example, given the same attack intensities and if $K_V < K_N$, to reach the same level of protection (i.e. $b_V = b_N$) control V requires less investment than control N. Hereafter we say control V is more (less) cost-effective than control N if $K_V < K_N$ ($K_V > K_N$). To rule out the uninteresting case of no firm investment on security controls, we assume $\max\{K_N, K_V\} < D_F$.

For any given positive attack intensity, the negative exponential form of the breach probability function implies the marginal investment needed to reduce b_i by a unit increases in b_i -- in other words, the firm faces a convex security cost function. Furthermore, it ensures that b_i falls into region $[0,1]$. This functional form also implies that, for any given positive attack intensity, perfect security (i.e. $b_i = 0$) is unattainable. The negative exponential function has been used by others in modeling security breach probabilities (Zhao et al. 2009). For notational succinctness, we slightly abuse the notation and treat $\exp(-m_i / (K_i t_i))$ as $\lim_{\tau \rightarrow 0} \exp(-m_i / (K_i \tau))$ when $t_i = 0$, and therefore use $b_i = \exp(-m_i / (K_i t_i))$ for any non-negative t_i instead of the conditional form in (1).

Let function $\omega(b_V, b_N)$ denote the probability that attackers successfully compromise the digital asset or service. We can then write the firm's expected utility as:⁸

$$U_F = V_F - \omega(b_V(m_V, t_V), b_N(m_N, t_N))D_F - m_V - m_N . \quad (2)$$

Three Security Configurations

We next describe the relationship between the two security controls and the security of the digital asset. We consider three basic and commonly-seen relationships -- which we refer to as *security configurations* -- depending on the nature of the security attacks and how the controls are interconnected.

Information security attacks can lead to two broad categories of detrimental consequences for businesses: unauthorized access of information and service disruptions (Loch et al. 1992). If a firm's security concern is on unauthorized access, naturally the firm would like to plug all possible loopholes through which threats may penetrate. Consider a scenario with two such loopholes, where breaching of either one can lead to unauthorized

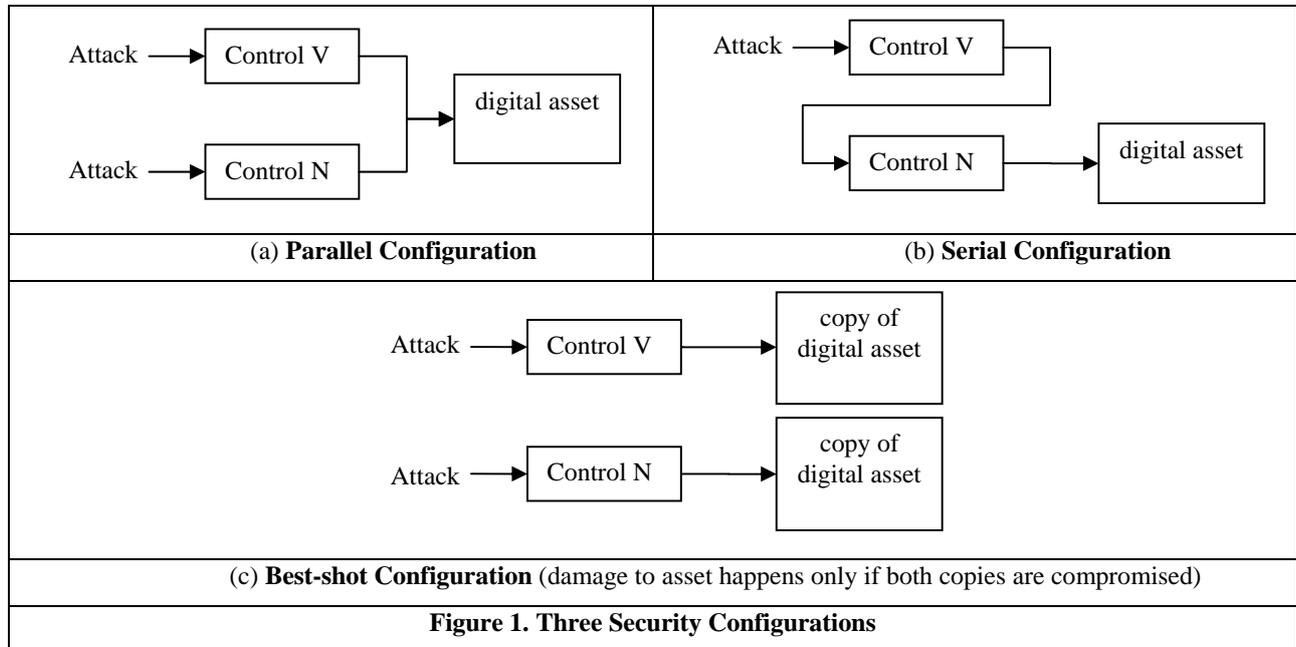
system to protect the confidentiality, integrity, and availability of the system and its information." (csrc.nist.gov/publications/nistpubs/800-53-Rev2/sp800-53-rev2-final.pdf, page 1)

⁶ Shortly we will see that "V" stands for "verifiable control," and "N" stands for "unverifiable control."

⁷ Or the collective effort of multiple attackers.

⁸ Again, note that D_F includes opportunity costs and any transfer payments upon a security breach.

access. The firm can then deploy one security control to each loophole -- called the *parallel configuration* -- as shown in Figure 1a. In other words, parallel configuration refers to the case where the digital asset is compromised when either of the two security controls is breached. One commonly seen example of the parallel configuration is a corporate network that is linked to the Internet at multiple access points, whereas each access point is secured by a separate firewall -- a widely used type of security control. Breaching any such firewall will then expose internal data to an attacker. Under parallel configuration, $\omega(b_V, b_N) = 1 - (1 - b_V)(1 - b_N) = b_V + b_N - b_V b_N$.



An alternative scenario -- the *serial configuration* -- is when the firm has only one security loophole, and the firm deploys two or more controls sequentially to defend against this loophole, as shown in Figure 1b. In other words, serial configuration refers to the case where the digital asset is compromised only when both security controls are breached sequentially. One prominent example is the deployment of both a firewall and an intrusion detection system (IDS) on a single Internet access point: if the firewall fails to catch a threat, the IDS can serve as an additional safeguard. Under serial configuration, $\omega(b_V, b_N) = b_V b_N$. Moreover, because the first security control in a serial configuration (e.g. V in Figure 1b) already filters out some attacks, the second security control (N) faces an often much-reduced attack intensity than the first one.

When a firm's security concern is on service disruption instead of on unauthorized access, a popular defense method is to create redundant and distributed copies of the same data or service, and then to protect every copy. For example, Denial-of-Service (DoS) attacks are a frequent type of disruption attacks to web services (Geng et al. 2002). A popular defense for many web service operators, such as CNN.com and MTV.com, is to deploy their services to multiple web servers so that if one server experiences service outage due to attacks, other redundant servers can takeover and resume the service.⁹ Formally, *best-shot configuration* refers to the case where digital asset security depends only on the strongest link between the two controls, as illustrated in Figure 1(c). Another example of the best-shot configuration is the popular practice of using Disaster Recovery Plan (DRP) to address possible natural or man-made disasters that destroy IT data or infrastructure: unrecoverable data or service loss can be avoided as long as at least one backup is not affected by a disaster. The breach probability function under best-shot configuration has the same form as the one under serial configuration, i.e. $\omega(b_V, b_N) = b_V b_N$. Nevertheless, these two security configurations differ significantly in that, under best-shot configuration, neither control filters out attacks for the other.

⁹ For e-commerce a business does not have to build redundant servers all by self. The Content Distribution Network (CDN) industry, where Akamai is a market share leader, provides rental service of redundant servers.

Note that in business practice, security configurations can be a complex combination of the aforementioned basic ones. As a first theoretical exploration on understanding the impact of security configurations on standardization in the presence of an unverifiable control, we focus on basic security configurations.

Also note that we do not include the *weakest-link configuration* -- under which digital asset security depends only on the weakest link between the two controls (i.e. $\omega(b_V, b_N) = \max\{b_V, b_N\}$) -- in the base model. Weakest-link implies that attackers will first strategically identify which of the two controls is weaker before attacking. Therefore we postpone the detailed discussion of this configuration to Section 5 when we analyze strategic attackers.

Non-Strategic and Strategic Attacks

Attacks against the security controls can be broadly classified into two categories: ones that are independent of the security investments made by the firm, and ones that are dependent. We refer to the former as non-strategic attacks and the latter as strategic attacks.

Intuitively, a security attack can be most effective when it is against a firm's weakest point of defense. Therefore, a strategic attacker may find it beneficial to first analyze a firm's security investments in order to identify the weakest control before taking any action. We will analyze such "weakest-link" attack strategy in Section 5.

There are, nevertheless, two other widely applicable cases where security risks are non-strategic. First, it is common for hackers to blanket the Internet with automated attacks, such as Port Scan Attacks.¹⁰ The frequency with which a firm receives Port Scan Attacks to any of its security controls has little to do with the relative strength among these security controls given the automated nature of the attacks. Second, many security risks are due to non-strategic factors such as equipment deterioration, natural disasters, accidental man-made disasters or adverse environmental conditions (e.g. power outage). We consider non-strategic attacks in Section 4.

The Policy Maker and Verifiability of Security Controls

The policy maker's objective is to maximize the expected social welfare, U_{SW} as shown below, via security standardization.

$$U_{SW} = V_{SW} - \omega(b_V(m_V, t_V), b_N(m_N, t_N))D_{SW} - m_V - m_N. \quad (3)$$

While the direct control of security investments is in the hands of the firm, the policy maker can indirectly affect firm investments through regulatory standards (such as PCI-DSS) on any verifiable security control. In this paper we are interested in the case where security control V is verifiable to the policy maker while N is not. For example and in the context of reducing firewall breaches, control V can be the frequency of external review of firewall rule sets that is contractually verifiable and thus enforceable by the policy maker;¹¹ control N can be a firm's managerial effort spent on discouraging employees from visiting external websites that are irrelevant to their jobs, whereas such effort is hard to monitor and quantify.

As a result, the policy maker can only mandate a standard s for control V. A standard for control V is an investment threshold that the firm must match or exceed.¹² For the scope of this paper, we focus on security standards that have strict enforcement power, so that the affected firm has to unconditionally confirm. Two widely applicable examples are NIST security standards and PCI-DSS: NIST standards are mandatory for all affected US governmental agencies (Keblawi and Sullivan 2007); PCI-DSS is mandatory for all merchants that "accepts, transmits or stores any (credit or debit) cardholder data."¹³

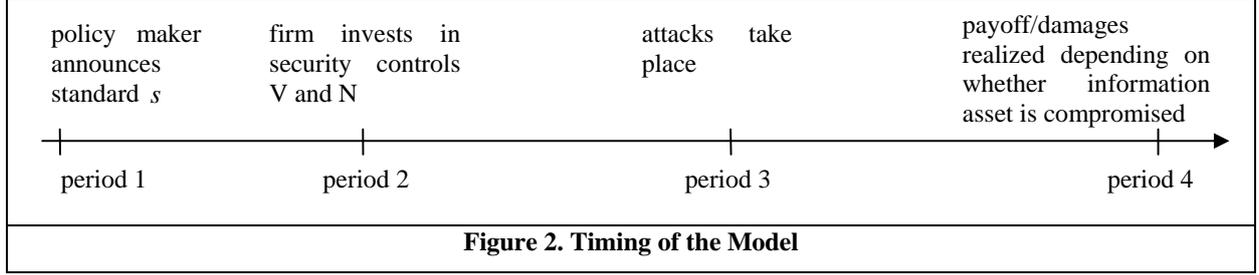
¹⁰ See <http://www.javvin.com/networksecurity/PortScanAttack.html> . Some botnets also attempt to indiscriminately infiltrate computers on the Internet.

¹¹ This is item 1.1.6 in in PCI-DSS version 1.2.1.

¹² For example, item 1.1.6 in in PCI-DSS version 1.2.1 requires a firm to "review firewall and router rule sets at least every six months."

¹³ <http://www.pcicomplianceguide.org/pcifaqs.php#2> .

Figure 2 shows the timing of the model. The policy maker first announces the standard, s , for control V. The firm then chooses its investments m_V and m_N on the security controls. Possible security attacks then take place.



Standardization Under Nonstrategic Attacks

In this section we consider nonstrategic attacks and study how the existence of the unverifiable security control N affects firm investments and the optimal security standard on the verifiable security control V.

As the attacks are nonstrategic, both t_V and t_N are exogenously given under either parallel configuration or best-shot configuration, which we consider as constants. Without loss of generality, we normalize both t_V and t_N to constant one under these two security configurations.¹⁴ Under serial configuration, let the attack intensity to the first security control be one, whereas the attack intensity to the second control will be lower and will depend on the effectiveness of the first control in blocking attacks.

Next we analyze security standardization and firm response for each of the three security configurations.

Parallel Configuration

We use backward induction to analyze parallel configuration: first, for any given standard s_{PC} on security control V, we analyze the firm's optimal investments on both controls V and N in period 2; second, we analyze the policy maker's optimal standard in period 1. Throughout this subsection, subscript "PC" means "parallel configuration."

$\omega(b_V, b_N) = b_V + b_N - b_V b_N = \exp(-m_V / K_V) + \exp(-m_N / K_N) - \exp(-m_V / K_V - m_N / K_N)$ under parallel configuration. In period 2 and given any arbitrary standard s_{PC} for control V that is imposed by the policy maker, the firm's optimization problem is:

$$\max_{m_V, m_N} U_F = V_F - \left(\exp\left(-\frac{m_V}{K_V}\right) + \exp\left(-\frac{m_N}{K_N}\right) - \exp\left(-\frac{m_V}{K_V} - \frac{m_N}{K_N}\right) \right) D_F - m_V - m_N,$$

$$\text{s.t. } m_V \geq s_{PC}.$$

For notational convenience, denote $\tilde{b}_V \equiv (D_F + K_V - K_N - \sqrt{(D_F - K_V + K_N)^2 - 4K_N D_F}) / (2D_F)$ and $\tilde{b}_N \equiv (D_F - K_V + K_N - \sqrt{(D_F - K_V + K_N)^2 - 4K_N D_F}) / (2D_F)$. \tilde{b}_V (\tilde{b}_N) is the firm's optimal choice of breach probability on control V (N) under parallel configuration when there is no security standard. The next lemma presents the firm's optimal investments.

Lemma 1: *Under parallel configuration and given standard s_{PC} for control V:*

- i. If $s_{PC} < K_V \ln(1/\tilde{b}_V)$, $m_V^* = K_V \ln(1/\tilde{b}_V)$ and $m_N^* = K_N \ln(1/\tilde{b}_N)$.

¹⁴ Any heterogeneity between the two attack intensities can be captured by the possibly different K_V and K_N .

ii. If $s_{PC} \geq K_V \ln(1/\tilde{b}_V)$, $m_V^* = s_{PC}$ and $m_N^* = K_N \ln \frac{D_F(1 - \exp(-s_{PC}/K_V))}{K_N}$.

Proofs are in the Appendix. Lemma 1 shows that, for the security standard s_{PC} to have impact on firm investments, it has to be high enough (i.e. more than $K_V \ln(1/\tilde{b}_V)$). Given $s_{PC} \geq K_V \ln(1/\tilde{b}_V)$, a higher standard not only directly forces the firm to invest more in the verifiable control, it also indirectly incentivizes the firm to invest more on the unverifiable control. We capture this important observation in the following proposition.

Proposition 1: *Under parallel configuration, the firm's investment on the unverifiable security control increases in the standard on the verifiable control when the standard is high enough ($s_{PC} \geq K_V \ln(1/\tilde{b}_V)$).*

Intuitively, under parallel configuration the firm's investments on the two security controls are complementary to each other -- high security investment on one control is effective in protecting the digital asset only if the investment on the other control is not disproportionately low. When a higher standard directly forces the firm to invest more heavily on the verifiable control, Proposition 1 shows that the firm finds the marginal return from investing on the unverifiable control increases accordingly -- thus it invests more on the unverifiable control.

We next analyze optimal standard decision by the policy maker in period 1. Analytically, because results under $s_{PC} < K_V \ln(1/\tilde{b}_V)$ are equivalent to the one under $s_{PC} = K_V \ln(1/\tilde{b}_V)$, it is sufficient for us to only consider $s_{PC} \geq K_V \ln(1/\tilde{b}_V)$. The policy maker's optimization problem is

$$\max_{s_{PC}} U_{SW} = V_{SW} - (\exp(-\frac{m_V}{K_V}) + \exp(-\frac{m_N}{K_N}) - \exp(-\frac{m_V}{K_V} - \frac{m_N}{K_N}))D_{SW} - m_V - m_N,$$

where $s_{PC} \geq K_V \ln(1/\tilde{b}_V)$, $m_V = s_{PC}$ and $m_N = K_N \ln \frac{D_F(1 - \exp(-s_{PC}/K_V))}{K_N}$. It turns out the policy maker will

always choose a standard high enough so that the firm is forced to invest more on both controls (than it would under no standard):

Lemma 2: *Under parallel configuration, the socially optimal standard on control V is*

$$s_{PC}^* = K_V \ln(1/b_{PC}^*), \text{ where } b_{PC}^* \equiv (D_{SW} + K_V - K_N - \sqrt{(D_{SW} - K_V + K_N)^2 - 4K_N D_{SW}}) / (2D_{SW}).$$

It is now worthwhile for us to compare the results in Lemma 2 to the ones under a *complete-information benchmark*. Consider, for a moment, the scenario where the policy maker can impose and enforce standards on both security controls -- we call this scenario the complete-information benchmark, and the policy maker's optimal standards under this benchmark *complete-information standards*. In other words, complete-information standards are the optimal standards when both security controls are verifiable. From equation (3) and $\omega(b_V, b_N) = b_V + b_N - b_V b_N$ it is straightforward to verify that the complete-information standard on control V is exactly s_{PC}^* . Therefore:

Proposition 2: *Under parallel configuration, the policy maker should simply impose the complete-information standard for security control V.*

Proposition 2 implies that, even though the policy maker is facing a complex situation where not all security controls are verifiable, its optimal choice of standard is nevertheless simple under parallel configuration: the policy maker can simply design socially-optimal standards as if all controls are verifiable, and then impose it wherever feasible.

There are, nevertheless, two caveats to this result on adopting a complete-information standard. First, though the firm's investment on the unverifiable control N is indirectly pushed up because of the high complete-information standard on control V, this investment is still lower than the socially-optimal level. As a result, social-optimality (as in the complete information benchmark) is not attainable. Second and as we will see shortly, this simple policy of standardization applies only to the parallel configuration, as optimal standards under the other two security configurations are sharply different.

Best-Shot Configuration

Similar to the last sub-section, we use backward induction to analyze best-shot configuration. Throughout this subsection, subscript "BC" means "best-shot configuration." Under best-shot configuration, $\omega(b_V, b_N) = b_V b_N = \exp(-\frac{m_V}{K_V} - \frac{m_N}{K_N})$. For any given standard s_{BC} imposed on control V, the firm's optimization problem in period 2 is:

$$\begin{aligned} \max_{m_V, m_N} U_F &= V_F - \exp(-\frac{m_V}{K_V} - \frac{m_N}{K_N}) D_F - m_V - m_N, \\ \text{s.t. } m_V &\geq s_{BC}. \end{aligned}$$

Lemma 3: *Under best-shot configuration and given standard s_{BC} on security control V:*

- i. *If $K_V \leq K_N$, $m_V^* = \max\{K_V \ln(D_F / K_V), s_{BC}\}$ and $m_N^* = 0$.*
- ii. *If $K_V > K_N$ and $s_{BC} \geq K_V \ln(D_F / K_N)$, $m_V^* = s_{BC}$ and $m_N^* = 0$.*
- iii. *If $K_V > K_N$ and $s_{BC} < K_V \ln(D_F / K_N)$, $m_V^* = s_{BC}$ and $m_N^* = K_N \ln \frac{\exp(-s_{BC} / K_V) D_F}{K_N}$.*

Unlike parallel configuration, under best-shot configuration firm's investments depend critically on the relative cost-efficiency of the security controls. When the verifiable control V is more cost-efficient (in that $K_V \leq K_N$), the firm should give up the unverifiable control N and focus its investment on control V (thus the term "best-shot"). Intuitively, though the marginal cost of defense increases in the security level of any security control, the multiplicative form of the breach probability function ($\omega(b_V, b_N) = b_V b_N$) implies that the firm will find that control V will always have a lower marginal cost of defense than control N at any security level if $K_V < K_N$. Therefore, it is not worthwhile to invest in the unverifiable control N.

The story is slightly more complicated when the unverifiable control N is more cost-efficient (i.e. $K_V > K_N$). In this case, the firm is forced to invest in the non-efficient control V. Lemma 3(iii) shows that, if the standard is not high, the firm will abide by the standard, yet will also invest in control N to take advantage of its cost-efficiency. If the standard is very high, as shown in Lemma 3(ii), the firm is forced to invest heavily on control V to the point where it does not see any benefit from additional investment on control N even if the latter is more cost-efficient. It is obvious that, from the firm's perspective, a standard on the verifiable control leads to inefficient investment when the other control is more cost-efficient. The next proposition describes the impact of the standard on the unverifiable control.

Proposition 3: *Under parallel configuration, the firm's investment on the unverifiable security control*

- i. *decreases in the standard on the verifiable control if the verifiable control is less cost-efficient and the standard is low enough (i.e. $s_{BC} < K_V \ln(D_F / K_N)$);*
- ii. *is zero otherwise.*

We next describe the policy maker's optimal decision in period 1. For ease of exposition, define $f(r) \equiv -r + r \ln r + r \ln(K_N / D_{SW}) + D_{SW} / D_F - \ln(K_N / D_F)$ and let \hat{r} be the solution to $f(\hat{r}) = 0$.

Lemma 4: *$s_{BC}^* = 0$ if $K_V / K_N \geq \max\{\hat{r}, 1\}$, $s_{BC}^* = K_V \ln(D_{SW} / K_V)$ otherwise.*

To understand the intuitions behind Lemma 4, we now introduce a second benchmark scenario -- *the naive-information benchmark*, which refers to the scenario where the policy maker is unaware of the existence of the unverifiable control. In other words, the policy maker naively (and incorrectly) believes that $\omega(b_V, b_N) = b_V$. This can be the case, for example, if the policy maker simply ignores all security controls that it cannot monitor and regulate. Alternatively, the naive-information benchmark may arise even for a policy maker that pays due diligence if a new type of security control is invented after the policy maker has already published the standard. From (3) and $\omega(b_V, b_N) = b_V$, we know the optimal standard under the naive-information benchmark is $K_V \ln(D_{SW} / K_V)$, which we refer to as the *naive-information standard*.

Proposition 4: *Under best-shot configuration, the policy maker should either impose the naive-information standard (if $K_V / K_N < \max\{\hat{r}, 1\}$) or not impose any standard. The firm will accordingly invest on only one security control.*

The policy maker's decision problem is more complicated under best-shot configuration (as compared to parallel configuration) because it now has to judge when to impose a standard. When the verifiable control is more cost-efficient (i.e. $K_V < K_N$), a high standard induces the firm to make socially-optimal investment. Furthermore and interestingly, the policy maker may force the firm to invest in the verifiable control *even if* it is less cost-efficient as compared to the unverifiable control, as in the case $\max\{\hat{r}, 1\}K_N > K_V > K_N$. Intuitively, in this case the policy maker is trading-off two effects: on the one hand, forcing the firm to invest heavily in the less cost-efficient control hurts firm profit; on the other hand, a high standard benefits consumer surplus -- as lacking a standard the firm will not invest as high even in the cost-efficient control. When the efficiency loss is not too high (i.e. K_V is upper-bounded by $\max\{\hat{r}, 1\}K_N$), the second effect dominates the first one from the policy maker's perspective.

Once the question of when to impose a standard is answered, the standard itself is remarkably simple: it is the naive-information standard. This finding under best-shot configuration contrasts sharply with the finding regarding the optimality of the complete-information standard under the parallel configuration.

Under best-shot configuration and given optimal standards, the firm will always put all investment into the "best-shot" security control. This result is consistent with prior theoretical findings such as Varian (2004).¹⁵ This is a unique characteristic of this security configuration as in all other security configurations, such as the serial configuration which we next discuss, we will see the firm investing in and balancing both security controls.

Serial Configuration

Serial and best-shot configurations are similar in that they have the same breach probability function:

$\omega(b_V, b_N) = b_V b_N = \exp(-\frac{m_V}{K_V} - \frac{m_N}{K_N})$. In other words, to compromise the digital asset and cause damage, in both

security configurations attackers have to breach both security controls. That said, a key difference between these two security configurations is that, under serial configuration, the first security control (e.g. V in Figure 1b) filters and blocks some attacks before traffic arrives at the second control (N in Figure 1b). As a result, one should expect a lower attack intensity -- conditional on how secure the first control is -- toward the second control in serial configuration.

Therefore, unlike in previous sub-sections where we normalize both attack intensities t_V and t_N to constant one, in this subsection only the first security control has a normalized attack intensity of one. In this paper we further restrict our attention to the case where the first security control is verifiable (as in Figure 1b). Thus $t_V = 1$. t_N is assumed as follows:

$$t_N = \alpha b_V \quad (4)$$

where α is a constant in (0,1]. The above linear equation is the simplest formula to capture the idea that, the better protection the first security control offers (thus a lower breach probability b_V), the less likely attacks can sneak through this first control and arrive at the second control. (4) can be rewritten as $t_N = t_N(m_V) = \alpha \exp(-m_V / K_V)$.

Similar to previous sub-sections, we use backward induction to analyze serial configuration. Subscript "SC" means "serial configuration." For any given standard s_{SC} imposed on control V, the firm's optimization problem in period 2 is:

¹⁵ Our best-shot configuration is mathematically a special case of total-effort in Varian (2004). Varian shows that firm strategies under total-effort and best-shot are identical. Note that this mathematical similarity stops at the best-shot configuration: our parallel, serial and later-to-be-discussed weakest-link configurations are all characteristically different.

$$\max_{m_V, m_N} U_F = V_F - \exp\left(-\frac{m_V}{K_V} - \frac{m_N}{K_N t_N(m_V)}\right) D_F - m_V - m_N,$$

s.t. $m_V \geq s_{SC}$.

Lemma 5: Under serial configuration and given standard s_{SC} on security control V:

- i. If $s_{SC} \geq K_V \ln \frac{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}{K_V}$, $m_V^* = s_{SC}$ and $m_N^* = K_N \alpha \exp(-s_{SC} / K_V) \ln(D_F / (K_N \alpha))$.
- ii. Otherwise, $m_V^* = K_V \ln \frac{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}{K_V}$ and $m_N^* = K_V (1 - \frac{1}{1 + \ln(D_F / (K_N \alpha))})$.

There are two similarities between serial and parallel configurations in terms of the firm's response to a security standard: first, a low enough standard has no impact on firm investments; second, if the standard is high enough, the firm's investment on the verifiable control will simply match the standard. These two security configurations, nevertheless, differs fundamentally in how the standard incentivizes the firm's investment on this unverifiable security control:

Proposition 5: Under serial configuration, the firm's investment on the unverifiable security control decreases in the standard on the verifiable control when the standard is high enough (i.e. when $s_{SC} \geq K_V \ln \frac{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}{K_V}$).

Under serial configuration, a high standard on the verifiable control results in a low probability of any attack passing through this control. As a result, t_N will be significantly lower than t_V (which is normalized to 1), which reduces the need to have strong security on the unverifiable control, as shown in Proposition 5.

That said, this reduction in the investment on control N is not as extreme as the one in the best-shot configuration: in the latter there is no investment at all on control N when the standard is high enough; while in the former investment on N is always positive. This is because a reduction in attack intensity t_N improves the marginal benefit of each unit of investment on control N because breach probability b_N is an increasing function of t_N . This improvement turns out to be significant enough: even if the verifiable control is ex ante more cost-efficient (i.e. $K_V < K_N$), from Lemma 5(i) it is straightforward to verify that $K_N t_N(m_V^*) < K_V$ always holds. Therefore, ex post and due to the reduction in attack intensity on control N, investment on this control becomes cost-efficient. Next we describe the policy maker's optimal decision in period 1 that maximizes social welfare.

Proposition 6: Under serial configuration, the policy maker should impose a standard of

$$s_{SC}^* = \max\left\{K_V \ln\left(\frac{K_N \alpha}{K_V} \left(\frac{D_{SW}}{D_F} + \ln \frac{D_F}{K_N \alpha}\right)\right), 0\right\}.$$

In this section we show that, when an unverifiable security control exists, the policy maker's decision on security standard (over the verifiable control) should be contingent on security configurations -- namely, how the verifiable and unverifiable controls work together in protecting the digital asset. This is because a (high enough) standard not only directly affects firm investment on the verifiable control, it also indirectly affects firm investment on the unverifiable control. Our analysis shows that this indirect effect on the unverifiable security control differs significantly among the three basic security configurations: under parallel configuration, a high enough standard *increases* firm investment on the unverifiable control; under best-shot configuration, whenever a high enough standard is imposed, it *does not affect* firm investment on the unverifiable control; under serial configuration, a high enough standard *decreases* firm investment on the unverifiable control. The policy maker thus should set different standards for different security configurations.

Another insight is that the right security standard needs not be always complicated. We find that the simple complete-information standard applies to the parallel configuration, and that the simple naive-information standard applies to the serial configuration (unless the cost-efficiency of the verifiable control is too low -- in this case no standard is optimal).

Standardization Under Strategic Attacks

In this section we consider strategic attacks, in which case the representative attacker strategically chooses her attack contingent on the characteristics (K_V and K_N) of the two security controls as well as her expectation of the security investments (m_V and m_N) taken by the firm.¹⁶ We limit our attention to the parallel configuration. We also limit our attention to attackers who strategically target the *weakest link* -- the security control that is most likely to be breached for any given attack intensity. In our model setup, the weakest link is the security control with the lowest ratio m_i / K_i , $i \in \{V, N\}$. To clearly differentiate the analysis in this section from the parallel configuration with non-strategic attacks in the previous section, hereafter we refer to the parallel configuration with strategic attacks as the *weakest-link configuration*.¹⁷

Accordingly, period 3 in the model timeline (Figure 2) now consists of two steps. In step 1, the representative attacker analyzes which security control is the weakest link. Because m_V and m_N are not observable to the attacker, we will look for a Perfect Bayesian Equilibrium (PBE) in period 1 in which the attacker forms unbiased beliefs over firm investments. In step 2, attackers concentrate their attacks on the weakest link with intensity T (and ignore the other security control). As we normalize attack intensity at 1 for both security controls under parallel configuration and nonstrategic attacks in the previous section, it is reasonable to assume $T \geq 1$ for this section as concentrated effort by the attacker is likely to be more dangerous for the weakest link as compared to non-discretionary attacks.

Using backward induction, we first analyze the PBE for any given security standard s_{WL} , where subscript "WL" stands for "weakest-link." Let p denote the attacker's probability of attacking the verifiable control V. Then in a PBE $p=1$ if the attacker expects control V to be the weakest link, i.e. $m_V / K_V < m_N / K_N$. $p=0$ if $m_V / K_V > m_N / K_N$. If $m_V / K_V = m_N / K_N$, the attacker can deem either security control as the weakest link, therefore mixed-strategy equilibria may arise in which $p \in [0,1]$. As shown in the next lemma, there is a unique mixed-strategy PBE if the standard s_{WL} is not too high:

Lemma 6: *Under weakest-link configuration and for any given standard s_{WL} for control V, firm and attacker strategies are:*

- i. If $s_{WL} < K_V T \ln \frac{D_F}{K_V T + K_N T}$, $m_V^* = K_V T \ln \frac{D_F}{K_V T + K_N T}$, $m_N^* = K_N T \ln \frac{D_F}{K_V T + K_N T}$, $p^* = \frac{K_V}{K_V + K_N}$.
- ii. If $K_V T \ln \frac{D_F}{K_V T + K_N T} \leq s_{WL} < K_V T \ln \frac{D_F}{K_N T}$, $m_V^* = s_{WL}$, $m_N^* = \frac{K_N}{K_V} s_{WL}$,
 $p^* = 1 - \frac{K_N T}{\exp(-s_{WL} / (K_V T)) D_F}$.
- iii. If $s_{WL} \geq K_V T \ln \frac{D_F}{K_N T}$, $m_V^* = s_{WL}$ and $m_N^* = K_N T \ln \frac{D_F}{K_N T}$, $p^* = 0$.

Parts (i) and (ii) in Lemma 6 reveal how the firm can best respond to a strategic attacker when the security standard is not too high: while the attacker tries to find and target the weakest link between the two security controls, the firm can simply eliminate any clear weakest link simply by setting the same ratio m_i / K_i for both controls $i = V$ and

¹⁶ While an attacker can often collect information relevant to cost-efficiency parameters K_V and K_N , such as prevailing market prices of various security products and security consulting services, it is much harder for the attacker to gauge specific investments a firm makes on their security controls, such as which specific security products are adopted, whether they are properly setup, and the IT labor assigned to monitor and maintain the security products. Accordingly, we assume m_V and m_N to be private knowledge to the firm.

¹⁷ Under either best-effort configuration or serial configuration, the representative attacker has to breach both security controls in order to compromise the digital asset. Therefore the concept of weakest-link does not apply.

$i = N$. As such, the attacker randomizes her attack on the two controls according to p^* .¹⁸ Another way of understanding the intuitions behind parts (i) and (ii) in Lemma 6 is to ask what happens if the firm does not select $m_V / K_V = m_N / K_N$ in equilibrium. If, for example, $m_V / K_V < m_N / K_N$ in equilibrium, the attacker's optimal strategy is then to set $p = 1$, i.e. to always attack control V. But anticipating the attacker's concentrated attack on control V, it is optimal for the firm to accordingly put all investment on control V and leave zero investment on control N, which however contradicts the assumption $m_V / K_V < m_N / K_N$.

Part (iii) in Lemma 6 carries an important observation regarding how an improperly high security standard can negatively affect the firm's defense against a strategic attacker. As the security standard is now very high, the firm is forced to invest heavily on the verifiable control, and finds it no longer worthwhile (in terms of investment needed) to improve the unverifiable control up to the same security level. As a result, the attacker correctly expects the unverifiable control to be the weakest link (i.e. $m_V / K_V > m_N / K_N$), and thus concentrates her attack on this control. We summarize this important observation in the following proposition.

Proposition 7: *The unverifiable control is the weakest-link if security standard on the verifiable control is sufficiently high (i.e. $s_{WL} \geq K_V T \ln \frac{D_F}{K_N T}$).*

Whenever the policy maker imposes an overly high security standard on the verifiable control in hope of improving security, Proposition 7 shows that, ironically, the verifiable control becomes irrelevant under strategic attacks because attackers will now completely focus on attacking the unverifiable security control. This finding is consistent with a number of recent security incidences. For example, in recent years the PCI Security Standards Council have imposed stricter standards on how merchants should secure up their databases in order to protect credit card information stored in them. Some industrial analysts have subsequently found PCI-DSS, such as internal corporate networks (Krebs 2009).

Therefore, given strategic attacks, the policy maker's optimal standard is capped by $K_V T \ln(D_F / K_N T)$. This cap is significant especially when the damage to social welfare, D_{SW} , resulting from a security breach, is high. To see this point, first note that this cap is irrelevant to D_{SW} . Second, if attacks are not strategic, as in the parallel configuration, from Lemma 2 we know that the standard will increase in D_{SW} and does not have an upper-bound. We next present the optimal standard decision by the policy maker under the weakest-link configuration:

Proposition 8: *Under weakest-link configuration, the policy maker should impose standard $s_{WL}^* = K_V T \ln \frac{D_{SW}}{K_V T + K_N T}$ if $D_{SW} < (1 + K_V / K_N) D_F$, or $s_{WL}^* = K_V T \ln \frac{D_F}{K_N T}$ otherwise.*

Concluding Remarks

This paper is a first study, from a policy maker's perspective, on whether and how the existence of an unverifiable security control affects optimal security standard on another related and verifiable security control. We find that, except in some cases under the best-shot configuration, the unverifiable control will affect optimal standard on the verifiable control. We further show that the specific security configuration -- namely, how the two controls together protect a firm's digital asset -- plays a critical role in deciding the optimal standard. Parallel configuration calls for a high standard, serial configuration calls for a low standard, and under best-shot configuration the unverifiable control has no impact on the standard if this control is less cost-efficient than the verifiable control. We further show that, under parallel configuration, whether attacks are strategic or not also affects the optimal standard: when attacks are strategically targeting the weakest-link control, optimal standard is capped.

¹⁸ As shown in the proof, p^* is the unique probability (of attacking the verifiable control V) that can sustain a mixed-strategy PBE.

This first research on the relationship between security control verifiability and security standard can be extended in a number of ways. First, in practice security configurations can be more complicated than the three basic forms discussed in this paper, and can involve more than two controls. The question of whether a complicated security configuration can always be decomposed into the three basic forms is intriguing. Second, subject to data availability, our research offers a number of empirically testable results, such as the ones on how security configuration affects a firm's investment on unverifiable controls. A follow-up empirical study will be valuable as to our knowledge there are few research that empirically studies how security standards affect firm investment on security controls and attacker strategy.

References

- Adams, A. and Sasse, M.A., 1999. "Users are Not the Enemy," *Communications of the ACM* (42:12), pp. 41-46.
- Battigalli, P., and Maggi, G., 2002. "Rigidity, Discretion, and the Costs of Writing Contracts," *The American Economic Review* (92:4), pp. 798-817.
- Bernheim B. D. and Whinston, M. D., 1998. "Incomplete Contracts and Strategic Ambiguity," *The American Economic Review* (88:4), pp. 902-932.
- Cavusoglu, H., Mishra, B., and Raghunathan, S., 2005. "The Value of Intrusion Detection Systems in Information Technology Security Architecture," *Information Systems Research* (16:1), pp. 28-46.
- Cavusoglu, H., Raghunathan, S., and Cavusoglu, H., 2009. "Configuration of and Interaction Between Information Security Technologies: The Case of Firewalls and Intrusion Detection Systems," *Information Systems Research* (20:2), pp. 198-217.
- Crawford, V., 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions," *The American Economic Review* (93:1), pp. 133-149.
- Culnan, M.J. and Williams, C. C., 2009. "How ethics can enhance organizational privacy: Lessons from the choicepoint and TJX data breaches," *MIS Quarterly* (33:4), pp. 673-687.
- Dye, R. A., 1993, "Auditing Standards, Legal Liability, and Auditor Wealth," *The Journal of Political Economy* (101:5), pp. 887-914.
- Geng, X., Huang, Y. and Whinston, A.B. 2002. "Defending Wireless Infrastructure Against the Challenge of DDoS Attacks," *ACM Journal on Mobile Networking and Applications* (7:3), pp. 213-223.
- Grossklags, J, Christin, N., and Chuang, J., 2008. "Secure or Insecure? A Game-Theoretic Analysis of Information Security Games," *Proceedings of the 17th International World Wide Web Conference*.
- Hendricks, K. and McAfee, R. P., 2006. "Feints," *Journal of Economics & Management Strategy* (15:2), pp. 431-456
- Keblawi, F. and Sullivan, D., 2007. "The Case for Flexible NIST Security Standards," *IEEE Computer Society*, June, pp. 19-26
- Loch, K. , Carr, H., and Warkentin, M., 1992. "Threats to Information Systems: Today's Reality, Yesterday's Understanding," *MIS Quarterly* (16:2), pp. 173-186.
- Krebs, R. 2009. "Hackers Test Limits of Credit Card Security Standards," *Washington Post*, April 16, available at voices.washingtonpost.com/securityfix/2009/04/the_number_scale_and_sophistic.html.
- Miller, A. R. and Tucker, C. E. 2010. "Encryption and Data Loss," *The Ninth Workshop on the Economics of Information Security, Harvard University, USA*, page 29
- Morse, E.A. and Raval, V, 2008. "PCI DSS: Payment card industry data security standards in context," *Computer Law & Security Report* 24 pp. 540-554
- Narasimhan, H. Varadarajan, V., Rangan C. P., 2010. "Towards a Cooperative Defense Model Against Network Security Attacks," *Tenth Workshop on the Economics of Information Security*.

- Romanosk, S., Telang, R., and Acquisti, A., 2008. "Do Data Breach Disclosure Laws Reduce Identity Theft?," *Seventh Workshop on the Economics of Information Security*, June 25-28, 2008.
- Ross, R., 2007. "Managing Enterprise Security Risk with NIST Standards," *IEEE Computer Society*, August, pp. 88-91
- Rothke, B., Mundhenk, D. 2009. "Sue the Auditor and Shut Down the Firm," (July 9). Available at http://www.csoonline.com/article/496923/Sue_the_Auditor_and_Shut_Down_the_Firm
- Tirole, J., 2009. "Cognition and Incomplete Contracts," *The American Economic Review* (99:1), pp. 265-294.
- Schechter S. E. and Smith, M. D. 2003. "How Much Security is Enough to Stop a Thief?," *Lecture Notes in Computer Science* 2742 pp. 122-137
- Schwartz, R. 1997. "Legal Regimes, Audit Quality and Investment," *The Accounting Review* (72:3), pp. 385-406
- Varian, H., 2004. "System Reliability and Free Riding," *Economics of Information Security*, Kluwer, pp 1-15.
- Willekens, M., Steele, A., and Miltz, D., 1996. "Audit Standards and Auditor Liability: A Theoretical Model," *Accounting and Business Research* (26:3), pp. 249-264.
- Zhao, X, Xue, L., and Whinston, A. B., 2009. "Managing Interdependent Information Security Risks: A Study of Cyberinsurance, Managed Security Service and Risk Pooling," *International Conference on Information Systems*, Phoenix, AZ.
- Zetter, K. 2009. "In Legal First, Data-Breach Suit Targets Auditor," *Wired* (June 2) Available at http://www.wired.com/threatlevel/2009/06/auditor_sued/

Appendix (Given page limit, we only provide proofs for key steps)

Proof of Lemma 1 (parallel configuration):

The firm's decision problem can be solved by Kuhn-Tucker condition.

$$\max_{b_v, b_N} L_{pc} = V_F - (\exp(-\frac{m_v}{K_v}) + \exp(-\frac{m_N}{K_N}) - \exp(-\frac{m_v}{K_v} - \frac{m_N}{K_N}))D_F - m_v - m_N + \lambda_{pc}(m_v - s_{pc})$$

$$\frac{\partial L_{pc}}{\partial m_v} = D_F \left(\frac{1}{K_v} \exp(-\frac{m_v}{K_v}) - \frac{1}{K_v} \exp(-\frac{m_v}{K_v} - \frac{m_N}{K_N}) \right) - 1 + \lambda_{pc} = 0$$

$$\frac{\partial L_{pc}}{\partial m_N} = D_F \left(\frac{1}{K_N} \exp(-\frac{m_N}{K_N}) - \frac{1}{K_N} \exp(-\frac{m_v}{K_v} - \frac{m_N}{K_N}) \right) - 1 = 0$$

For any inner solution, λ_{pc} must have a zero value, i.e. $m_v > s_{pc}$. The solutions to the above two equations are $m_v^* = -K_v \ln(\tilde{b}_v)$ and $m_N^* = -K_N \ln(\tilde{b}_N)$. When the standard, s_{pc} , is greater than m_v , however, two equations have a corner solution, i.e., $m_v = s_{pc}$. Therefore, the solutions of two equations are $m_v^* = s_{pc}$ and $m_N^* = -K_N \ln \frac{K_N}{D_F(1 - \exp(-s_{pc}/K_v))}$. Q.E.D.

Proof of Lemma 2:

$$\begin{aligned} U_{sw} &= V_{sw} - (b_v + b_N - b_v b_N)D_{sw} - m_v - m_N \\ &= V_{sw} - (b_{pc} + \frac{K_N}{D_F})D_{sw} + K_v \ln b_{pc} + K_N \ln \frac{K_N}{D_F(1 - b_{pc})} \end{aligned}$$

$$\text{From } \frac{\partial U_{sw}}{\partial b_{pc}} = -D_{sw} + \frac{K_v}{b_{pc}} + \frac{K_N}{(1 - b_{pc})} = 0 \text{ we have } b_{pc}^* = \frac{D_{sw} + K_v - K_N - \sqrt{(D_{sw} + K_v - K_N)^2 - 4D_{sw}K_v}}{2D_{sw}}.$$

We can transform b_{PC}^* into the following: $b_{PC}^* = \frac{2K_V}{(D_{SW} + K_V - K_N + \sqrt{(D_{SW} + K_V - K_N)^2 - 4D_{SW}K_V})}$.

$\frac{\partial b_{PC}^*}{\partial D_{SW}} = \sqrt{(D_{SW} + K_V - K_N)^2 - 4D_{SW}K_V} + [(D_{SW} - K_V - K_N)]$. If $D_{SW} > K_V + K_N$, b_{PC}^* is a decreasing function of D_{SW} . Therefore, $b_{PC}^* < \tilde{b}_V$ which means s_{PC}^* is always larger than $-K_V \ln(\tilde{b}_V)$. Q.E.D.

Proof of Lemma 5 (serial configuration):

$L_{SC} = V_F - D_F \exp(-\frac{m_V}{K_V T} - \frac{m_N}{K_N t_N}) - m_V - m_N + \lambda_{SC}(m_V - s_{SC})$ where $\lambda_{SC} \geq 0$, $m_V \geq 0$, and $m_N \geq 0$.

$$\frac{\partial L_{SC}}{\partial m_V} = D_F \exp(-\frac{m_V}{K_V T} - \frac{m_N}{K_N t_N}) (\frac{1}{K_V T} - \frac{m_N t_N'(m_V)}{K_N t_N^2(m_V)}) - 1 + \lambda_{SC} = 0 \quad (5)$$

$$\frac{\partial L_{SC}}{\partial m_N} = D_F \exp(-\frac{m_V}{K_V T} - \frac{m_N}{K_N t_N}) \frac{1}{K_N t_N(m_V)} - 1 = 0 \quad (6)$$

$$\lambda_{SC}(m_V - s) = 0 \quad (7)$$

Given $t_N = \alpha \exp(-\frac{m_V}{K_V T})$, we can rewrite (5) as

$$\frac{\partial L_{SC}}{\partial m_V} = D_F \exp(-\frac{m_V}{K_V T} - \frac{m_N}{K_N t_N}) \frac{1}{K_V T} (1 + \frac{m_N}{K_N t_N(m_V)}) - 1 + \lambda_{SC} = 0 \quad (8)$$

From (6) and (8),

$$K_N t_N(m_V) = (1 - \lambda) K_V T / (1 + m_N / (K_N t_N(m_V))) \quad (9)$$

For inner solutions (i.e. $m_V > s$), we need $\lambda = 0$. From equation (9),

$m_N = K_V T - K_N \alpha \exp(-m_V / (K_V T))$. Plug this into (6), we have

$$K_N \alpha \exp(-m_V / (K_V T)) = K_V T / (1 + \ln(D_F / (K_N \alpha)))$$

Therefore, $b_V^* = \frac{K_V T}{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}$, $m_V^* = K_V T \ln \frac{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}{K_V T}$,

and $m_N^* = K_V T (1 - \frac{1}{1 + \ln(D_F / (K_N \alpha))})$. Next we check whether this is indeed an inner solution.

$m_V^* > s_{SC} \Leftrightarrow K_V T \ln \frac{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}{K_V T} > s_{SC}$. Therefore, if standard is low, we have the inner solution.

Otherwise, given $s_{SC} > K_V T \ln \frac{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}{K_V T}$, we have $m_V^* = s_{SC}$. From (6),

$$m_N^* = K_N \alpha \exp(-s_{SC} / (K_V T)) \ln(D_F / (K_N \alpha)).$$

Based on the results above, we know the optimal m_V is decreasing in s_{SC} . To calculate optimal s_{SC} , we only need to consider the case where $s_{SC} > K_V T \ln \frac{K_N \alpha (1 + \ln(D_F / (K_N \alpha)))}{K_V T}$. Given $b_V = \exp(-s_{SC} / (K_V T))$ and

$$b_N = \exp(-m_N^* / (K_N \alpha \exp(-s_{SC} / (K_V T)))) = \exp(-\ln(D_F / (K_N \alpha))) = K_N \alpha / D_F,$$

$$U_{SW} = V_{SW} - D_{SW} K_N \alpha / D_F \exp(-s_{SC} / (K_V T)) - s_{SC} - K_N \alpha \ln(D_F / (K_N \alpha)) \exp(-s_{SC} / (K_V T))$$

$$= V_{SW} - s_{SC} - \exp(-s_{SC} / (K_V T)) K_N \alpha [D_{SW} / D_F + \ln(D_F / (K_N \alpha))].$$

$$\frac{\partial U_{sw}}{\partial s_{sc}} = -1 + \exp\left(-\frac{s_{sc}}{K_v T}\right) \left(\frac{K_N \alpha}{K_v T}\right) \left[\frac{D_{sw}}{D_F} + \ln\left(\frac{D_F}{K_N \alpha}\right)\right].$$

If solution is inner, $s_{sc}^* = K_v T \ln \left[\left(\frac{K_N \alpha}{K_v T}\right) \left(\frac{D_{sw}}{D_F} + \ln\left(\frac{D_F}{K_N \alpha}\right)\right) \right]$.

Condition for inner solution: $K_v T \ln \left[\left(\frac{K_N \alpha}{K_v T}\right) \left(\frac{D_{sw}}{D_F} + \ln\left(\frac{D_F}{K_N \alpha}\right)\right) \right] \geq K_v T \ln \frac{K_N \alpha (1 + \ln D_F / (K_N \alpha))}{K_v T}$,

which is always true. Therefore, D_F . Q.E.D.