Resilience in Information Stewardship

Christos Ioannidis* David Pym† Julian Williams‡ Iffat Gheyas§

October 15, 2018

Abstract

Information security is concerned with protecting the confidentiality, integrity, and availability of information systems. System managers deploy their resources with the aim of maintaining target levels of these attributes in the presence of reactive threats. Information stewardship is the challenge of maintaining the sustainability and resilience of the security attributes of (complex, interconnected, multi-agent) information ecosystems. In this paper, we present, in the tradition public economics, a model of stewardship which addresses directly the question of resilience. We model attacker-target-steward behaviour in a fully endogenous Nash equilibrium setting. We analyse the occurrence of externalities across targets and assess the steward's ability to internalise these externalities under varying informational assumptions. We apply and simulate this model in the case of a critical national infrastructure example.

1 Introduction

The objective of information security managers is to protect the confidentiality, integrity, and availability of the information systems for which they are responsible, by adjusting their adopted security measures in response to the evolving threat environment. The optimal investment in such security measures has been studied initially by Gordon and Loeb (2002).

The dynamic responses of information security managers to evolving threats to confidentiality, integrity, and availability have been modelled by, among others, Ioannidis et al. (2013), August and Tunca (2006) and Arora et al. (2008). This paper builds on and substantially generalises the previous work undertaken in Ioannidis et al. (2013) to include hidden action on behalf of the target. By allowing hidden action, the space of outcomes that a policy maker must account for increases markedly.

*Aston Business School, Aston University, Birmingham B4 7ET, UK c.ioannidis@aston.ac.uk
†University College London, Dept. of Computer Science, London, WC1E 6BT, UK d.pym@ucl.ac.uk
‡Durham University Business School, Durham, DH1 3LB, UK julian.williams@durham.ac.uk
§Durham University Business School, Durham, DH1 3LB, UK iffat.gheyas@durham.ac.uk

This objective must be pursued in the context of an information ecosystem that is subject to finite degradation of performance because of internal and external influences. In the information ecosystem, threats to the confidentiality, integrity, and availability of individual components of the ecosystem can be transmitted to others, impacting negatively on their security status. In such an environment, the role of the information steward is to maintain the sustainability and resilience of the ecosystem's nominal operating capacity, so delivering the managers' desired levels of confidentiality, integrity, and availability.[1]

In many domains the steward is a public policy maker regulating individual and collections of individuals and organizations defining sets of common rules and norms that should be observed to promote the sustainability and resilience of the ecosystem. Examples of this type of regulation are the security of Network and Information Security (NIS) directive from the European Union and the Federal Information Security Management Act (FISMA) in the United States.

However, there are well documented tensions between subjects of regulation, usually a subset of society, and those imposing regulations on behalf of society at large. For instance, time preferences and risk bearing might have quite different tolerances for those in stewardship roles and those being stewarded. Moreover, the very nature of investment in information security is quite complex. Assets have varying degrees of exposure under different modalities of use. An illustrative example is the separation of operational assets (such as industrial control systems) and assets used in corporate information systems (email, billing and customer record management).[2]

Modern information systems have high degrees of functionality, hence organizations can run several aspects of their operations in many different ways. In some cases, assets might be transferable across different information systems, with varying degrees of transparency to the regulator. An example might be using hardwired and micro-wave transmission systems for communication between substations in bulk electricity transmission versus internet protocol (IP) and cellular wireless (3G) data transmission. Both techniques offer the required dual redundancy approach, but the IP and 3G approach have several common risk factors as the standard implementation in 3G requires some IP configuration. This is in contrast to the completely separate tracks for the hardwire and microwave alternatives. Furthermore, as the IP/3G approach can be easily embedded in the typical corporate information systems infrastructure, the degree of connectivity to potentially insecure components increases.[3]

The degrees of vulnerability of an information system is commonly referred to as the

[1]Empirical exploration of the information stewardship problem can be found in Baldwin et al. (2017) and Allodi (2012) as examples.

[2]See Dehning and Richardson (2002); Maruster et al. (2008); Piotrowicz and Cuthbertson (2009); Pym et al. (2011) and Pym and Sadler (2010) for the notion of stewards within a public policy and commercial information ecosystem settings.

[3]See EU Green Paper on Energy Policy (2006); European-Commission (2006, 2008, 2012) and Govt. (2013), for a sample of legal frameworks on cyber and information security.

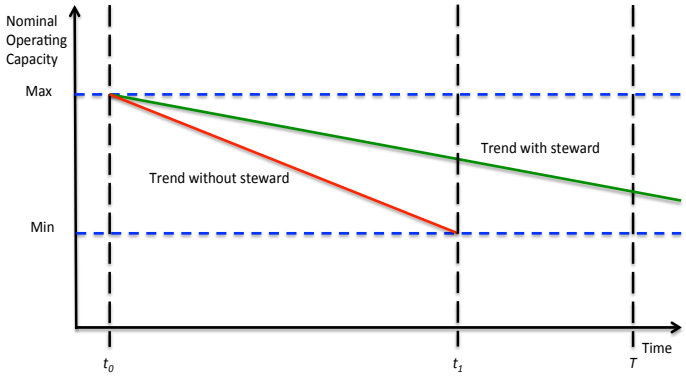

Figure 1: The concept of sustainability in information stewardship.

'attack-surface'. Indeed, audit requirements based on securing previous 'non-generic' information systems with specific physical features are redundant when a firm decides to shift to a new architecture. Hence, a potentially attractive feature of reshuffling an organizations information systems is to avoid costly regulation. It is to this dynamic aspect of stewardship that this paper is primarily directed our primary conceptual design will be based on the narrative concept of stewardship and specifically the notions of sustainability and resilience, which we will now briefly review.

1.1 Sustainability

By the sustainability of a system, subject to finite degradation caused by a persistent stream of attacks, we mean its tendency to remain within specified levels of nominal operating capacity. The graph in Figure 1 depicts this concept in static framework.

During successive tenures, stewards adopt policies to extend the system's lifetime. This is illustrated in Figure 2, in which interventions by the steward can be seen to maintain the system within its intended operating zone. From a game-theoretic point of view, managing the ecosystem to achieve this objective is a mechanism design problem for the steward.

In a recent paper, Ioannidis et al. (2013) have explored the steward's responses to maintain sustainability in the presence of endogenous attacks. They show that the presence of the steward increases investment in information security and, more importantly, reduces the number of attacks, thus retarding system degradation.

1.2 Resilience

In addition to the persistent stream of attacks to which a system is regularly subjected, occasionally the impact of an attack will be such that the system diverges from its predefined operational bounds. By resilience, we mean the ability of the system to return its

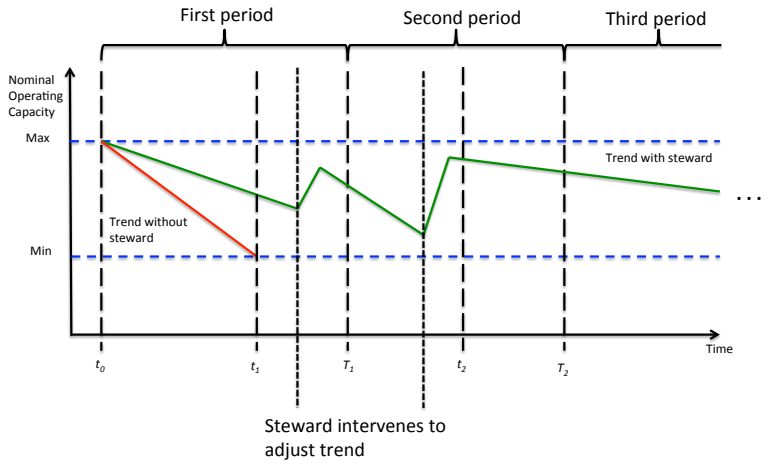

Figure 2: A multi-stage sustainability problem with periodic technology resets.

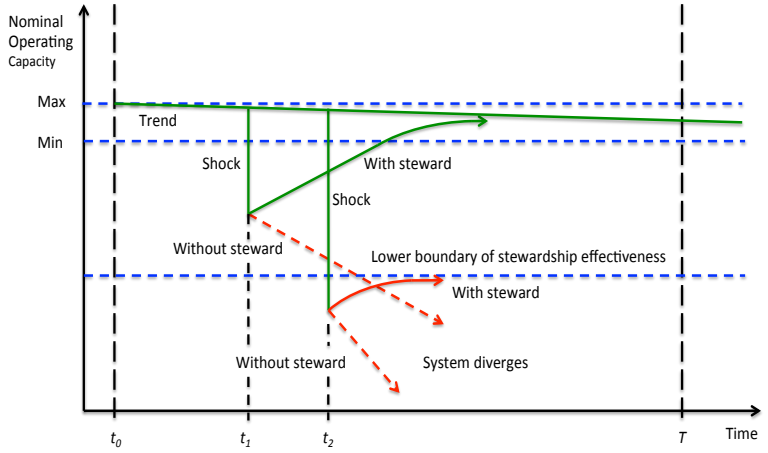

Figure 3: The concept of resilience in information stewardship.

operating capacity to within the specified bounds. Our notion of resilience is consistent with that discussed in Xie et al. (2005) and Hall et al. (2013).

The steward seeks to minimize, subject to a range of endogenous constraints, the time for which the system operates outside of the specified operating limits. Figure 3 depicts the concept of resilience in static framework in the presence of the information steward.

The remainder of this paper is organized as follows: in § 2, we present a model of resilience, with and without the information steward, with varying degrees of influence over the inhabitants of the ecosystem; in § 3, we present a detailed example of the model,

illustrated by simulations, and informed by the operations of Industrial Control Systems (ICSs); finally, § 4 provides a summary of our contribution and some directions for future research.

2 The Model

We have argued that the challenge for the information steward is to maintain the sustainability and resilience of target levels of confidentiality, availability, integrity, and investment for information ecosystems operating in potentially hostile environments. In previous work Pym et al. (2013) and Ioannidis et al. (2013), we have shown how to model sustainability, placing our discussion in the broader context of stewardship. We now extend our treatment of sustainability to account for resilience.

This extension is not merely incremental. Rather, it introduces significant additional and conceptual challenges. A key feature of our account of resilience is that we illustrate how thresholds for the effectiveness of stewards emerge from the underlying model of the response of information ecosystems to the hostility of the environment.[4]

As in our previous work on sustainability, see Ioannidis et al. (2013), our approach is one of mechanism design in which the targets of attacks are expected-loss-minimisers, subject to diminishing marginal returns on security investment. Similarly, attackers are modelled as rational agents. They are assumed to have utility functions, with well defined preferences, which can be used to capture their behavioural choices with respect to a variety of consumption goods, which maybe converted to monetary certainty equivalents, gained from successful attacks. We consider a set of N_T ex-ante identical targets choosing to allocate defensive resources that mitigate the harm from attacks. In a departure from previous models, the targets need to solve, simultaneously, a multi-dimensional resource allocation problem. Let the subscripts h and l represent to potential areas of allocation of assets, where h and l denote the areas of high and low security where information assets are held, and let $x_h \geq 0$ and $x_l \geq 0$ denote the one-off investments made at time t_0 in securing assets located in the corresponding areas. Finally, we define z to be a switching variable such that a fraction, $0 \leq z \leq 1$, of assets is allocated between h and l.

Our model depends crucially on two key (vector) parameters. First, we consider the elasticity of attacking intensity denoted by the vector α. This is the parameter that captures the marginal effectiveness of an additional attacker per target (η) entering the ecosystem. Whilst, in general, one would consider an ecosystem of participants having n types of assets of interest, without loss of representation of resilience, our model restricts is limited to two types of assets. In this case, we need to consider just two elasticities, α_l and α_h with corresponding η_l and η_h, which are associated with low and high levels of difficulty in securing assets. Second, we consider parameters ψ_l and ψ_h, which capture the relative

[4]Note by that the hostility of the environment we mean a representation of the capacity of attackers rather than simply the success or failure of an individual attack.

rate of risk reduction for additional security investments by targets in each asset class (x_l and x_h).

Let $\tilde{\sigma}_{i\in\{l,h\}} : \mathbb{R}_+ \to [0,1]$ be a function that determines the instantaneous time t risk for a fixed time-horizon, where $(t_0, T) = \{t \mid t_0 < t < T\}$. When properly specified we can interpret $\tilde{\sigma}$ as the instantaneous probability of a successful attack. We refer to z as the 'asset allocation' and the quantities x_l and x_h as the 'investment allocation', stated combinations of all three are referred as 'allocation bundles'.

Our assumption is that increased investment $x_{i\in\{l,h\}}$ reduces the probability of a successful attack; that is, $\partial\tilde{\sigma}_{i\in\{l,h\}}/\partial x_{i\in\{l,h\}} < 0$, ceteris paribus. However, along with increasing investment there is a decreasing marginal reduction in the probability of a successful attack, $\partial^2\tilde{\sigma}_{i\in\{l,h\}}/\partial x^2_{i\in\{l,h\}} > 0$. Similarly, with increased attacking intensity $\eta_{i\in\{l,h\}}$ on the particular area of allocation there should be a corresponding increase in the probability of a successful attack, $\partial\tilde{\sigma}_{i\in\{l,h\}}/\partial\eta_{i\in\{l,h\}} > 0$.

A functional form for $\tilde{\sigma}$ that satisfies these conditions is the following multiplicative model:

$$\tilde{\sigma}_i = e^{-\psi_i x_i}\eta_i^{\alpha_i}, \quad i \in \{l,h\}. \tag{1}$$

Under this formulation, there is an upper bound on $\eta_{i\in\{l,h\}}$ of $\ln\eta_i^* < \alpha_i^{-1}x_i\psi_i$, for $i \in \{l,h\}$, such that $\tilde{\sigma}_i$ may still be interpreted as probability of a successful attack. Here $\psi_{i\in\{l,h\}}$ is the relative marginal decrease in $\tilde{\sigma}_{i,i\in\{l,h\}}$ for a unit increase in $x_{i\in\{l,h\}}$ whilst $\alpha_{i\in\{l,h\}}$ is the elasticity of attack.

In this model, we assume that attacker externalities are driven by the diffuse-attacking-mass approach first suggested in Pym et al. (2013) and refined in Ioannidis et al. (2013). In this approach attackers are assumed to be ex-ante identical and randomly allocated to targets with identical probability $1/N_T$. Attackers are assumed to be able to make independent decisions on the type of attacks. A useful interpretation of the attacker cost per unit is that attackers need to develop an attacking tool at cost c each time they engage a target. The attacker then chooses the medium by which it seeks to monetize its attacking effort (in the case of terrorists, for example, monetization is via utility equivalents). An example could be corporate network information channels versus industrial control systems. Attackers, at inception, may not know which target they intend to attack. From the viewpoint of the steward in this setting, it is irrelevant who is attacking the targets. From the target-attacker transaction point of view, the salient point is the aggregate level of loss incurred in the presence of attacking intent.

Let the number of attackers for each asset area be $N_{A,i\in\{l,h\}}$. The ratio of attackers per target is the attacking intensity $\eta_{i\in\{l,h\}} = N_{A,i\in\{l,h\}}/N_T$. Let the reward $R > 0$ for a successful attack be proportional to the assets allocated in each area, h and l, and for simplicity let the fraction of reward $\zeta_{i,i\in\{l,h\}}$ from attacks be the same as the fractions within the asset allocations, hence $\zeta_{i=l} = z$ and $\zeta_{i=h} = 1 - z$. Set $\gamma = c/R$ to be the cost ratio of attack, where c is the unit cost of a single attack. When the attacker's time

preference is described by δ, the profit function for a single attacker is

$$\tilde{\Pi}_{A,i} = \int_{t_0}^{T} e^{-\delta t} \zeta_i \eta_i^{-1} \tilde{\sigma}_i \left(x_i, \eta_i\right) dt - \gamma, \quad i \in \{l, h\}. \tag{2}$$

We assume that attackers do not coordinate attacks (or are commissioned by a single attacker) and rewards are claimed on a first-winner-takes-all basis. Attackers are assumed to be drawn from a pool and make one-off entry decisions until marginal cost and marginal benefit are equal and hence $\tilde{\Pi}_{A,i} = 0$.

For the targets of such attacks, let $L > 0$ be an instantaneous value of assets at risk from attack and $\beta \in \mathbb{R}$ be a subjective discount rate determining the time preferences of all targets. The risk neutral expected loss over the time horizon $t_0 < t < T$, is given by

$$\tilde{V}_L = \int_{t_0}^{T} e^{-\beta t} \left(z \tilde{\sigma}_l \left(x_l, \eta_l\right) L + \left(1 - z\right) \tilde{\sigma}_h \left(x_h, \eta_h\right) L\right) dt + x_l + x_h. \tag{3}$$

The optimal allocation bundle $(z^{\diamond}, x_l^{\diamond}, x_h^{\diamond})$, when attacking intensity is exogenous, is the simultaneous solution of $\{\partial \tilde{V}_L / \partial x_l = 0, \partial \tilde{V}_L / \partial x_h = 0, \partial \tilde{V}_L / \partial z = 0\}$. By construction, if $\alpha_{i \in \{l,h\}} > 0$, $\psi_{i \in \{l,h\}} > 0$, $L > 0$, $\beta > 0$ and $z \in (0, 1)$, a minimum of this function exists. By assumption we set that the optimal allocation must be either $(x_{i \in \{k,h\}}) \in \mathbb{R}_+$ when $(\eta_{i \in \{k,h\}}) \in \mathbb{R}_+$ or, if the minimum lies at $x_{i \in \{l,h\}} < 0$, then $x_{i \in \{l,h\}}^{\diamond} = 0$. Similarly, we impose the inequality constraint that $0 \leq z^{\diamond} \leq 1$.

Assuming that targets and attackers have positive discount rates the appropriate time horizon, T, for empirical analysis, maybe determined endogenously. Let λ be an arbitrarily large, but not infinite, number. For a given discount rate, $\tilde{\theta} = \min(\delta, \beta)$, by construction

$$\lim_{T \to \infty} \int_{t_0}^{T} \tilde{\theta}^{-1} e^{-\theta t} dt = 1.$$

Therefore, the approximation of the time horizon $\tilde{T}$ covering the $1 - 1/\lambda$ proportion of the future losses is derived from $\tilde{T} = \ln(\lambda)/\tilde{\theta}$. In §(3) of this paper, we follow Ioannidis et al. (2013) and assume that $\beta > \delta$ and $\tilde{T} = \ln(\lambda)/\delta$, such that the interval t_0 to $\tilde{T}$ covers 90% of the expected present value; that is, $\lambda = 10$.

What is important, to the steward, is the overall mass of attacks against systems containing assets under the types l and h 'storage/operations' areas and this will be influenced by the aggregate behaviour of targets and attackers, rather than the microstructure of individual attack-defence interactions. The more attractive the ecosystem is to attackers, the greater the mass of attacks against its individual components.

Proposition 1 (Existence of Nash equilibria without the steward)**.** *In the absence of a steward a Nash equilibrium exists with the following equilibrium investment (1) and attacking intensity (2):*

1. *(Equilibrium Target Investment) Under the preceding assumptions, when* $\tilde{\sigma}_i = e^{-\psi_i x_i}\eta_i^{\alpha_i}$, *for* $i \in \{l, h\}$, *the Nash equilibrium allocations of* x_h, x_l *and* z *denoted* x_h^*, x_l^* *and* $z*$ *are*

$$\begin{aligned} x_i^* &= \frac{\alpha_i}{\psi_i}\ln\left(\frac{L\psi_i\psi_j^2\left(e^{\delta T}-1\right)^2}{\gamma\delta\beta(\psi_j+\psi_i)^2}\right) - \frac{\alpha_i\delta T}{\psi_i}, \quad i \in \{l,h\}, j \in \{l,h\}, j \neq i \\ z^* &= \frac{\psi_l}{\psi_h+\psi_l}. \end{aligned} \tag{4}$$

2. *(Equilibrium attacker intensity) Following from Part 1, above, the Nash equilibrium attacker intensities, denoted* η_l^* *and* η_h^* *are*

$$\eta_i^* = \left(\frac{\psi_j(e^{\delta T}-1)e^{-x_i^*\psi_i-\delta T}}{\gamma\delta(\psi_i+\psi_j)}\right)^{\frac{1}{1-\alpha_l}}, \quad i \in \{l,h\}, j \in \{l,h\}, j \neq i, \tag{5}$$

where ${x^*}_{i,\in\{l,h\}}$, *is the functional forms of the Nash equilibrium given in Part 1 (above).*

Proof. The proofs of Parts 1 and 2 are given in Appendix A.1. Note that in the multiplicative separably additive form of $\tilde{\sigma}_{i\in\{l,h\}}$ the Nash equilibrium allocation z^* is a simple function of $\psi_{i\in\{l,h\}}$ and when $\psi_l = \psi_h$ the allocation is equal. If we add the constraint $x_l + x_h = \tilde{x}$, where $\tilde{x}$ is a binding budget constraint, then the attacking effort in each asset area enters the solution for z. □

We demonstrate that, in this modelling approach, we do not have to impose an arbitrary constraint on $x_l + x_h$, to create conditions similar to the standard results obtained when optimizing under such budget restrictions.

2.1 Introducing the Steward

The subject of this paper is resilience, and why a system might not be resilient to security shocks through the choices of the individual components. The first stewardship action we evaluate replicates our previous work (on sustainability Ioannidis et al. (2013)) by postulating a Stackelberg policy framework in which the policy-maker stewarding the system sets rules relative to a target level of sustainability. When the steward is fully informed, our model reverts to the mechanism design problem discussed in Ioannidis et al. (2013), in which the steward is able to set a mandatory investment bundle (denoted by the lower bar) on the individual targets $(\bar{x}_l, \bar{x}_h)$, as well as imposing a specific asset allocation $\bar{z}$.

The Nash equilibrium allocations for the N_T targets assumes no social coordination. Therefore, the Nash equilibrium allocation (x_l^*, x_h^*, z^*) of defensive effort and corresponding attacking intensities (η_l^*, η_h^*) will not necessarily be the first best solution for Pareto efficiency. Let $(x_l^\dagger, x_h^\dagger, z^\dagger)$ be the Pareto efficient allocations for a given set of model parameters $(\alpha_{i\in\{l,h\}}, \beta, \gamma, \delta, \lambda, , \psi_{i\in\{l,h\}}, L)$.

A classical efficiency model, with the steward acting as a public policy-maker and imposing $(\bar{x}_l, \bar{x}_h, \bar{z})$ Ioannidis et al. (2013), demonstrates that Pareto efficiency is only guarantied when the subjective discount rate of the steward is equal to β, the common discount rate.

Indeed, the analysis in Ioannidis et al. (2013) illustrates that, from the subjective viewpoint described by targets' heterogeneous discount rates, the chosen values of $(\bar{x}_l, \bar{x}_h, \bar{z})$ cannot always be a Pareto efficient allocation, $(x_l^\dagger, x_h^\dagger, z^\dagger)$, when $\beta \neq \bar{\beta}$. However, there may exist constellations of parameters such the welfare of the individual agents have improved due to the presence of the steward despite their different discount rates. In this study, we do not explore such welfare comparisons.

2.2 The Fully Informed (x_l, x_h, z)-setting Steward

Let the steward's discount rate be $\bar{\beta}$. A fully informed steward sets a mandatory level of $(\bar{x}_l, \bar{x}_h, \bar{z})$ by minimizing the following loss function

$$\tilde{V}_P = \int_{t_0}^{T} e^{-\bar{\beta}t} \left(z\tilde{\sigma}_l \left(x_l, \eta_l^{\diamondsuit} \right) L + (1-z)\, \tilde{\sigma}_h \left(x_h, \eta_h^{\diamondsuit} \right) L \right) dt + x_l + x_h, \tag{6}$$

where $\eta_i^{\diamondsuit}(x_i, z)$ for $i \in \{l, h\}$ is the solution to

$$\int_{t_0}^{T} e^{-\delta t} \zeta_i \eta_i^{-1} \tilde{\sigma}_i \left(x_i, \eta_i \right) dt = \gamma, \quad i \in \{l, h\}, \tag{7}$$

in terms of (x_l, x_h, z). We can see that, by internalizing the attacker reaction curve, the fully informed policy-maker with identical time preferences to the homogenous targets $\bar{\beta} = \beta$ will set an allocation bundle $(\bar{x}_l, \bar{x}_h, \bar{z})$. Moreover, in Ioannidis et al. (2013), we show that for the one-dimensional investment case the allocation will be the Pareto efficient allocation from the point of view of both the steward and targets.

In the multi-allocation form of the model, where $\tilde{\sigma}_i = e^{-\psi_i x_i} \eta_i^{\alpha_i}$, for $i \in \{l, h\}$, proof that $(\bar{x}_l, \bar{x}_h, \bar{z}) = (x_l^\dagger, x_h^\dagger, z^\dagger)$, when $\bar{\beta} = \beta$ for all parameter combinations, is not possible because $\bar{z}$ does not have a closed form solution (other than in certain special cases; e.g., $\alpha_l = \alpha_h = \alpha$). One such case is to consider a constraint on weighting aspect of the bundle z across asset areas of the form: $z = \psi_h/(\psi_h + \psi_l)$, the Nash equilibrium allocation. Other constraints on z can be reasonably justified, as we subsequently demonstrate.

Proposition 2 (The fully informed steward)**.** *When the steward is fully informed the following results hold:*

1. *(Target investment with steward) When* $\tilde{\sigma}_i = e^{-\psi_i x_i}\eta_i^{\alpha_i}$ *and* $\bar{z} = \psi_h/(\psi_h + \psi_l)$*, the stewards optimal investment allocation* $(\bar{x}_l, \bar{x}_h)$ *is*

$$\begin{aligned}\bar{x}_i &= \frac{1}{\psi_i}\ln\left(\psi_j(\psi_i+\psi_j)^{\frac{1}{1-\alpha_j}}\right) + \frac{\alpha_i}{\psi_i}\ln\left(\frac{1}{\gamma}\delta\left(e^{\delta T}-1\right)\right) + \\ &\left(\frac{\bar{\beta}T(\alpha_i-1)}{\psi_i} - \frac{\delta T\alpha_i}{\psi_i}\right) + \frac{(\alpha_i-1)}{\psi_i}\ln\left(\frac{-\bar{\beta}(\alpha_j-1)}{L\psi_i\left(e^{\bar{\beta}T}-1\right)}\right), \\ &i\in\{l,h\}, j\in\{l,h\}, j\neq i \end{aligned} \tag{8}$$

2. *(Attacking intensity) Following from Part 1, above, the attacker intensity* $\eta_{i\in\{l,h\}}$ *is*

$$\bar{\eta}_i = \left(\frac{\psi_i\left(e^{\delta T}-1\right)e^{-\bar{x}_i\psi_i-\delta T}}{\gamma\delta(\psi_j+\psi_i)}\right)^{\frac{1}{1-\alpha_i}}, \quad i\in\{l,h\}, j\in\{l,h\}, j\neq i \tag{9}$$

where $\bar{x}_i$ *is given in Equation 8.*

Proof. The proofs of Parts 1 and 2 are given in Appendix A.2. The solution is again subject to an upper bound of $\eta_i^* < e^{\alpha_i^{-1}x_i\psi_i}$, for $i\in\{l,h\}$. We can compare the solutions in Equations 8 and 9 for the fully informed steward versus those in Equations 4 and 5. □

Proposition 3 (The Steward's Improvement)**.** *If* $\tilde{\sigma}_i = e^{-\psi_i x_i}\eta_i^{\alpha_i}$*, with* $\beta \geq \bar{\beta}$*, and* $\alpha_{i\in\{l,h\}} > 0, \psi_{i\in\{l,h\}} > 0, \gamma > 0, \delta > 0, L > 0$ *and the asset allocation is constrained to* $\bar{z} = \psi_h/(\psi_h+\psi_l)$*, then the steward's mandated investment* $\bar{x}_{i\in\{l,h\}}$ *is always greater than or equal to the Nash equilibrium investment bundle* ${x^*}_{i\in\{l,h\}}$*.*

Proof. The proof is obtained by substituting the expressions $\bar{x}_{i\in\{l,h\}}$ and ${x^*}_{i\in\{l,h\}}$ in Equations 4 and 8 into the functional form $\bar{x}_{i\in\{l,h\}} \geq {x^*}_{i\in\{l,h\}}$ and subject to the constraint $\beta \geq \bar{\beta}$. By solving the two inequalities simultaneously for each parameter relative to its own constraint, that is $\alpha_{i\in\{l,h\}} > 0, \psi_{i\in\{l,h\}} > 0, \gamma > 0, \delta > 0, L > 0$, by inspection the constraint $\beta > \bar{\beta}$ is never violated. The complete set of steps of the proof is relatively simple, albeit long algebraic manipulation. □

A useful by-product of the comparison between Propositions 1 and 2 is that we can define an upper bound on $\beta \geq \bar{\beta}$ such that the steward does at least as well as the Nash equilibrium even when the steward weights potential near-term losses more than the targets do). Again, this is covered in more detail for the one-dimensional case in Ioannidis et al. (2013).

The attacker intensities follow from the functional form of the Nash equilibrium, except with $\bar{x}_{i\in\{l,h\}}$ replacing $x^*_{i\in\{l,h\}}$, as in Equation 9. From the chosen functional form of $\tilde{\sigma}_{i\in\{l,h\}}$, $\bar{\eta}_{i\in\{|,h\}}$, we know that overall loss decreases with increasing $x_{i\in\{l,h\}}$, ceteris paribus, and we know, by construction, that $\bar{x} > x^*$ when we constrain using $\bar{z} = \psi_h/(\psi_h+\psi_l)$ and $\beta \geq \bar{\beta}$.

Following Ioannidis et al. (2013), we also consider an non-discounted metric $\tilde{V}_A$ that measures total cost from attacks and investment. We consider a detailed functional form in § 2.6. If $x_{i\in\{l,h\}}$ is set by the fully informed steward minimizing the objective function set out in Equation 6 and if $\tilde{\sigma}_{i\in\{l,h\}}$, $\bar{\eta}_{i\in\{|,h\}}$, with $\bar{z} = \psi_h/(\psi_h+\psi_l)$ and $\beta \geq \bar{\beta}$, then $\tilde{V}_A(\bar{x}_{i\in\{l,h\}})$ will be lower than $\tilde{V}_A(\bar{x}^*_{i\in\{l,h\}})$ for all combinations of $\alpha_{i\in\{l,h\}} > 0, \psi_{i\in\{l,h\}} > 0, \gamma > 0, \delta > 0$ and $L > 0$. Although, by construction, $\tilde{V}_A$ is not an objective function (its minima is unbounded in $x_{i\in\{l,h\}}$), the functional form of $\tilde{V}_A$ is useful in measuring the effect of the transition from $x^*_{i\in\{l,h\}}$ to $\bar{x}_{i\in\{l,h\}}$ free from the subjective discount rates β and $\bar{\beta}$.

2.3 Reducing the Steward's Abilities

The preceding notion of the steward assumed that it has the ability to impose $(\bar{x}_l, \bar{x}_h, \bar{z})$ on the targets and thus achieve a lower loss in $\tilde{V}_P$ than the Nash equilibrium allocation of (x^*_l, x^*_h, z^*). This result is useful, if unsurprising. The steward acts as a classic public policy-maker and sets the mechanism so that any attacking externalities are internalised by the targets. A less intuitive fact is that it is the steward's discount rate $\bar{\beta}$ that determines if, from the viewpoint of the targets with discount rate β, a Pareto efficient solution has been achieved.

For some parameter combinations of α and ψ, with $\beta \neq \bar{\beta}$, a natural tension will exist between the targets and the steward. If the steward requires, periodically, say, to have its power to set $(\bar{x}_l, \bar{x}_h, \bar{z})$ ratified by the targets, then it is likely that $\bar{\beta} \to \beta$. However, if the individual targets have heterogeneous discount rates, then the steward will never be able to attain the Pareto dominant solution unless each target is allowed to state its own discount rate. When this issue occurs, targets may overstate their discount rates — we can consider the security resource allocation to be part of a wider investment bundle for the targets — and their allocation bundles will simply tend back to the Nash equilibrium. We leave extended discussion of this effect to future work.

Moving back to the simplified ex-ante identical targets example, further interesting cases can be analysed by restricting either the action set and/or the information set of the steward. Indeed, these cases present the type of situations where the steward is unable to maintain the resilience of the ecosystem of targets in the presence of shocks to specific parameters (we focus on shocks to the technology parameters $\alpha_{i\in\{l,h\}}$ and $\psi_{i\in\{l,h\}}$). In the next section, we analyse the cases of the fully informed steward with limited action and, finally, the partially informed steward with limited action.

2.4 Full Information with Limited Action: Majority and Minority Cases

First, consider the case in which the steward can observe $x_{i\in\{l,h\}}$ and z, but can only impose constraints on x_h and z. We designate this the *majority-action-case*; that is, the steward controls the majority of variables affecting the allocation bundle (two variables) and the individual agents control a minority of it (one variable).

A similar case occurs when the the steward can only impose constraints on x_h and x_l, but observes z, the results are intuitively identical. In this case for the targets of attacks, x_h and z are now exogenous and their problem reduces to a one-dimensional optimization problem seeking to minimize

$$\tilde{x}_l(z, x_h, \eta_l, \eta_h) = \\ \arg\min_{x_l} \int_{t_0}^{T} e^{-\beta t} \left(z\tilde{\sigma}_l\left(x_l, \eta_l\right) L + (1-z)\tilde{\sigma}_h\left(\bar{x}_h, \eta_h\right) L\right) dt + x_l + x_h, \tag{10}$$

where $\tilde{x}_l(z, x_h, \eta_l, \eta_h)$ is the target's optimal solution for x_l as a function of the now imposed values of x_h, z, and the attacker intensity choices η_l and η_h. The intuition behind this approach is that the steward sets some collection of rules that identify the allocation z and then imposes some investment on that allocation x_h. The optimal bundle of (x_h, z) from the viewpoint of the steward is denoted $(\bar{x}_h, \bar{z})$. The steward therefore solves the other two thirds of the allocation using the following objective function:

$$(\bar{x}_h, \bar{z}) = \\ \arg\min_{x_h, z} \int_{t_0}^{T} e^{-\bar{\beta} t} \left((1-z)\tilde{\sigma}_l\left(\tilde{x}_l\left(z, x_h, \eta_l^{\diamondsuit}, \eta_h^{\diamondsuit}\right), \eta_l^{\diamondsuit}\right) L + z\tilde{\sigma}_h\left(x_h, \eta_h^{\diamondsuit}\right) L\right) dt \\ + \tilde{x}_l\left(z, x_h, \eta_l^{\diamondsuit}, \eta_h^{\diamondsuit}\right) + x_h, \tag{11}$$

where $\eta_{i\in\{l,h\}}^{\diamondsuit}$ is the solution to the attacker intensities given in Equation 7 As the steward anticipates the reaction of the target into the objective function for x_l, in this instance, almost all of the steward's objectives in (x_l, x_h, z) can be achieved. The the steward can impose itself on two out of the three degrees of freedom in the model. We can also see that if $\bar{\beta} = \beta$ (i.e., when the steward and targets have aligned time preferences), then the steward will achieve a risk profile broadly similar to the case when the steward controls all of the degrees of freedom (x_l, x_h, z). Whilst the steward can attain its desired risk expenditure trade-off, it can do so only at a lower level of efficiency (in terms of total initial cost $x_l + x_h$) than if the steward controls (x_l, x_h, z). Unless an arbitrary upper bound is placed on $x_h + x_l$, the steward can achieve a global minimum, for any given combination of $\alpha_{i\in\{l,h\}}$ and $\psi_{i\in\{l,h\}}$, by imposing a shift of assets (if necessary) into the high security domain. In the extreme case, in which $\bar{z} \to 1$, the steward has control over all assets and sets an unbounded investment in protection of $\bar{x}_h$ as an essentially one-dimensional optimization problem.

Reducing the steward's action space to only one of the three allocation variables (which we call the minority action case) provides a far greater limitation to its action space and substantially impairs the steward's ability to internalize the attacker externalities and adjust the total level of risk in response to a change in $\alpha_{i\in\{l,h\}}$ or $\psi_{i\in\{l,h\}}$. However,

the circumstances in which a steward would be able to observe behaviour, but have no direct influence on it, violate one of the presumed key roles of the steward in the ecosystem leave the motivation and analysis of this fully informed, but substantively limited, steward to future work.

2.5 The Partially Informed Steward with Limited Action: Minority Case

We skip case of the fully informed steward with limited action and move directly to a partially informed steward with minority action. This, in theory at least, is the most interesting case as it illustrates both the limitations of the steward's actions in response to changes in $\alpha_{i\in\{l,h\}}$ or $\psi_{i\in\{l,h\}}$ and also that, with limited information, the presence of the steward can in fact lead to a worse global outcome than the Nash equilibrium .

Let the steward observe and enforce only x_h. The steward can observe and internalize the externality in η_h, but cannot observe or enforce z or x_l. The targets then choose the investment and allocation bundle (x_l, z) following

$$(\tilde{x}_l, \tilde{z}; x_h, \eta_l, \eta_h) = \\ \arg\min_{x_l, z} \int_{t_0}^{T} e^{-\beta t} \left(\bar{z}\tilde{\sigma}_l\left(x_l, \eta_l\right) L + (1-\bar{z})\tilde{\sigma}_h\left(\bar{x}_h, \eta_h\right) L\right) dt + x_l + x_h. \tag{12}$$

The steward now solves the following minority optimization, with the steward's given information set:

$$\bar{x}_h(\eta_h) = \arg\min_{x_h} \int_{t_0}^{T} e^{-\beta t} \left(\hat{L}\tilde{\sigma}_h\left(\bar{x}_h, \eta_h^{\diamondsuit}\right)\right) dt + x_h, \tag{13}$$

where $\eta_h^{\diamondsuit}$ is the solution to the attacker entry problem from Equation 2, but only for the h asset class.

From the steward's point of view this is now

$$\int_{t_0}^{T} e^{-\delta t} \tilde{\zeta}_h \eta_h^{-1} \tilde{\sigma}_h\left(x_h, \eta_h\right) dt = \gamma. \tag{14}$$

Note that the steward now takes for given $\hat{L}$ as the value of losses; this is because the steward can no longer identify zL and $(1-z)L$, the steward is simply given $\hat{L}$ by the targets at an a-priori stage and is assumed to be exogenous. Similarly, whilst $\tilde{\zeta}_h$ is equal to z from the viewpoint of attackers and targets, it is simply a parameter unrelated to the overall asset allocation of the targets from the point of view of the steward. The steward is now unwittingly, not a Stackelberg policy maker, but in a Nash equilibrium with the targets and attackers, as the Steward is not able to observe the hidden action of the target.

The attackers are also solving their entry and exit decision for assets in allocation l, following $\int_{t_0}^{T} e^{-\delta t}\tilde{\zeta}_l \eta_l^{-1}\tilde{\sigma}_l\left(x_l,\eta_l\right)dt = \gamma$. This is unobserved by the steward, but is accounted for as part of a Nash equilibrium by the targets. For tractability, we assume that, from the viewpoint of the attackers, $\tilde{\zeta}_h$ is set exogenously and at a fixed ratio to $\hat{L}$. We are interested in the reaction of targets setting x_l and attackers choosing η_l, in order to demonstrate the natural limits that appear in the game and to analyse this case, we assume without loss of generality that $\hat{L}$ is exogenous by construction and $\tilde{\zeta}_h$ is already set in a pre-optimization between the attackers and the steward.

Proposition 4 (Attackers and Steward)**.** *When the steward has a) only partial actions and b) partial information.*

1. *(Asset Class h) If $\tilde{\sigma}_i = e^{-\psi_i x_i}\eta_i^{\alpha_i}$, for $i \in \{l,h\}$, the steward's objective function is as stated in Equation 13, and the attacker dynamics are as given in Equation 14, then the steward's optimal mandated investment allocation is*

$$\begin{aligned}\bar{x}_h &= \frac{1-\alpha_h}{\psi_h}\ln(\hat{L}(e^{bT}-1)\psi_h) - \frac{\alpha_h}{\psi_h}\ln(\gamma\delta(\tilde{\zeta}e^{\delta T}-\tilde{\zeta})) \\ &\quad \frac{1}{\psi_h}(\ln(\bar{\beta}\alpha_h-\bar{\beta})(1-\alpha_h)+\alpha_h T(\bar{\beta}-\delta)-\bar{\beta}T). \end{aligned} \tag{15}$$

   Following from the steward's choice, the attacker intensity, given the steward's actions $\bar{\eta}_h$, is given by

$$\bar{\eta}_h = \left(\frac{\tilde{\zeta}\left(e^{\delta T}-1\right)e^{-\bar{x}_h\psi_h-\delta T}}{\gamma\delta}\right)^{\frac{1}{1-\alpha_h}}, \tag{16}$$

   where $\bar{x}_h$ is as defined in Equation 15.

2. *(Asset Class l) We now consider the targets' and attackers' new equilibrium: if $\tilde{\sigma}_i = e^{-\psi_i x_i}\eta_i^{\alpha_i}$, for $i \in \{l,h\}$, and the targets' objective is as specified in Equation 12, then the equilibrium allocation bundle x_l, z will be*

$$\begin{aligned} x_l^{\ddagger} &= -\frac{1}{\psi_l}\ln(\bar{\eta}_h^{\alpha_h}) + \frac{\alpha_l}{\psi_l}(\ln(\bar{\eta}_h^{\alpha_h}) + \ln(\beta(e^{\delta T}-1)\bar{\eta}_h^{-\alpha_h}) - \\ &\quad \ln(\gamma\delta\psi_l(e^{\beta T}-1)L) + \beta T - \delta T) + \bar{x}_h\psi_h \end{aligned} \tag{17}$$

$$z^{\ddagger} = \frac{\beta\bar{\eta}_h^{-\alpha_h}e^{\bar{x}_h\psi_h+\beta T}}{L\psi_l(e^{\beta T}-1)}, \tag{18}$$

   and the attacker intensity η_l is given by

$$\eta_l^{\ddagger} = \left(\frac{z\left(e^{\delta T}-1\right)e^{-x_l^{\ddagger}\psi_l-\delta T}}{\gamma\delta}\right)^{\frac{1}{1-\alpha_l}}. \tag{19}$$

Proof. The proof of Parts 1 and 2 are given in Appendix A.3. Note that the steward's choice is effectively determined by three variables $\hat{L}$, $\tilde{\zeta}_h$, and $\bar{\beta}$. We assume that these are, a priori, in the steward's information set. It is worth reiterating that, at this stage, decisions regarding z, x_l, and η_l are, by construction, not included in this optimization. However, we do not have to impose these restrictions, as the derivative with respect to x_h of the steward's objective function, given the multiplicative form of $\tilde{\sigma}_i = e^{-\psi_i x_i}\eta_i^{\alpha_i}$, for $i \in \{l, h\}$, does not include x_l and η_l. So, the only implicitly restricted information is replaced by $\hat{L}$, $\tilde{\zeta}_h$. □

Note that we use the ‡ to denote this new equilibrium for the targets as it is not strictly a Nash equilibrium solution, but rather is Bayes-Nash equilibrium, in which the steward has prior values for $\hat{L}$ and $\tilde{\zeta}$. See Fudenberg and Tirole (1991) for an explanation of the difference between Nash and Bayes-Nash equilibria.

2.6 Measuring Resilience

Measuring the impact of technological shocks to $\alpha_{i,i\in\{l,h\}}$ and $\psi_{i,i\in\{l,h\}}$ and economic shocks to $\bar{\beta}$, β, δ, L, and γ is a challenging task and requires the creation of an arbitrary metric. In this paper, we combine the equilibrium values of $x_{i\in\{l,h\}}$, z, and $\eta_{i\in\{l,h\}}$ using a total non-discounted loss function for the risk component only. This is given as follows:

$$\tilde{V}_A(\tilde{v}, \tilde{u}) = \int_{t_0}^{\tilde{T}} \tilde{z}\tilde{\sigma}_l\left(\tilde{x}_l, \tilde{\eta}_l\right) L + \left(1 - \tilde{z}\right)\tilde{\sigma}_h\left(\tilde{x}_h, \eta_h\right) L dt \tag{20}$$

$$\tilde{v} = \left(\tilde{z}, \tilde{x}_{i\in\{l,h\}}, \tilde{\eta}_{i\in\{l,h\}}\right) \tag{21}$$

$$\tilde{u} = \left(\alpha_{i,i\in\{l,h\}}, \psi_{i,i\in\{l,h\}}\right), \tag{22}$$

where $\tilde{T} = -\ln(\lambda)/\theta$ and $\theta = \min(\bar{\beta}, \beta, \delta)$, for an arbitrary λ tending to zero. By construction, Equation 20 gives an un–discounted loss function, so that the value of the critical parameter $\tilde{T}$, which represents the step-size of the periods considered in the model (cf. Figure 2, for a multi-period sustainability model), is finite. $\tilde{v}$ is the collection of choice variables under the various stewardship options. $\tilde{u}$ is the collection of parameters that are subject to the technology shocks under consideration.

For a single period, resilience will be measured by a response function to shocks to the parameters $\tilde{u}$. Our choice of response function to technology shocks allows for shocks across the set of parameters $\tilde{u}$ either simultaneously of individually. It is given by the numerical evaluation of the following ordinary differential equation:

$$\tilde{I}(\tilde{u}) = \int_{t_0}^{\tilde{T}} \frac{\partial \tilde{z}}{\partial \tilde{u}}\tilde{\sigma}_l\left(\frac{\partial \tilde{x}_l}{\partial \tilde{u}}, \frac{\partial \tilde{\eta}_l}{\partial \tilde{u}}\right) L + \frac{\partial\left(1 - \tilde{z}\right)}{\partial \tilde{u}}\tilde{\sigma}_h\left(\frac{\partial \tilde{x}}{\partial \tilde{u}}, \frac{\partial \tilde{\eta}}{\partial \tilde{u}}\right) L dt, \tag{23}$$

$$\tilde{u} = \{\alpha_{i\in\{l,h\}}, \psi_{i\in\{l,h\}}\},$$

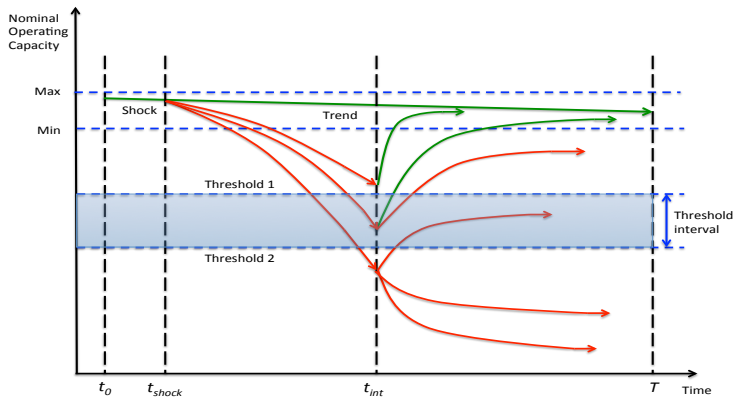

Figure 4: Resilience with an incompletely informed steward.

where each case has a set of functional forms for $\tilde{z}$, $\tilde{x}_{i\in\{l,h\}}$ and $\eta_{i\in\{l,h\}}\}$. We have denoted the three cases as follows: $\tilde{v} = v^*$ and $\tilde{u} = u^*$ for the Nash equilibrium, $\tilde{v} = \bar{v}$ and $\tilde{u} = \bar{u}$ for the fully informed steward and $\tilde{v} = \bar{v}^\ddagger$ and $\tilde{u} = \bar{u}^\ddagger$ for the partially informed steward with minority action case.

We are interested in establishing the thresholds, illustrated in Figure 4, which describe levels of system operating capacity, as measured by loss, for differing degrees of steward's effectiveness. We attempt to establish whether the system restores, through co-ordinated investment, to the target zone or not.

In our model, these thresholds reveal themselves as discontinuities, relative to shock size, in the solutions to Equation 23, below. Such discontinuities can be seen in our simulations as the asymptotes in Figures 6 and 8, for fully and partially informed stewards.

Let $\tilde{V}_A(v^*, u^*)$ and $\tilde{I}(u^*)$ be, respectively, the total non-discounted loss for the risk component under the Nash equilibrium and the corresponding collection of response functions. Similarly let $\tilde{V}_A(\bar{v}, \bar{u}), \tilde{I}(\bar{u})$ and $\tilde{V}_A(\bar{v}^\ddagger, \bar{u}^\ddagger), \tilde{I}(\bar{u}^\ddagger)$ be, respectively, the same pair of functions and collection of functions for the fully informed steward and the partially informed steward with minority action cases.

We can measure the effectiveness of the steward by comparing $\tilde{V}_A(v^*, u^*)$ to $\tilde{V}_A(\bar{v}, \bar{u})$. We can also evaluate the erosion in risk reduction caused by restricting the stewards information set and action space by pairwise evaluation of $\tilde{V}_A(v^*, u^*)$ and $\tilde{V}_A(\bar{v}, \bar{u})$ with $\tilde{V}_A(\bar{v}^\ddagger, \bar{u}^\ddagger)$.

To examine the impact of shocks and measure resilience we compare the response functions $\tilde{I}(u^*)$ and $I(\bar{u})$ to evaluate the impact of the fully informed steward. Finally, we can compare the resilience of the system when the stewards information set is restricted by comparing $\tilde{I}(u^*)$ and $I(\bar{u})$ to $I(\bar{u}^\ddagger)$, for varying sizes of shocks in $\tilde{u}$. In particular, we

focus on $\alpha_{i\in\{l,h\}}$.

3 Application to ICS, SCADA, and Corporate Networks

Industrial control systems (ICS) are ubiquitous in most large industrial firms and related organisations. A further common type of ICS are Supervisory Control and Data Acquisition (SCADA) systems. These systems are designed to automatically or semi-automatically control industrial processes. Examples of such systems can be found in petroleum exploration and processing, gas distribution, bulk electricity transmission, various parts of the nuclear industry and most manufacturing processes. Similar, or identical types of systems may be found in defensive applications, such as automatic air defence systems. ICS/SCADA systems are often very complex and deal with a large number of different types of sensors and actuators affecting the various aspects of the system in question. ICS/SCADA systems and the security of ICS/SCADA systems is not a new topic however, when many of the ICS/SCADA systems were first installed they were viewed as standalone assets and as such the major security concern was physical access to the control system or by physically tapping directly into the data acquisition sensors and/or the control communications to actuators. For our purposes, this distinction between ICS and SCADA is not critical and we refer generically to ICS/SCADA as a single type of assets within a target organization.

Our main question centres on whether a firm would seek to adjust its declared mix of ICS/SCADA and corporate information assets (we explicitly do not include physical assets in this example) to avoid costly regulation. We will assume that there exists some legacy regulation of certain types of ICS/SCADA systems and that firms can choose to replace some or all of the information architecture of theses systems with analogous technologies run on an unregulated corporate network. In terms of the model presented here, we have following set-up:

	Investments	Allocation	Risk-reduction rate	Attacker elasticity
ICS/SCADA	x_h	$1-z$	ψ_h	α_h
Corporate	x_l	z	ψ_l	α_l

In this paper, we run simulations for this model in which we shock the α_l; that is, the elasticity of attacker intensity against assets in the corporate network. Here we are modelling the situation in which we assume that the primary source of vulnerability is associated with the corporate network because of its more direct exposure to the Internet, with all of the associated vulnerabilities. As ICS/SCADA systems are increasingly interconnected with corporate systems, these vulnerabilities potentially affect core CNI systems; that is, the high-value (h) assets. Clearly, Equation 23 allows for a wider variety of experiments, which would allow us to explore different assumptions about the sources of vulnerabilities.

In the US, 1,900 bulk power system operators are regulated by The North American Electric Reliability Corporation (NERC), a not-for-profit organization with the role of coordinating the individual operators. This regulated ecosystem — of interconnected organizations — provides us with some indicative parameters for our experiments. Each operator will have a ICS/SCADA system that manages the bulk electricity transmission in their area. This will be a network of communications that monitor the physical network of power cables, transformers, and substations.

In addition to the ICS/SCADA assets, the various operators have corporate networks that provide on-going information services for the normal business activities for each operator. The corporate network has many of the same features as the ICS/SCADA system and there are elements of substitutability between the two. For instance, an operator could phase out using expensive fibre optic cables to communicate between ICS/SCADA systems and substations and replace this with a IP or 3G type communications.

A successful penetration of a corporate network that is integrated with an ICS/SCADA now provides attackers with a potentially more effective means of attacking the ICS/ SCADA system. The attacker can sit an learn the systems properties via sampling and observation of the ICS/SCADA systems normal operation and then use this information to either provide a priori information to improve the chance of success of a physical attack or actually attack the ICS/SCADA system directly through the corporate network.

As a community of targets, systematic underinvestment across all targets leads to increased attacking intensity and this provides a negative externality that requires coordination across targets in order to internalize this cost. We will illustrate three cases for this example, first where targets are unregulated and choose investment using the Nash equilibrium approach in § 2. We will then demonstrate the improvement that can be achieved by the fully informed steward. Finally, we will illustrate the deterioration in security when targets can shift assets from the oversight of the steward and the steward can no longer mandate investment. In each case, we illustrate the change in total risk with shocks to attacker elasticities and why targets may find it attractive to move assets from a regulated to an unregulated environment.

3.1 An Example Simulation

This simulation is designed to provide an overview of the intuition of our model and is not supposed to provide specific quantification for our proposed application. However, we have tried to stay close to real data when possible.

Let us assume that targets have a discount rate of 20% per annum ($\beta = \ln(6/5)$ continuous growth rate), in this case when $\lambda = 10$, the target time overall horizon is around $T = 12.3$ years. This appears to be a reasonable assumption for the amortization of information assets within a firm see, for example, the survey in Baldwin et al. (2005). For electricity transmission in the United States, the difference between physical and information assets can be found in NERC-Publications (2013) and FERC-Policy-Statement (2009).

The market risk premium for the US used to benchmark returns on investment in the private sector averages between 8% and 11% (the long run average return for equity over the stock market). Hence, a return of 20% is not unreasonable for a specific component of a firms asset base.

We assume that the societal discount rate used by the steward is much lower and ranges from $\bar{\beta} \to 0$ to $\bar{\beta} \to 1/10$, which is in line with social discount factors. In Ioannidis et al. (2013), we outline the various debates on the appropriate social discount factor to be applied in public policy scenarios. For certain areas of public policy debate such as climate change discount rates approaching zero a common for certain economic arguments relating to low carbon policies. For information stewardship the requirement is not so acute but significant differences between firm discount rates and societal discount rates remain, see for instance Nordhaus (2007).

For our starting numerical example, we assume that $\psi_l = \psi_h$; that is, the relative marginal risk reduction from investment in both asset classes is identical and fixed we assume that it is 1/100, 1/10 and 1/2, to represent low, medium and high effectiveness bands. This is a more difficult assumption to justify as there is very little literature on the efficacy of investment in security in this area, therefore our simulation covers a wide range of reasonable bands. In practice, placing a definitive upper and lower bound on the values of ψ_l and ψ_h is quite difficult. We know that attackers are present, hence the values cannot be too high. However, we do not observe overwhelming numbers of breaches, so we would expect the ability of firms to protect themselves must make reducing the numbers of attacks on information systems to average one or two successes per annum.

We arbitrarily fix $L = \$1\text{M}$, as an example, and divide all losses by L to give a per-dollar-at-risk measure. $\hat{L}$ is assumed to be half L. Starting from the Nash equilibrium assumption, if $\psi_l = \psi_h$, it follows that $z^* = 1/2$.

We set the attackers' discount rate to be $\delta = \ln(11/10)$, or a 10% discrete rate of return. From the viewpoint of attackers, the discount rate is analogous to an investment, as opposed to depreciation and amortization from the viewpoint of the targets. The most difficult parameter to set in the simulation is γ, as almost no data exists on the cost per attack to reward ratio. When $\gamma \to 0$, the cost per attack divided by reward indicates that either the rewards are very high or that the cost per attack is very low. When $\gamma = 0$, attacking intensity is infinite. This has not been observed, therefore we stick to finite values of $\gamma = 1/10$ or a 10% cost-reward ratio. The shock of interest is that to the elasticity of attack $\alpha_{i \in \{l,h\}}$ and, in particular, shocks to α_l.

Figures 5 and 6 illustrate the differing effects of shocks to the attacker elasticities, in the presence of a fully informed steward. Recall that by increasing the elasticity the attacker chances of success increase substantially, ceteris paribus.

However, the ecosystem will react to shocks to this elasticity, for the Nash equilibrium in the absence of a steward this will simply be an adjustment to the ratios illustrated in Proposition 1. However, for the fully informed steward there is a reactive coordinating entity, balancing current period investment with future uncertain losses. As shocks to the

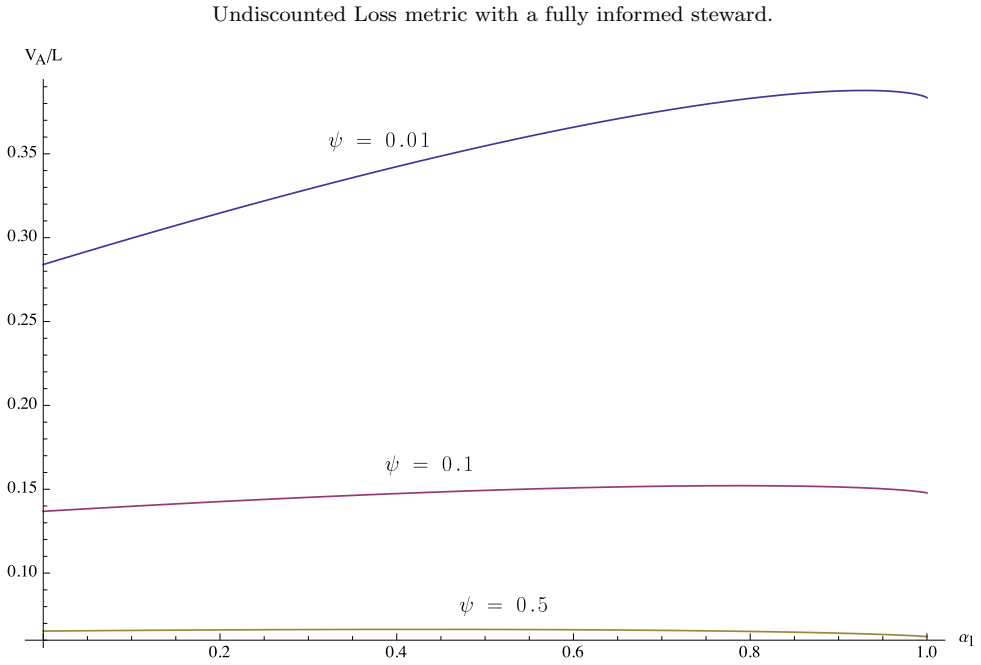

Figure 5: Steward's total non-discounted loss function, $\tilde{V}_A$, as a function of α_l. An important point to note is that this does not include the deterministic up-front investment, so this curve can actually slope downwards, even with increasing α_l. The upper (blue) curve represents $\psi_l = \psi_h = \psi = 0.01$, the middle (red) curve $\psi = 0.1$ and the lower (yellow) curve is $\psi = 0.5$. These values of ψ represent, respectively, low, medium, and high rates of risk reduction for additional investment.

attacking elasticity α_l increase, the steward utilizes this collective component to reduce the attacking intensity (rather than keeping the risk down by defensive effort x_l). The derivative of $\partial\bar{\eta}_l/\partial\bar{x}_l$ is now more important than $\partial\tilde{\sigma}_l/\partial\bar{x}_l$, where η_l is constant. The steward therefore finds an optimum by driving away all the attackers (as even small numbers are now very effective).

We see that, for all values of ψ, the fully informed steward provides a lower total non-discounted loss than the Nash equilibrium. This illustrates the beneficial effect of the steward. However, with larger values of ψ, the absolute effect decreases. The major benefit of the steward is in suppressing and adjusting the ecosystem to shocks and this effect is demonstrable for all three values of ψ.

Finally, we move to the partially informed steward with minority action, the total non-discounted loss $\tilde{V}_A$ and response function $\tilde{I}$ for shocks in α_l are plotted in Figures 7 and 8. In this case, the pattern is similar to the Nash equilibrium for small shocks. The targets, however have costly regulation in the h asset class and are under investing in the

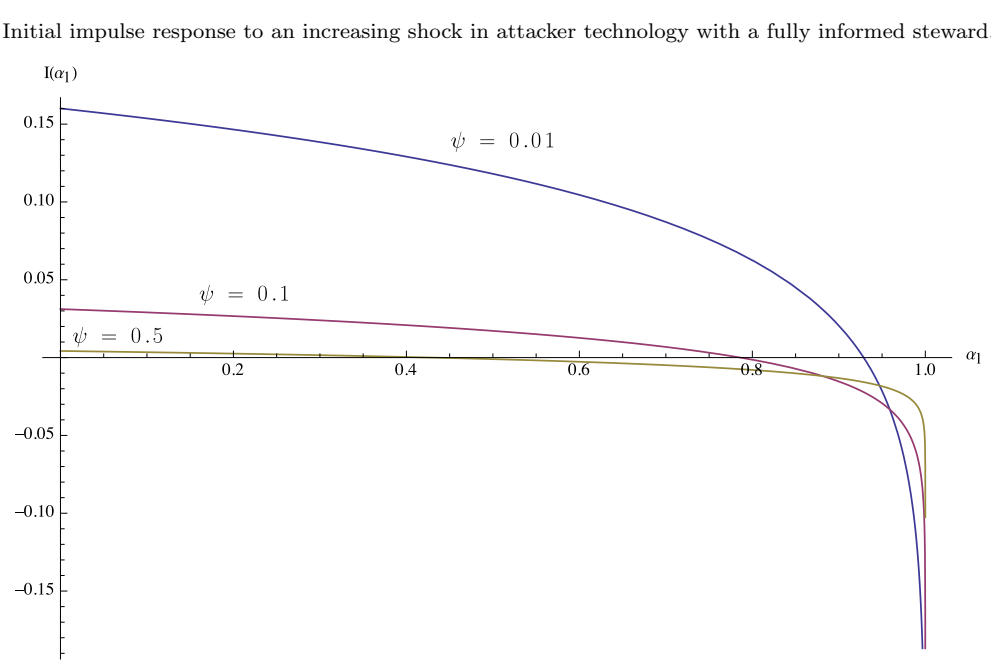

Figure 6: Steward's response function, $\tilde{I}(\alpha_l)$, as a function of an increasing shock in α_l, the abscissa values. Note that the steward now takes a positive action and seeks to manage the direction of the shock, as α_l becomes very large, the steward tolerates almost no attacking intensity and this effect is illustrated by the change in sign of the response.

l asset class. Unfortunately, in this case there is a discontinuity at $\alpha_l = 1$, so the total loss spikes prior to the shift in assets from l to h. This is a de facto boundary, as illustrated in Figure 4. We can see that before the steward can regulate the assets, the total risk will traverse the discontinuity, before the steward can actually manage the majority of assets that the targets have not declared. Here, we can see a case of an ecosystem that is not resilient and lies within the feasible boundaries of our example parameter sets.

3.2 Robustness of the Modelling Assumptions

The various forms of the model that we have proposed assume that targets are ex-ante identical. This is, of course, a simplifying assumption to lend tractability to the derivation and illustration of the specific effects that we are attempting to identify. However, this assumption is not as limiting as might be suspected.

The issue with the heterogeneity of the types of target — in terms of vulnerability or magnitude of loss — is that once we assume a steward in the role of a policy-maker determining mandatory investments, this steward would necessarily have to identify each target's Pareto efficient investment. For a large cross section of targets, this could poten-

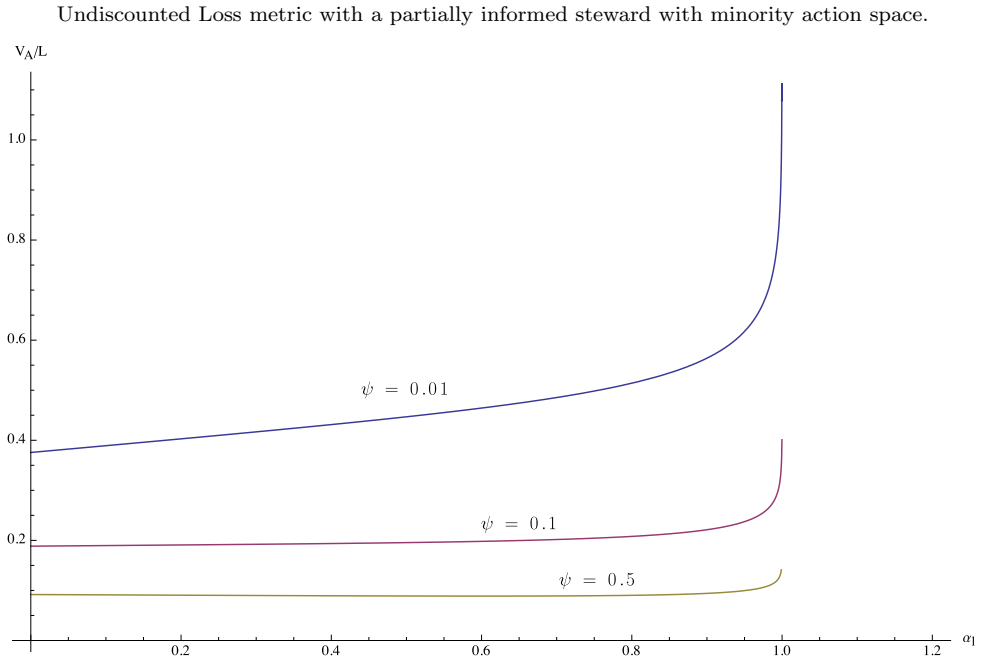

Figure 7: Partially informed steward with minority action total non-discounted loss function $\tilde{V}_A$ as a function of α_l. In this case, the targets maintain assets in the increasingly risky l class to avoid ostly regulation in h, however a discontinuity exists at α_l causing the loss function to spike before the assets are shifted back to the regulated domain.

tially be a costly information-gathering exercise.

Targets have the incentive to under-disclose their characteristics (e.g., because of budget pressures) and the remediation action of the steward is therefore rendered ineffective. A standard approach to this is contingent audit; see, for example, early research in this area in Kunreuther et al. (1985); Arrow (1983) and later work in Bohn and Deacon (2000); Benabou and Tirole (2006). Targets are asked to declare their characteristics — in terms of vulnerability and magnitude of loss — by the steward. In the event of an incident, there is a chance of audit (with known likelihood) and a large penalty (necessarily large enough for incentive compatibility) for incorrect prior identification to the steward. If the target has correctly identified their characteristics then no fine is levied. For the types of model proposed in Pym et al. (2013); Ioannidis et al. (2013), this approach would allow the steward to coordinate and mandate investment allocations with targets declaring their own vulnerability and loss characteristics. The allocation would therefore be Pareto efficient from the viewpoint of the steward. However, the allocation will not necessarily be Pareto efficient from the viewpoint of the target as the steward and target time preferences may be divergent. This is further exaggerated when the targets have the ability to hide assets

Initial impulse response to an increasing shock in attacker technology with a partially informed steward with minority action.

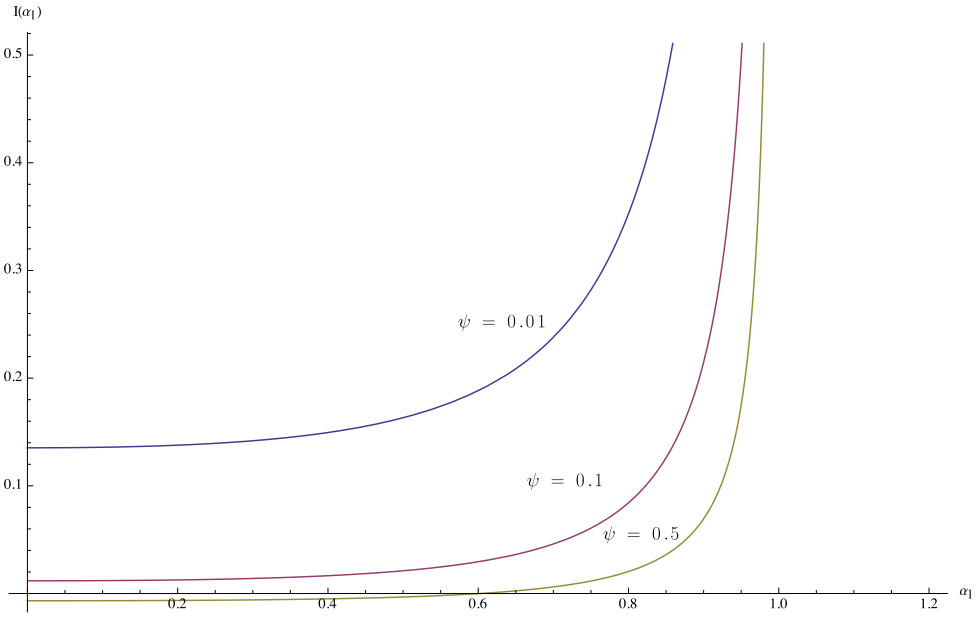

Figure 8: Partially informed steward with minority action response function $\tilde{I}(\alpha_l)$ as a function of an increasing shock in α_l, the abscissa values. Note that after a shock of $\alpha_l \to 1$, the function $\tilde{I}(\alpha_l)$ is not defined.

from the steward. If the steward's discount rate is very low relative to the targets' rates, then, under certain cases of the model, targets will move their assets to the class labelled l, by decreasing z substantially towards zero. This leaves very few assets in class h regulated by the steward.

When shocks — to the elasticity of the technology of attack in class l denoted α_l, say — result in a higher level of viable attacking intensity in equilibrium, targets can either choose to shift their assets to h by decreasing z or try to cope with the increasing attacks in l. Unfortunately, the game between attackers and targets in l results an equilibrium with externalities. Moreover, for certain versions of the model, the total risk when the steward takes action without observing x_l and z may be substantially higher (by orders of magnitude) than if the targets and attackers achieved a Nash equilibrium in the absence of the steward. We have discussed this case in §3.

Several rationales can be put forward to explain why the common knowledge assumption of z and x_l might not be shared with the steward by the targets. First, if β is much larger than $\bar{\beta}$, then the targets do not share the sustainability objectives of the steward, defined in terms of their time preferences (the targets are far shorter term than the policy-maker), therefore the targets may make a strategic choice, in an initial sub-game, to hide z and x_l from the steward. Second, an alternative explanation, that does not require another

mechanism to explain it, is that the targets and steward initially entered into a Stackelberg arrangement that is binding to the steward (to accomplish some sustainability target and internalize externalities in x_h). The steward sets x_h within the framework of the original agreement and this optimization rule continues through the life of the ecosystem, even when potentially new assets x_l exist. Indeed, the steward may simply not have sufficient information processing power to supervise all assets and then to cover them under relevant tort law liability conditions for the targets self-revelation approach to work. If there are a very large number of targets with highly diverse information assets, then the full audit may not be possible. Clearly, the model assumes the types of organization in x_l are ex-ante homogeneous.

One can postulate a set of regulations (in the form of fixed rules) designed by the steward and requiring the disclosure of targets' assets such that the investment x_h internalizes attacker externalities across targets (on the assumption that this is the complete set of assets). However, after a time, new assets not covered by the rules appear, or methods that allow targets to de-recognize these assets from the steward may exist.

4 Summary

This paper will make grim reading for any governmental, supra-governmental agency or firm that needs to act in a stewardship capacity over a complex information ecosystem. We illustrate two contrasting issues that complicate the management of this type of ecosystem. First, for almost all conceivable target–attacker interactions the presence of a steward is beneficial to overall risk reduction, by acting as a social coordinator and mandating investment that internalises externalities. Second, it is unlikely, however, that the time preferences of the steward, acting on behalf of society, and the targets will be aligned and as such the targets may not have the correct incentive to reveal their true type to the steward. In our framework this is in the form of hiding assets in an alternative unregulated asset class.

If the steward is able to observe these assets and mandate the majority of the investment bundle then the steward can still perform a beneficial role. However, when the steward acts on minority information and has limited action, the effect can be far worse than the Nash equilibrium when the steward is not present. Targets, maybe incentivized to store assets in increasingly insecure areas and this can substantially degrade the resilience of the ecosystem.

We have also provided a short example of this model using parameters designed to approximate the choice between holding information assets in a regulated ICS/SCADA system versus redeployment to a standard corporate information network. We demonstrate that a catastrophic scenario predicted by the model solutions under certain parameter configurations is possible for the domain of shocks assumed choices in this example.

Our major conclusions are also backed by qualitative analysis of the types of contracts

and regulations needed to ensure that the stewards information set is sufficient to maintain the information ecosystem. The types of of regulatory structures outlined herein are already beginning to be implemented in practice. Critically, the emphasis is often on specific audit schedules rather than placing the emphasis on targets to identify critical components with tort based penalties for failures to comply. Given the flexible nature of information systems prescriptive audit schedules will likely be made redundant as targets innovate around them to reduce inflexible costs.

The NIS directive includes mandated security monitoring and begins down the road to security audits. In the US, the NERC-CIP regulations are a specific set of audit and compliance models that require the identification of assets and an analysis of specific vulnerabilities so that some federal indemnities can be accessed. Indeed, the process of integrating NERC-CIP assets into corporate networks outside the domain of the federal regulator is one of the major drivers for this theoretical analysis.

Acknowledgments. We gratefully acknowledge support from the European Commission FP7-funded project 'Seconomics' and from National Grid plc. We are grateful to the reviewers and participants of the WEIS 2014 conference for their comments and advice on completing the first version of the paper. We also gratefully acknowledge the assistance of the referees and editor, Immanuel Bomze, of the European Journal of Operational Research for detailed and insightful comments in preparing the final version.

References

Allodi, L. (2012). The dark side of vulnerability exploitation. In G. Barthe and B. Livshits (Eds.), *Proc. International Symposium on Engineering Secure Software and Systems*.

Arora, A., R. Telang, and H. Xu (2008). Optimal policy for software vulnerability disclosure. *Management Science 54*(4), 642–656.

Arrow, K. (1983). *Behavior Under Uncertainty and Its Implications for Policy*. Institute for Mathematical Studies in the Social Sciences, Stanford University.

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

Baldwin, A., I. Gheyas, C. Ioannidis, D. Pym, and J. Williams (2017). Contagion in cybersecurity attacks. *Jnl of the Oper Res Soc 68*(7), 780–791.

Baldwin, J., G. Gellatly, M. Tanguay, and A. Patry (2005). Estimating depreciation rates for the productivity accounts. Technical report, OECD Micro-Economics Analysis Division Publication.

Benabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Bohn, H. and R. T. Deacon (2000). Ownership risk, investment, and the use of natural resources. *American Economic Review 90*(3), 526–549.

Dehning, B. and V. J. Richardson (2002). Returns on investments in information. technology: A research synthesis. *Journal of Information Systems 16*(1), 7 – 30.

EU Green Paper on Energy Policy, Commission of the European Communities, B. (2006). A European Strategy for Sustainable, Competitive and Secure Energy. Technical Report Com SEC(2006), Register of Commission Documents.

European-Commission (2006). Communication on a European Programme for Critical Infrastructure Protection. Technical Report [COM/2006/786], Register of Commission Documents.

European-Commission (2008). Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. Technical report, Register of Commission Documents.

European-Commission (2012). Position Paper of the TNCEIP on EU Policy on Critical Energy Infrastructure Protection. Technical report, Register of Commission Documents.

FERC-Policy-Statement (2009). Smart grid policy. Technical report, Federal Energy Regulatory Commission.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. MIT Press.

Gordon, L. and M. Loeb (2002). The economics of information security investment. *ACM Transactions on Information and Systems Security 5*(4), 438–457.

Govt., U. (2013). Executive Order 13636–Improving Critical Infrastructure Cybersecurity. Technical Report Federal Register, Vol. 78 No. 33, February 19, 2013, United States Library of Congress.

Hall, C., R. Anderson, R. Clayton, E. Ouzounis, and P. Trimintzios (2013). Resilience of the Internet Interconnection Ecosystem. In *The Twelfth Workshop on the Economics of Information Security (WEIS 2013)*. George Mason University.

Ioannidis, C., D. J. Pym, and J. M. Williams (2013). Sustainability in information stewardship: Time preferences, externalities, and social co-ordination. In *The Twelfth Workshop on the Economics of Information Security (WEIS 2013)*.

Kunreuther, H., J. Linnerooth, P. Knez, and R. Yaksick (1985). *Risk Analysis and Decision Processes: The Siting of Liquified Energy Gas Facilities: in Four Countries*. Springer–Verlag, London.

Maruster, L., N. R. Faber, and K. Peters (2008). Sustainable information systems: a knowledge perspective. *Journal of Systems and Information Technology 10*(3), 218 – 231.

NERC-Publications (2013). Second draft 2014 business plan and budget. Technical report, North American Electric Reliability Corporation.

Nordhaus, W. D. (2007). The "Stern Review" on the Economics of Climate Change. Technical Report w12741, NBER Working Paper.

Piotrowicz, W. and R. Cuthbertson (2009). Sustainability a new dimension in information systems evaluation. *Journal of Enterprise Information Management 22*(5), 492–503.

Pym, D. and M. Sadler (2010). Information Stewardship in Cloud Computing. *International Journal of Service Service, Management, Engineering, and Technology 1*(1), 50–67.

Pym, D., M. Sadler, S. Shiu, and M. C. Mont (2011). Information Stewardship in the Cloud: A Model-based Approach. In *Proceedings of CloudComp 2010*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (LNICST). Springer. To appear.

Pym, D. J., J. Swierzbinski, and J. Williams (2013). The need for public policy in information security. Working draft at `http://www.cs.ucl.ac.uk/staff/D.Pym/InfoSecPubPol.pdf`.

Xie, L., P. Smith, M. Banfield, H. Leopold, J. Sterbenz, and D. Hutchinson (2005). Towards resilient networks using programmable networking technologies. In *IFIP IWAN*.

A Internet Appendix: Extended Proofs

This is the internet appendix for *Resilience in Information Stewardship*. The following sections contain the extended proofs for propositions 1, 2 and 4 in the main paper.

A.1 Proof of Proposition 1

Part 1: Equilibrium Target Investment

Let $\tilde{\sigma}_{i\in\{l,h\}} : \mathbb{R}_+ \to [0,1]$. Evaluating the non-stochastic integral of loss over $t_0 = 0$ to T we find an analytic form the loss function,

$$V_L = \frac{1}{\beta} L \left(e^{\beta T} - 1\right) e^{-x_h\psi_h - x_l\psi_l - \beta T} \left(z e^{x_h\psi_h} \eta_l^{\alpha_l} - (z-1)\eta_h^{\alpha_h} e^{x_l\psi_l}\right) + x_h + x_l. \tag{24}$$

Differentiating with respect to $\tilde{x}_l$, $\tilde{x}_h$, and $\tilde{z}$ yields

$$\frac{\delta \tilde{V}_L}{\delta x_l} = 1 - \frac{1}{\beta} L z \psi_l \left(e^{\beta T} - 1\right) \eta_l^{\alpha_l} e^{-x_l\psi_l - T\beta} \tag{25}$$

$$\frac{\delta \tilde{V}_L}{\delta x_l} = 1 - \frac{1}{\beta} L (1-z) \psi_h \left(e^{\beta T} - 1\right) \eta_h^{\alpha_h} e^{-x_h\psi_h - T\beta} \tag{26}$$

$$\frac{\delta \tilde{V}_L}{\delta z} = \frac{1}{\beta} L \left(e^{\beta T} - 1\right) e^{-x_h\psi_h - x_l\psi_l - T\beta} \left(e^{x_h\psi_h} \eta_l^{\alpha_l} - \eta_h^{\alpha_h} e^{x_l\psi_l}\right). \tag{27}$$

Setting $\delta\tilde{V}_L/\delta x_l = 0$, $\delta\tilde{V}_L/\delta x_h = 0$ and $\delta\tilde{V}_L/\delta z = 0$, and solving simultaneously, we derive the unconstrained optimal allocation $(x_l^\diamond, x_h^\diamond, z^\diamond)$. When attacking intensity (η_l, η_h) is exogenous, this is analytically derived as

$$x_l^\diamond(\eta_l) = \frac{1}{\psi_l} \ln\left(\frac{(\lambda-1)L\psi_h\psi_l\eta_l^{\alpha_l}}{\beta\lambda\left(\psi_h + \psi_l\right)}\right) \tag{28}$$

$$x_h^\diamond(\eta_h) = \frac{1}{\psi_h} \ln\left(\frac{(\lambda-1)L\psi_h\psi_l\eta_h^{\alpha_h}}{\beta\lambda\left(\psi_h + \psi_l\right)}\right) \tag{29}$$

$$z^\diamond = \frac{\psi_l}{\psi_h + \psi_l}. \tag{30}$$

Note that $z^\diamond$ is a simple ratio of ψ_h and ψ_l. In this model, we apply no total budget constraint on x_h and x_l; that is, $x_h + x_l = x$, so no Lagrange multiplier needs to be added at this stage.

Part 2: Equilibrium Attacker Intensity

Following from the target decision-making process, we derive the attacker intensity function. Attackers enter the market for attacks in each asset class until they break even. When $\Pi_l = 0$ and $\Pi_h = 0$, we assume that attackers are randomly assigned targets, with identical probability $1/N_T$ for each attack, and that the first successful attacker wins the reward R.

Let $\gamma = c/R$, the cost of attack to reward. When $\tilde{\sigma}_{i\in\{l,h\}} : \mathbb{R}_+ \to [0,1]$, the profit functions for the attacker are as follows:

$$\Pi_l = \frac{1}{\delta} z \lambda^{-\frac{\delta}{\beta}} \left(\lambda^{\delta/\beta} - 1\right) \eta_l^{\alpha_l - 1} e^{-x_l\psi_l} - \gamma \tag{31}$$

$$\Pi_h = \frac{1}{\delta} (1-z) \left(\lambda^{\delta/\beta} - 1\right) \lambda^{-\frac{\delta}{\beta}} \eta_h^{\alpha_h - 1} e^{-x_h\psi_h} - \gamma. \tag{32}$$

Solving each function for the break-even attacking intensities $\eta_l^\diamond(x_l)$ and $\eta_h^\diamond(x_h)$, we compute the aggregate attacker reaction functions:

$$\eta_l^{\diamond}(x_l) = \left(\frac{z\lambda^{-\frac{\delta}{\beta}} \left(\lambda^{\delta/\beta} - 1\right) e^{-x_l \psi_l}}{\gamma\delta} \right)^{\frac{1}{1-\alpha_l}} \tag{33}$$

$$\eta_h^{\diamond}(x_h) = \left(\frac{(1-z)\lambda^{-\frac{\delta}{\beta}} \left(\lambda^{\delta/\beta} - 1\right) e^{-x_h \psi_h}}{\gamma\delta} \right)^{\frac{1}{1-\alpha_h}}. \tag{34}$$

The simultaneous Nash equilibrium is the best reply of the target to the best reply of the attacker (and vice versa), which is the simultaneous solution of $\{x_l^{\diamond}, x_h^{\diamond}, z^{\diamond}, \eta_l^{\diamond}, \eta_h^{\diamond}\}$.

Setting the Nash equilibrium defensive allocation (targets) and attacking intensity (attacker) as $\{x_l^*, x_h^*, z^*, \eta_l^*, \eta_h^*\}$, we obtain

$$\begin{aligned} x_l^* &= \frac{\alpha_l}{\psi_l}\left(-\ln\left(\gamma\delta L\psi_h\psi_l\left(e^{\beta T}-1\right)\right)+\ln\left(\beta\psi_h\left(e^{\delta T}-1\right)\right)+\beta T-\delta T\right) \\ &\quad +\frac{1}{\psi_l}\ln\left(\frac{L\psi_h\psi_l\left(e^{\beta T}-1\right)}{\beta\left(\psi_h+\psi_l\right)}\right)-T\beta \end{aligned} \tag{35}$$

$$\begin{aligned} x_h^* &= \frac{\alpha_h}{\psi_h}\left(-\ln\left(\gamma\delta L\psi_h\psi_l\left(e^{\beta T}-1\right)\right)+\ln\left(\beta\psi_l\left(e^{\delta T}-1\right)\right)+\beta T-\delta T\right) \\ &\quad +\frac{1}{\psi_h}\ln\left(\frac{L\psi_h\psi_l\left(e^{\beta T}-1\right)}{\beta\left(\psi_h+\psi_l\right)}\right)-T\beta \end{aligned} \tag{36}$$

$$\eta_l^* = \left(\frac{\beta\left(e^{\delta T}-1\right)e^{\alpha_l\left(\ln\left(\gamma\delta L\psi_h\psi_l\left(e^{\beta T}-1\right)\right)-\ln\left(\beta\psi_h\left(e^{\delta T}-1\right)\right)+\beta(-T)+\delta T\right)+T(\beta-\delta)}}{\gamma\delta L\psi_l\left(e^{\beta T}-1\right)}\right)^{\frac{1}{1-\alpha_l}} \tag{37}$$

$$\eta_h^* = \left(\frac{\beta\left(e^{\delta T}-1\right)e^{T(\beta-\delta)+\alpha_h\left(\ln\left(\gamma\delta L\psi_h\psi_l\left(e^{\beta T}-1\right)\right)-\ln\left(\beta\psi_l\left(-\left(e^{\delta T}-1\right)\right)\right)-\beta T+\delta T\right)}}{\gamma\delta L\psi_h\left(e^{\beta T}-1\right)}\right)^{\frac{1}{1-\alpha_h}}, \tag{38}$$

where $z^* = z^{\diamond}$. Assuming that $\alpha_{i\in\{l,h\}} > 0$, $\psi_{i\in\{l,h\}} > 0$, $L > 0$, $T > 0$, $\gamma > 0$, $\delta > 0$ and $\beta > 0$, then Equations 35 and 36 simplify to the result given in Proposition 1 (Part 1) and Equations 37 and 38 simplify to the equations given in Proposition 1 (Part 2). □

A.2 Proof of Proposition 2

Part 1: Target Investment with Steward

For the fully informed steward, setting $\bar{x}_{i\in\{h,l\}}$ and $\bar{z}$ the steward's objective is to minimize total aggregate loss for all targets. For our derivation, the targets are all assumed to be identical therefore the steward seeks to minimize

$$\tilde{V}_P = N_T \int_{t_0}^{T} e^{-\bar{\beta}t}\left(z\tilde{\sigma}\left(x_l, \eta_l^{\diamond}\right) + (1-z)\,\tilde{\sigma}\left(x_h, \eta_h^{\diamond}\right)\right) dt + N_T x_h + N_T x_l$$

where $\tilde{\sigma}_{i\in\{l,h\}} : \mathbb{R}_+ \to [0,1]$, ${\eta^{\diamond}}_{i\in\{l,h\}}$ is derived from Equations 33 and 34. The asset allocation z does not have a tractable analytic solution in this case, so for exposition purposes we focus on x_l and x_h when z is fixed. In this case, let us fix z to the Nash equilibrium solution, therefore $\bar{z} = z^{\diamond}$, from the proof in Proposition 1 (Part 1). Evaluating the integral from $t_0 = 0$ to T and eliminating N_T yields:

$$\begin{aligned} \tilde{V}_P &= \frac{L\left(e^{\beta T}-1\right)e^{-x_h\psi_h-x_l\psi_l-\beta T}\psi_h e^{x_l\psi_l}}{\beta\left(\psi_h+\psi_l\right)}\left(\frac{\psi_h\left(e^{\delta T}-1\right)e^{-x_h\psi_h-T\delta}}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{\alpha_h}{1-\alpha_h}} \\ &\quad +\frac{L\left(e^{\beta T}-1\right)e^{-x_h\psi_h-x_l\psi_l-\beta T}\psi_h e^{x_l\psi_l}\psi_l e^{x_h\psi_h}}{\beta\left(\psi_h+\psi_l\right)}\left(\frac{\psi_l\left(e^{\delta T}-1\right)e^{-x_l\psi_l-\delta T}}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{\alpha_l}{1-\alpha_l}} \end{aligned} \tag{39}$$

This is now a two-dimensional unconstrained optimization problem, where

$$\frac{\partial \tilde{V}_P}{\partial x_l} = \frac{L\psi_l^2\left(e^{\beta T}-1\right)e^{-x_l\psi_l-\beta T}}{\beta\left(\alpha_l-1\right)\left(\psi_h+\psi_l\right)}\left(\frac{\psi_l\left(e^{\delta T}-1\right)e^{-x_l\psi_l-\delta T}}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{\alpha_l}{1-\alpha_l}} \tag{40}$$

$$\frac{\partial \tilde{V}_P}{\partial x_h} = \frac{L\psi_h^2\left(e^{\beta T}-1\right)e^{-x_h\psi_h-\beta T}}{\beta\left(\alpha_h-1\right)\left(\psi_h+\psi_l\right)}\left(\frac{\psi_h\left(e^{\delta T}-1\right)e^{-x_h\psi_h-\delta T}}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{\alpha_h}{1-\alpha_h}}. \tag{41}$$

Setting $\partial\tilde{V}_P/\partial x_l = 0$ and $\partial\tilde{V}_P/\partial x_h = 0$ and solving for $\bar{x}_l$ and $\bar{x}_h$, we obtain the steward's solution:

$$\bar{x}_l = \frac{-\left(1-\alpha_l\right)}{\psi_l}\times \tag{42}$$
$$\ln\left(\frac{\left(1-\alpha_l\right)\beta\gamma^{-\frac{\alpha_l}{\alpha_l-1}}\delta^{-\frac{\alpha_l}{\alpha_l-1}}\psi_l^{\frac{1}{\alpha_l-1}-1}\left(\psi_h+\psi_l\right)^{\frac{1}{1-\alpha_l}}\left(e^{\delta T}-1\right)^{\frac{1}{\alpha_l-1}+1}e^{\beta T-\frac{\delta\alpha_l}{\alpha_l-1}T}}{L\left(e^{\beta T}-1\right)}\right)$$

$$\bar{x}_h = \frac{-\left(1-\alpha_h\right)}{\psi_h}\times \tag{43}$$
$$\ln\left(\frac{\left(1-\alpha_h\right)\beta\gamma^{-\frac{\alpha_h}{\alpha_h-1}}\delta^{-\frac{\alpha_h}{\alpha_h-1}}\psi_h^{\frac{1}{\alpha_h-1}-1}\left(\psi_h+\psi_l\right)^{\frac{1}{1-\alpha_h}}\left(e^{\delta T}-1\right)^{\frac{1}{\alpha_h-1}+1}e^{T\left(\beta-\frac{\delta\alpha_h}{\alpha_h-1}\right)}}{L\left(e^{\beta T}-1\right)}\right)$$

Simplification of Equations 42 and 43 yields the solutions given in Proposition 2 (Part 1).

Part 2: Attacking Intensity

The attacker intensities under the fully informed steward are obtained by substituting the optimal expenditures $\bar{x}_l$ and $\bar{x}_h$ into Equations 33 and 34; that is,

$$\bar{\eta}_l = \left(\frac{\psi_l\left(e^{\delta T}-1\right)e^{\delta(-T)-\bar{x}_l\psi_l}}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{1}{1-\alpha_l}} \tag{44}$$

$$\bar{\eta}_h = \left(\frac{\psi_h\left(e^{\delta T}-1\right)e^{\delta(-T)-\bar{x}_h\psi_h}}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{1}{1-\alpha_h}} \tag{45}$$

Setting $i = \{h, l\}$ and $j = \{h, l\}$ for $j \neq i$ yields Equation 9 in Proposition 2 (Part 2). □

The analytic forms of Equations 44 and 45, as functions of the model parameters, are as follows:

$$\bar{\eta}_l = \left(\frac{\psi_l e^{\delta(-T)}\left(e^{\delta T}-1\right)}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{1}{1-\alpha_l}}\times \tag{46}$$
$$\left(-\frac{\beta\left(\alpha_l-1\right)\gamma^{-\frac{\alpha_l}{\alpha_l-1}}\delta^{-\frac{\alpha_l}{\alpha_l-1}}\psi_l^{\frac{1}{\alpha_l-1}-1}\left(\psi_h+\psi_l\right)^{\frac{1}{1-\alpha_l}}\left(e^{\delta T}-1\right)^{\frac{1}{\alpha_l-1}+1}e^{T\left(\beta-\frac{\delta\alpha_l}{\alpha_l-1}\right)}}{L\left(e^{\beta T}-1\right)}\right)$$

$$\bar{\eta}_h = \left(\frac{\psi_h e^{\delta(-T)}\left(e^{\delta T}-1\right)}{\gamma\delta\left(\psi_h+\psi_l\right)}\right)^{\frac{1}{1-\alpha_h}}\times \tag{47}$$
$$\left(-\frac{\beta\left(\alpha_h-1\right)\gamma^{-\frac{\alpha_h}{\alpha_h-1}}\delta^{-\frac{\alpha_h}{\alpha_h-1}}\psi_h^{\frac{1}{\alpha_h-1}-1}\left(\psi_h+\psi_l\right)^{\frac{1}{1-\alpha_h}}\left(e^{\delta T}-1\right)^{\frac{1}{\alpha_h-1}+1}e^{T\left(\beta-\frac{\delta\alpha_h}{\alpha_h-1}\right)}}{L\left(e^{\beta T}-1\right)}\right).$$

A.3 Proof of Proposition 4

The final case we consider in this paper considers the case when a steward can only observe and mandate one of the elements of the investment allocation, x_h. The targets have discretion to signal a value $\hat{L}$, however the steward does not know the true value of L or z.

The attackers signal a value $\tilde{\zeta}$, which we assume is actually $1-z$. Targets, still have to choose their asset allocation, but they can potentially hide it from a potentially costly investment allocation. For tractability, we will assume this is in two steps, a signal of $\hat{L}$ and $\tilde{\zeta}$ and then an adjustment. This is done for tractable exposition, although the simultaneous model also has an analytic solution and provides a similar result, whilst being algebraically more complex.

Part 1: Asset Class h

Let $\tilde{\sigma}_{i\in\{l,h\}}:\mathbb{R}_+\to[0,1]$ and, for the targets, let x_h be exogenous. Targets minimize

$$\tilde{V}_T=\frac{1}{\beta}L\left(e^{\beta T}-1\right)e^{-x_h\psi_h-x_l\psi_l+\beta(-T)}\left(ze^{x_h\psi_h}\eta_l^{\alpha_l}-(z-1)\eta_h^{\alpha_h}e^{x_l\psi_l}\right)+x_h+x_l \tag{48}$$

Setting $\partial\tilde{V}_T/\partial x_l=0$ and $\partial\tilde{V}_T/\partial z=0$ and solving for x_l and z we obtain

$$x_l^{\diamond} = \frac{x_h\psi_h-\ln\left(\eta_h^{\alpha_h}\eta_l^{-\alpha_l}\right)}{\psi_l} \tag{49}$$

$$z^{\diamond} = \frac{\beta e^{\beta T}\eta_l^{-\alpha_l}}{L\psi_l\left(\eta_h^{\alpha_h}\eta_l^{-\alpha_l}e^{\beta T-x_h\psi_h}-\eta_h^{\alpha_h}e^{-x_h\psi_h}\eta_l^{-\alpha_l}\right)} \tag{50}$$

Note that the both the optimal asset allocation $z^{\diamond}$ and the optimal investment $x_l^{\diamond}$ are now functions of x_h and are both subject to an upper bound of $\eta_i^*<e^{\alpha_i^{-1}x_i\psi_i}$.

Part 2: Asset Class l

The steward has received a information on $\hat{L}$ and $\tilde{\zeta}$, which in this derivation we treat as exogenous. However, the optimal initial bid of $\hat{L}$ from the targets to the steward can be obtained by numerical analysis. The steward sets a mandatory investment level of $\bar{x}_h$, from a restricted information set by minimizing

$$\tilde{V}_P=\frac{N_T}{\bar{\beta}}\hat{L}\left(e^{\bar{\beta}T}-1\right)\eta_h^{\diamond\alpha_h}e^{-\bar{\beta}T-x_h\psi_h}+N_Tx_h \tag{51}$$

where

$$\eta_h^{\diamond}=\left(\frac{\tilde{\zeta}\left(e^{\delta T}-1\right)e^{-x_h\psi_h-\delta T}}{\gamma\delta}\right)^{\frac{1}{1-\alpha_h}} \tag{52}$$

solving the single equation and single unknown $\partial V_P/x_h=0$, yields

$$\bar{x}_h=\frac{1}{\psi_h}\ln(\mathcal{A})+\frac{\alpha_h}{\psi_h}\left(T(\bar{\beta}-\delta)-\ln(\mathcal{B})\right)-\frac{\bar{\beta}T}{\psi_h} \tag{53}$$

Note that x_h is now a function of $\hat{L}$, $\tilde{\zeta}$ and the structural parameters δ, γ, $\psi_{i\in\{l,h\}}$, $\alpha_{i\in\{l,h\}}$ and T. Simplification of Equation 53 results in the steward component of Proposition 4 (Part 1). Substitution of $\bar{x}_h$ into Equation 52 provides the functional form of the attacker intensity $\bar{\eta}_h$ of Proposition 4 (Part 1). The solution in terms of the model parameters is as follows:

$$\bar{\eta}_h=\mathcal{B}^{\frac{1}{1-\alpha_h}}e^{\alpha_h\left(\ln(-\mathcal{A})+T(\delta-\bar{\beta})+T(\bar{\beta}-\delta)\right)^{\frac{1}{1-\alpha_h}}} \tag{54}$$

where

$$\mathcal{A} = \frac{\hat{L}\left(1-e^{bT}\right)\psi_h \gamma^{\frac{1}{\alpha_h-1}+1}\delta^{\frac{1}{\alpha_h-1}+1}\tilde{\zeta}^{\frac{\alpha_h}{1-\alpha_h}}\left(e^{\delta T}-1\right)^{\frac{\alpha_h}{1-\alpha_h}}}{\bar{\beta}\left(\alpha_h-1\right)} \tag{55}$$

$$\mathcal{B} = \frac{\bar{\beta}\left(1-\alpha_h\right)\gamma^{\frac{1}{1-\alpha_h}-2}\delta^{\frac{1}{1-\alpha_h}-2}\tilde{\zeta}^{\frac{1}{\alpha_h-1}+2}\left(e^{\delta T}-1\right)^{\frac{1}{\alpha_h-1}+2}}{\hat{L}\left(e^{bT}-1\right)\psi_h} \tag{56}$$

To derive the target allocation and attacker intensity $\eta_l^{\ddagger}$, we now simply need to substitute the functional forms of $\bar{x}_h$ and $\bar{\eta}_h$ into Equations 49 and 50 and simplify functional forms in Proposition 4 (Part 2). For $x_l^{\ddagger}$ $z^{\ddagger}$ and $\eta_l^{\ddagger}$,

$$x_l^{\ddagger} = \frac{1}{\psi_l}\alpha_h(T(\bar{\beta}-\delta)-\ln(\mathcal{A}))+\ln(\mathcal{A})-\bar{\beta}T \tag{57}$$
$$-\frac{1}{\psi_l}\ln\left(\left(\frac{\beta\left(e^{\delta T}-1\right)e^{T(\beta-\delta)}}{\gamma\delta L\psi_l\left(e^{\beta T}-1\right)}\right)^{-\alpha_l}\left(e^{\alpha_h(\ln(\mathcal{A})-\bar{\beta}T+\delta T)+T(\bar{\beta}-\delta)}\right)^{\frac{\alpha_h}{1-\alpha_h}}\right)$$

$$z^{\ddagger} = \frac{\beta\mathcal{A}}{L\psi_l\left(e^{\beta T}-1\right)}e^{T(\beta-\bar{\beta})-\alpha_h(\ln(\mathcal{A})+T(\delta-\bar{\beta}))}\times \tag{58}$$
$$\left(\mathcal{B}e^{\alpha_h(\ln(\mathcal{A})+T(\delta-\bar{\beta}))+T(\bar{\beta}-\delta)}\right)^{\frac{-\alpha_h}{1-\alpha_h}}$$

$$\eta_l^{\ddagger} = \frac{\beta\left(e^{\delta T}-1\right)e^{T(\beta-\delta)}}{\gamma\delta L\psi_l\left(e^{\beta T}-1\right)} \tag{59}$$

which simplify to the equations in Proposition 4 (Part 2). □