Information Security Trade-offs and Optimal Patching Policies

Christos Ioannidis[1], David Pym[2⋆], and Julian Williams[3]

[1] University of Bath
Department of Economics
Bath BA2 7AY
England, U.K.
c.ioannidis@bath.ac.uk
[2] University of Aberdeen
School of Natural and Computing Sciences
King's College
Aberdeen AB24 3UE
Scotland, U.K.
d.j.pym@abdn.ac.uk
[3] School of Business
University of Aberdeen
King's College
Aberdeen AB24 3QY
Scotland, U.K.
julian.williams@abdn.ac.uk

Abstract. We develop and simulate a basic mathematical model of the costly deployment of software patches in the presence of trade-offs between confidentiality and availability. The model incorporates representations of the key aspects of the system architecture, the managers' preferences, and the stochastic nature of the threat environment. Using the model, we compute the optimal frequencies for regular and irregular patching, for both networks and clients, for two example types of organizations, military and financial. Such examples are characterized by their constellations of parameters. Military organizations, being relatively less cost-sensitive, tend to apply network patches upon their arrival. The relatively high cost of applying irregular client patches leads both types of organization to avoid deployment upon arrival.

1 Introduction

Software for computer networks, systems, and applications is typically subject to information security flaws, which, if exploited, may lead to substantial losses for the host organization. As vulnerabilities appear, software vendors periodically release patches in response. For large organizations, with tens or even hundreds of thousands of network devices, the deployment of patches is a costly exercise, impacting significantly on system availability, with consequences for properties of business processes, for credibility, and revenue. Failure to deploy a patch, however, risks exposing the host organization to exploitations of vulnerabilities.

The host organization's information security management team must make a judgement regarding the appropriate timing of the deployment of patches, in the light of the organization's policies. As in other areas of information security operations, decisions to deploy patches involve trade-offs between protecting the confidentiality of the system and maintaining its availability.

In recent years, there has been a good deal of research in the economics of information security. For example, Anderson et al. [1, 3, 2] have presented wide-ranging discussions of the issues, whilst Gordon and Loeb [21, 22] have employed a microeconomic analysis of the costs and benefits of defences against given vulnerabilities.

[⋆] This work has also been supported by the Cloud and Security Lab, HP Labs, Bristol.

Recent work by the present authors [25] has considered how to apply ideas and methods from utility theory and dynamic optimization to investment in information security. More specifically, by way of an illustration of the methodology, we have presented a dynamic model of trade-offs between confidentiality, availability, and investment in information security.[4] Our analysis has been motivated by those situations — including a detailed example in Beautement et al. [8], based on the use of USB memory sticks, as well as the work of Beres et al. [9, 10] — in which the corruption of data (i.e., integrity) is a relatively minor issue. Here we intend confidentiality to refer to the system's state of protection against breaches of confidentiality, rather than the state of exposure of any particular data item. Similarly, we intend availability to refer to the system's readiness to supply its intended service. Our use of this example does not exclude the applicability of the methods we employ to situations in which integrity plays a major role. Such situations will be considered elsewhere. Moreover, in specific practical applications, it will typically be necessary to build richer models that incorporate more domain-specific details, such as the criticality of various information system components to business processes.

In this paper, we develop a model that is based on the confidentiality–availability trade-off model presented in [25], in which patch arrivals are interpreted as shocks to confidentiality and availability. We use this model to derive patching strategies in (large) organizations. We consider in detail the optimal timing of both client and network patching. As in [25], the purpose of this paper is to illustrate the application of modelling and reasoning methods from utility theory and macroeconomics to questions in information security management that involve trade-offs between attributes of interest.

The remainder of the paper is organized as follows: in Section 2, we discuss related work; in Section 3, we develop our basic mathematical set-up, which draws upon methods from utility theory and dynamic optimization as deployed in economic and financial modelling, explaining how we model optimal responses in the presence of shocks to confidentiality and availability; in Section 4, we study an example of a model of the kind described in Section 3, to which to introduce a cost structure for implementing patches, and in which shocks to confidentiality and availability are given by the arrival (according to a Poisson process) of patches whose severity (or intensity) is drawn from a log-normal distribution; in Section 5, we describe the appropriate instance of the space of parameters employed in the model, and present and comment upon the results of our numerical simulations; finally, in Section 6, we explain our findings and their consequences for patching policies.

2 Related Work

Beres et al. [9] provide a process model (written in the Demos 2K modelling language [16], now superseded for our purposes by the Gnosis modelling language [13, 19]) of vulnerability management policies in a large organization, and explore the effectiveness of both standard (or regular) patch-management and emergency (or irregular) escalation-based policies. In designing their model, which is concerned with patching clients, they examine the decision making process followed by the security operations managers of several large organizations, together with the different mitigation and patching measures that might be selected. They also identify external threat environment events that influence the type of mitigations that are deployed and the time at which they are deployed. They focus on examining the 'risk exposure window', defined as the time from public vulnerability disclosure to when an organization believes the risk is mitigated, as a measure of the effectiveness of these processes. By designing a model of these processes and running stochastic simulations, they examine the effectiveness of security operations processes and protection mechanisms based on external environment events.

In [23], it is postulated that both attackers and defenders behave strategically, whilst Beres et al. [9] seeks, treating attackers exogenously, to enable the decision-makers in IT security to predict the outcome of investment decisions or changes in policy in advance of putting them into effect.

[4] Modelling multiple trade-offs can be accommodated within the same methodology.

Their results show the impact of increasing the effectiveness of early mitigations and of speeding up patch deployment on reducing the risk exposure window.

The importance of timely patching in networks in the presence of externalities has been addressed by August and Tunca [6], in which they develop a set of incentive structures for users to implement effective patch management when their actions impact upon the welfare of other users. They show that software vendors can offer rewards to encourage timely patching when vulnerabilities occur in both proprietary software and freeware and, given the differential costs of patching to users, conclude ([6], p. 1718) that 'a "one-size-fits-all" approach is unlikely to be an immediate remedy'.

The timing of vulnerability disclosures by vendors is modelled formally by Arora, Telang, and Xu (2008) [5], where it is shown that, with no regulation, the vendor releases a patch less frequently than is socially optimal.

The relationship between the release of patches by vendors and their implementation has been studied recently by Cavusoglu, Cavusoglu, and Zhang [12]. They classify patching cycles into time-driven and event-driven. They show that social loss is minimized when vendor releases are synchronized with the time-driven cycles of the system operator. Their analysis is done in the context of single vendor and a single system operator. When such synchronization cannot be achieved because it is costly, the imposition of liability of the vendor for delayed release cannot achieve socially optimal disclosures.

When system operators employ a variety of applications, patch arrivals to the system operator will appear as random events, without apparent periodicity. In this paper, we capture patch arrivals as a Poisson process, and we decompose patching implementation into time-driven and event-driven incidents.

The 'Vulnerability Timeline', reproduced from Beres et al. [9]in Figure 1, is a reference point for many studies of patching policies.

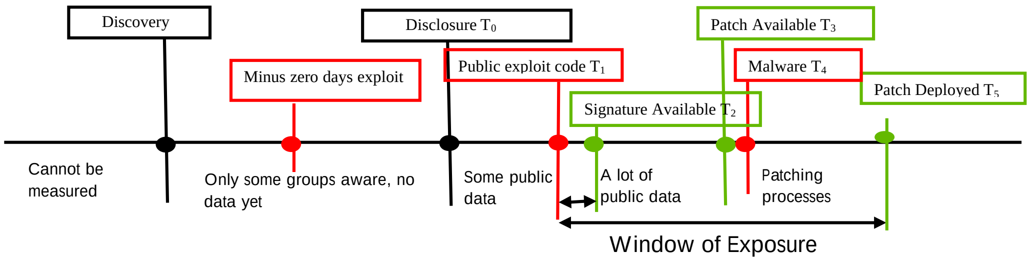

Fig. 1. Vulnerability Timeline: the sequencing of events in this timeline is not fixed; the aim is to illustrate the various stages in the vulnerability life cycle (Beres et al. [9])

The timeline provides a detailed description of the sequence of events from the discovery of a vulnerability to the deployment of a patch. Similar accounts of such a timeline have been given by other authors; for example, Arbaugh et al. [4] and Schneier [34]. Of particular interest is Frei et al. [17], which illustrates the distributions and frequencies of vulnerabilities, using data from several large databases. The vulnerability arrival timeline given by Frei et al. is augmented in Beres et al. [9]. Arora, Telang, and Xu [5] calculate the socially optimal time interval between discovery and disclosure, T_0. August and Tunca [6] calculate, in the presence of externalities, the optimal period Patch Available to Patch Deployed, T_3 to T_5, when vendors offer incentives to the system operator. Cavusoglu, Cavusoglu, and Zhang [12], calculate the socially optimal window of exposure an decompose the patching process into time- and event-driven incidents.

Beattie et al. [7] explore the factors affecting the best time to apply patches so that organizations minimize disruptions caused by defective patches. Their results indicate that patching during the period of 10 to 30 days after first patch release date is the optimal time for minimizing the disruption caused by defective patches. The adoption of a real options methodology for determining choices of the appropriate technology in the presence of multiple sources of uncertainty and market entry has been addressed by [11, 32]. In a similar vein, Gordon et al. [20] offer a framework around which decisions to delay the implementation of patches are integrated into a financial model that exploits deferment.

From the arguments discussed above, it is apparent that the timing of patch deployment matters because their deployment subjects organizations to serious costs and mis-timing may exacerbate the impact of costs.

In our economic model, developed in Sections 3.2 and 4, we are concerned with just one segment of the timeline: Patch Available to Patch Deployed. We develop a model of regular (i.e., time-driven) and irregular (i.e., event-driven) patching in the presence of stochastic patch arrivals (at T_3) that differ in frequency and intensity and impact upon confidentiality and availability. We take account of the system manager's preferred trade-off between protecting confidentiality and availability.

3 Introduction to the Modelling Method

Organizations deploy systems technologies in order to achieve their business objectives. Typically, it will be necessary for an organization to invest in deploying information security policies, processes, and technologies in order to protect the confidentiality, C, integrity, I, and availability, A, of its business processes.

Defences deployed against each of C, I, and A may compromise the others. For example, the process of deploying a network patch, in order to protect the confidentiality of the organization's system, may compromise the availability of network resources, such as filestores and databases to client devices. Moreover, the deployment of such defences is costly. It follows that security managers with limited budgets must determine an allocation of investments, K, to defences for C, I, and A that is appropriate to their priorities. Different types of organization will have different priorities and examples exist of trade-offs between all of C, I, A, and K.

The purpose of this paper is to illustrate the methodology of modelling and reasoning about the dynamics of the trade-offs between the quantities of interest to the managers using methods that are commonly deployed in macroeconomics. To that end, the situation we study, though informed by reflective interaction with operational security managers in several large organizations, is somewhat simplified for illustrative purposes and brevity. For example, for the purposes of this paper, we consider just trade-offs between investments to protect confidentiality and availability. This simplification should in no way be taken as an indication that integrity is not a concern that our methodology is able to address. Moreover, practical applications of our approach will typically require richer models that incorporate a good detail of domain-specific detail.

A basic modelling framework for addressing trade-offs of this kind has been given by Ioannidis, Pym, and Williams [25].

3.1 A Dynamic Model of Trade-offs

Following Ioannidis, Pym, and Williams [25], we assume — for simplicity and illustrative purposes, implying no restriction of the method — that our storage and processing technologies do not corrupt data and study the trade-off between confidentiality and availability. This simple situation is intuitively appealing: disks, DVDs, and memory sticks are rarely corrupted, at least in contexts similar to that studied in Beautement et al. [8]. Client and network patches, as modelled by Beres et al. [9], only rarely corrupt users' data. It should be emphasized, however, that there are many situations in which the suppression of integrity is not valid. In such cases, models of the kind described herein must be enriched with components handling integrity.

Technically, within the framework described by Giannoni and Woodford [18], we consider a utility maximizing decision-maker with convex preferences, $U(t)$ over a fixed time horizon, $[t_0, T]$.

The general linear stabilization policy problem can be expressed as a solution to the following control problem, in which the economic interaction structure of the state variables is given in terms of a linear system of the form

$$\varGamma \begin{bmatrix} Z_{t+1} \\ E_t z_{t+1} \end{bmatrix} = \varPsi_1 \begin{bmatrix} Z_t \\ z_t \end{bmatrix} + \varPsi_2 r_t + \varPsi_3 u_t \tag{1}$$

where z_t denotes a vector of endogenous variables (e.g., inflation, unemployment, consumption) and the vector of pre-determined variables is given by Z_t (lagged values of the dependent and current and lagged values of the independent variables), $\varGamma$, $\varPsi_1$, $\varPsi_2$, and $\varPsi_3$ represent the structure of the system as determined by the behaviour of the agents in the economy and the resource constraint, and E_t denotes the conditional expectations operator. The instrument available to the authorities is given by r_t and the system is disturbed from its original equilibrium position due to the existence of shocks u_t. The objective of the policy is to minimize the quadratic objective function in terms of squared deviations of the variables of interest $\varTheta$ from some a-priori specified target values $\varTheta^*$ by choosing the appropriate value of r given the structure of the system, the loss function,

$$\varLambda = E_t \left\{ \sum_{t=0}^{T} \frac{\delta^{-t}}{2} \left(\varTheta - \varTheta^*\right)^{\mathsf{T}} \varOmega \left(\varTheta - \varTheta^*\right) \right\} \tag{2}$$

where the vector of variables denoted by $\varTheta$ includes values of both z_t and r_t. The matrix $\varOmega$ denotes the variance covariance matrix of the system and δ is the authorities' discount factor. The conditional (on all available information) expectations operator is E_t.

The equilibrium characterization of the system consists of a set of time invariant equations:

$$z_t = \beta_0 + \beta_1 \bar{Z}_t + \beta_2 \bar{u}_t \tag{3}$$

where $\bar{\cdot}$ indicates that the structure of the relevant vectors can differ from the one denoted in Equation 1, and the parameters β_0, β_1, and β_2 represent the optimal responses from Equation 2 and are derived as combinations of the structural parameters in Equation 1.

The imposition of rational expectations requires that the model's predictions of the endogenous variables are equal to the agents' forecasts.

Nobay and Peel [31] accommodate the absence of symmetric loss in the presence of deviations by employing, in $\varLambda$, the Linex function whose asymmetry depends upon the choice of the parameter a:

$$g(u_t) = \left\{\exp(au_t) - au_t - 1\right\} / a^2. \tag{4}$$

In our case, we restrict our analysis to quadratic loss functions but we allow for unequal weights to be applied to its different arguments.

The analysis given by Giannoni and Woodford [18], together with refinements of the kind suggested by the work of Nobay and Peel [31], provides a very general framework for capturing the dynamics of investments and trade-offs in information security within which the choices of security and investment properties to be modelled, along with associated organizational preferences, can be captured.

3.2 A Specific Dynamic Model of Trade-offs

In our context, the decision-maker wishes to minimize the following loss function, defined in terms of the time t levels of confidentiality, $C(t)$, availability $A(t)$, and investment $K(t)$, and their respective targets $(\bar{C}, \bar{A}, \bar{K})$ and a control vector $\mathbf{x}$:

$$\mathfrak{H}\left(C(t), A(t), K(t;\mathbf{x})\right) = \omega_1 \left(C\left(t\right) - \bar{C}\right)^2 + \omega_2 \left(A\left(t\right) - \bar{A}\right)^2 + \omega_3 \left(K\left(t;\mathbf{x}\right) - \bar{K}\right)^2 \tag{5}$$

The system's solution will be of the form

$$\Im\left(\mathbf{x}^*\right) \triangleq \min_{\mathbf{x}} \int_t^T \exp(-\delta t) \mathfrak{H}\left(C(t), A(t), K(t;\mathbf{x})\right)\, dt \tag{6}$$

where δ is a discount factor operating over the investment horizon t, T and $\mathbf{x}^*$ is the optimal policy response. The weights in the loss function (5) characterize the type, or preferences, of the organization; for example, military and deep-state organizations might put a great deal of weight on C compared to A, whilst a retailer or welfare distributor might place greater value on A compared to C. A bank, when considering the potential impact of network patching on its client devices' access to network services must make a more delicate judgement: its customers expect their data to be adequately protected, but also expect to access their accounts via ATMs at all times.

Finally, the weight on $(K(t;\mathbf{x}) - \bar{K})^2$ reflects the system's loss when managers are forced to compromise 'budgets'. In practice, we expect managers' preferences for investment to be asymmetric; that is, that they will be more averse to overspending.

The effectiveness of the managers' investment responses is constrained by the time evolution of confidentiality and availability. There are described by a system of equations (7) which express their dynamics in terms of parameters that characterize the behaviour of the system.

$$\begin{aligned} C(t) &= -\alpha(P)\left(\int_{t_0}^{t} \dot{A}\, dt \left(\beta \int_{t_0}^{t'} \dot{K}\, dt'\right)^{-1}\right) + C_0 \\ A(t') &= \gamma\left(\int_{t_0}^{t'} \dot{R}\, dt'\right) + \delta\left(\int_{t_0}^{t'} \dot{K}\, dt'\right) - \epsilon\left(\int_{t_0}^{t'} \dot{C}\, dt'\right) - \zeta(Q) \end{aligned} \tag{7}$$

where $t' < t$, denoting that current shocks, $\alpha(P)$, to confidentiality subsequently affect availability. We explain the set-up below.

Shocks (reductions) to availability are denoted by the stochastic process Q, and shocks (breaches) in confidentiality are denoted by a second the stochastic process P. In both cases, the shocks enter the system linearly, and the process is defined as being non-decreasing. Their impact is measured by α and ζ, respectively. The system's attack surface is modelled by the availability

$$\int_{t_0}^{t} \dot{A}\, dt, \tag{8}$$

and amplifies the influence of breaches of confidentiality, whilst increases in the capital stock of information security[5],

$$\frac{1}{\int_{t_0}^{t} \dot{K}\, dt}, \tag{9}$$

mitigates against the severity of such shocks. The effectiveness of this mitigation is measured by the value of the positive parameter β.

The availability of the system depends positively of the system's inter-connectedness,[6] $\int_{t_0}^{t'} \dot{R}\, dt$ and the capital stock of information security. Increases in confidentiality are expected to exert a

[5] For simplicity of exposition of the initial properties of the model, we do not allow for depreciation in the capital stock of information security.

[6] The system response to deviations in confidentiality is given by (where x is an element of the control vector $\mathbf{x}$)

$$\dot{R} = x\,(C - \bar{C})$$

where R is defined as

$$R = \frac{1}{1-\xi} \tag{10}$$

for $\xi \in [0, 1)$. R is thus considered to be a (very crude) measure of the interconnectedness of the system.

This notion of *interconnectedness* should not be confused with *interdependence*. In the theory of distributed systems (see, for example, [14]), system availability is promoted by reducing the *interdependence* of the distributed components: system availability is increased by increasing the extent to which the distributed components are able to operate independently of one another. In contrast, this negative influence on the system's availability. The positive parameters γ, δ, ϵ, and ζ measure the impact of these factors on the system's availability.

Capital stock in information security is determined by 'unexpected' fluctuations in availability and by the arrival of software and system upgrades, such as security patches. The time dynamics of this is expressed in Equation 11 (K_0 is an initial value):

$$K = -\eta A + \mathcal{P}(x(t)) + K_0 \tag{11}$$

Here $x(t)$ is a component of the vector $\mathbf{x}$ in Equation 5. In the presence of patches, with associated deployment costs given by $\mathcal{P}(x)$, IT managers must decide on the optimal timing of such deployment to minimize costs. Note that, as $t' \to \infty$, the system stabilizes.

As formulated here, the model shocks both confidentiality and availability. The model considered by the present authors in [25] shocks only confidentiality. A richer model might also shock investment. Such a model would need to be formulated with an additional control instrument, so that there would be an instrument corresponding to each shocked dimension. The general framework within which models such as these are formulated is discussed briefly in Section 3.1.

IT managers will respond to decreases in availability by increasing investment in information security (11). The managers' response is measured by the parameter, η. In the presence of deviations of confidentiality from its target, IT managers respond by manipulating the system's inter-connectedness. Such response is calculated optimally given the architecture of the system, as captured by the parameters, the managers' preferences, and behaviour as captured by w_1, w_2, w_3, and η, given the choice of targets $\bar{C}$, $\bar{A}$, and $\bar{K}$.

Knowledge of the existence of patches creates an expectation of a loss of both C and A. This negative expectation triggers the patching response. Managers act upon expectations of patch arrivals, $\mathcal{P}(x(t))$, and plan for their implementation.

4 A Patching Model

In this section, following the methodology of Section 3.1, for the system given in Section 3.2, we describe the threat environment represented by the arrival of patches, in the presence of fixed and variable costs, and compute the optimal responses to approximate Equation 6. In this context, the vector control variate is the number of regular and irregular patches.

Patches are interpreted as shocks to the confidentiality and availability of the system. That is, the arrival of a patch (which is intended to be applied to the system) signals the existence of a vulnerability in the system's confidentiality or availability, because it admits the possibility of a breach.

Patches are considered to be a non-decreasing Cox-type point process [15], y_t, with Poisson-type arrivals, $\Phi\left(t\right)|\theta^{\Phi}$, and bivariate log-normal intensities $\Psi\left(t\right)|\theta^{\Psi}$.

The Poisson process is a standard way of modelling independent arrivals and the log-normal is a single-tailed distribution which captures a random variable that arises as a product positive independent random increments. Our choices represent a simplification of reality, but we believe it is a reasonable one. The vulnerabilities signalled by each patch have a an impact on confidentiality or availability described by Equation 12.

$$\begin{bmatrix} y_{1,t+\Delta t} - y_{1,t} \\ y_{2,t+\Delta t} - y_{2,t} \end{bmatrix} = \begin{bmatrix} \pi_{1,1} \; \pi_{1,2} \\ \pi_{2,1} \; \pi_{2,2} \end{bmatrix} \begin{bmatrix} \psi_{1,t} \\ \psi_{2,t} \end{bmatrix} \phi_t \tag{12}$$

The parameter vector associated with the system consists of μ_1, μ_2, σ_1^2, σ_2^2, and the correlation coefficient ρ_{12}. The matrix Π, with components, $\pi_{i,j}$, linearly decomposes the signal of the patch

(albeit crude) measure of interconnectedness simply captures the extent to which the components of the system are able to communicate with one another, thereby permitting data/information located at one component available to other components, but also permitting the propagation of malware.

For $\xi = 0$, $R = 1$, so that the system amounts to a single isolated device. As $\xi \to 1$, $R \to \infty$, and the system amounts to a highly interconnected structure. We include the definition of R here only for completeness of presentation of the underlying model: we make no further use of it in this paper.

arrival to consequences for the availability and confidentiality of the system. The time t states of confidentiality C_t and availability A_t — the discrete-time equivalents of the continuous-time measures $C(t)$ and $A(t)$ — are based on a fully patched system, $[C_0, A_0]^{\mathrm{T}}$. In the presence of patches, confidentiality and availability evolve according to Equation 13:

$$\begin{bmatrix} C_t \\ A_t \end{bmatrix} = \begin{bmatrix} C_0 \\ A_0 \end{bmatrix} - f \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} \tag{13}$$

The function f is a rescaling function to ensure that the patch information, contained in the evaluation of the intensities (ψ_1, ψ_2) at t, matches the appropriate scale of confidentiality and availability.

If the mapping of the vulnerabilities to confidentiality and availability is fully determined, then such stochastic decomposition is not required, thus significantly reducing the number of parameters to be estimated.

4.1 Co-optimization and Decision Making

We consider a patching optimization problem, whereby the decision-maker has two instruments, a long instrument, $x_1(t\,|t_0, E(y_t)\,)$, which is a regular patching cycle taken at evenly spaced points in the time interval $[t_0, T]$, and set prior to t_0, and a short instrument, $x_2\,(t)$, the decision to patch early, taken within the interval $t \in [t_0, T]$.

At time t, the non-decreasing sequence of confidentiality and availability is as follows (notation: $|\_$ denotes dependency and $\|$ is read as ‘or’):

$$t'\,|E(y_t) \,\in \left[t_0, t_0 + \tfrac{T}{x_1}, t_0 + \tfrac{2T}{x_1}, \ldots, T\right] \tag{14}$$

$$C_{t+\varDelta t} = \begin{cases} C_t + \varDelta C_t\,|y_t & \text{iff } (t \neq t')\,\|(x_2 = 0) \\ \bar{C} & \text{if } \; t = t' \\ \bar{C} & \text{if } \; x_2 > 0 \end{cases} \tag{15}$$

$$A_{t+\varDelta t} = \begin{cases} A_t + \varDelta A_t\,|y_t & \text{iff } (t \neq t')\,\|(x_2 = 0) \\ \bar{A} & \text{if } \; t = t' \\ \bar{A} & \text{if } \; x_2 > 0 \end{cases} \tag{16}$$

In the first cases of Equations 15 and 16, the system is vulnerable because patches have been installed neither by the utilization of the long nor the short instruments. All other cases denote that the system has been patched.

The long instrument, x_1, is an non-negative integer defining the number of regular patch implementations during the planning period $[t_0, T]$. This process is considered to be the *regular patching cycle* and the associated required increases in information security capital stock are given as

$$\mathcal{P}_1(x_1(t)) = \nu \exp x_1 \tag{17}$$

where ν is the cost of implementing each patch.[7]

Implementation of the short instrument, $x_2\,(t)$, has two expenditure components: a fixed component and an additional variable reflecting the extra expenditure requirements for patching either side of the regular cycle. If $x_2(t) = 0$, then no additional expenditure is required; that is, $\mathcal{P}_2(x_2(t)) = 0$; otherwise, a convenient representation is given by the following equation:

$$\mathcal{P}_2(x_2(t)) = \upsilon + \alpha'\,(t'' - x_2)^2 + \beta'\,(x_2 - t')^2 \tag{18}$$

where t'' is the time of the next regular patch, t' is the timing of the previous regular patch, and α' and β' are structural parameters. In the case $\alpha' = \beta' = 0$, there is no additional penalty for timing the patch outside of the regular cycle. Patching under the short instrument is considered

[7] For $x_1 = 0$, we define $\mathcal{P}_1(x_1) = 0$.

to be patching outside the regular cycle, constituting the *irregular cycle*: practitioners often refer to it as 'out-of-cycle' patching.

$\mathcal{P}_1$ and $\mathcal{P}_2$ are the components of $\mathcal{P}$ that enter Equation 11 and increasing their size might lead to deviations from the target $\bar{K}$.

4.2 Analysis of the Model

Our model offers a simple representation of patch arrivals and jump intensities, as the decision process is a continuous path decision problem (akin to the exercise choice of an American option). We assume that the jumps are drawn from a log-normal distribution, as described in Equation 12, and they are decomposed according to their impact on confidentiality and availability, as described in Equation 13.

Clients face frequent, and low impact threats, in contrast to networks that encounter high impact, low frequency threats. In both environments, implementing patches for both clients and networks incurs fixed and variable costs. The fixed costs of irregular patching are described in Equation 18 and the variable costs of regular patching are described in Equation 17. The fixed costs are higher for networks. Similar models, with costly patching cycles, have been developed in terms of real options analysis; see, for example, the discussion by Gordon, Loeb, and Lucyshyn [15]. Within their modelling framework the threat environment characteristics are maintained, reducing the decision-making trade-off to a simple minimization of costs associated with regular and event-driven patching cycles. An innovative feature of our model is that IT managers' preference weightings, as expressed in Equation 5, reflecting the natures of their organizations, are important in determining patching policy.

When vulnerabilities are disclosed and patches arrive, IT managers, in deciding the implementation of patching operations, are taking into account the operational status of the system, as expressed in Equation 7, the differential costs of regular (fixed time interval or fixed frequency, Equations 17 and 18) and irregular (ad hoc time interval or irregular frequency) patching, and the impact of the threats on the system's characteristics (Equation 13). The managers' responses are captured by Equation 11, that incorporates the state of the system, in the presence of threats, and the cost of mitigations.

By choosing appropriate parameter constellations, we present numerical simulations based on a two organizational types (military and financial) and two particular cases of the patching problem (client and network). These examples are intended to demonstrate how the model may be used to guide policy, and are based on observations and interactions with practitioners.

We proceed, in Section 5, by providing a detailed account of the vectors of parameters and the calculation of optimal patching frequencies for the given cases.

We have chosen to solve the model by simulation for the following reasons: the managers' reaction, as represented by Equation 11, is path-dependent (and so does not have closed-form solutions) as managers react to the continuous arrival of shocks; also, Equations 17 and 18 admit non-convex forms.

5 Numerical Simulations

For any given patch arrival path, and parameter constellation, we begin by searching for the optimal number of regular patching cycles. Conditional upon such finding, we compute the number irregular cycles. We repeat this process B (e.g., 1000, 10 000,...) times, indexed by i, for each bootstrap iteration there n patches, indexed by j, drawn from the process given in Equation 12 over T time intervals, indexed by t. We select the combination of x_1 and x_2 that provides the minimum expected loss as follows:

$$\min_{x} \tfrac{1}{B} \sum_{j=1}^{B} \sum_{i=1}^{n} \sum_{t=1}^{T} \tfrac{1}{2} \delta^{-t} \mathfrak{H}_{i,j,t} \qquad (19)$$

In Ioannidis, Pym, and Williams [25], we explain how different constellations of the parameters in the loss function and state equations, (5), (7), and (7), identify generic types of organizations. In this paper, we adopt the constellations corresponding to the financial firm and the military establishment, presented by Ioannidis, Pym, and Williams [25].

The additional parameters required in this paper, along with the system properties and the decision-maker's preferences, are given in Table 1. Our choices of parameters are informed by extensive, reflective discussions with experienced practitioners. Certain parameters are easily identified, such as the rate of patch arrival, however the impact of vulnerabilities is more difficult to identify. Part of ongoing research is to infer the impact of vulnerabilities from observing practitioner reactions to certain types of vulnerability and patching events. We emphasize that the extensive empirical studies (with all their well-documented attendant difficulties [24, 26–30]) that would be necessary to obtain a more rigorous elicitation of preferences are beyond the scope of the present paper. For now, we require a plausible way to proceed and test the feasibility and value of the overall framing and modelling methods in the decision process, deferring consideration of more rigorous reference-elicitation methodologies to another occasion.

Category	Parameters	Description
Threat Environment	θ^{Φ}	Poisson process hyper-parameter
	$\theta^{\Psi}_1 = \mu_1$	Mean of log-normal factor process 1
	$\theta^{\Psi}_2 = \mu_2$	Mean of log-normal factor process 2
	$\theta^{\Psi}_3 = \sigma^2_1$	Variance of log-normal factor process 1
	$\theta^{\Psi}_4 = \sigma^2_2$	Variance of log-normal factor process 2
	$\theta^{\Psi}_5 = \rho_{1,2}$	Correlation of factor process 1 and 2
System Properties	ν	Cost per regular patching cycle node
	υ	Cost of implementing an out-of-cycle patch
	α'	Early patch penalty parameter
	β'	Early patch penalty parameter
	$\pi_{1,1}, \ldots, \pi_{2,2}$	Patch intensity to confidentiality and availability vulnerabilities
Decision-maker's Preferences	ω_1	Loss function parameter: confidentiality
	ω_2	Loss function parameter: availability
	ω_3	Loss function parameter: investment
	$\bar{C}$	Target level of confidentiality
	$\bar{A}$	Target level of availability
	$\bar{K}$	Target level of investment

Table 1. Parameter Categorization with Definitions

The parameters in the various categories given in Table 1 are set out as follows:

- Threat Environment: The parameters can be estimated form observing histories of patch arrivals and intensities;
- System Properties: These are intrinsic to the configuration of the system;
- Decision-maker's Preferences: These are determined by the managers' of the system to be aligned with the relevant business processes.

In Table 5, we give numerical values of the parameters for both the financial and military organizations. For each type of organization, we assign values for both network and client patching environments. For implementation, the model is discretized using an Euler scheme[33], similar to that employed in [25]. We compute by simulation the optimal patching frequencies (both regular and irregular) for both the client and network patching cases. We characterize the differential threat environment for client and network as suggested in Table 2.

Client patching is characterized by a relatively high value of parameter θ^{Φ} (i.e., high-frequency arrivals) and relatively low values of θ^{Ψ}_1 (i.e., low impact events). In addition, the differential of

Table 2. Suggested Parameter Hierarchies

$\theta^{\Phi}(\text{client}) \gg \theta^{\Phi}(\text{network})$	$\sigma_2(\text{network}) \gg \sigma_2(\text{client})$
$\mu_1(\text{network}) \gg \mu_1(\text{client})$	$\rho_{1,2}(\text{client}) = \rho_{1,2}(\text{network}) = 0$
$\mu_2(\text{network}) \gg \mu_2(\text{client})$	$(\nu - \upsilon)(\text{network}) \gg (\nu - \upsilon)(\text{client})$
$\sigma_1(\text{network}) \gg \sigma_1(\text{client})$	

the fixed costs of regular and irregular patching is negligible. Networking patching, however, is characterized by substantial cost differentials, low frequency, and high impact.

To simplify the model, we adopt a simple decomposition of the patching signal into availability and confidentiality as follows: in both cases $\pi_{1,1} = \pi_{2,2} = 1$ and $\pi_{1,2} = \pi_{2,1} = 0$. Specifically, the impact on a system of an unpatched vulnerability results in a cost directly proportional to the jump sizes of C and A.

The model is simulated over a period of 365 days, with patches arriving daily. Figures 2, 3, and 4 depict the sample paths for the doubly stochastic process for network and client patches, together with their decompositions into the confidentiality and availability factors. Over the period of the simulation, network patching arrivals have low periodicity, with the decomposition of the factors being essentially symmetric. In contrast, client patching arrivals are approximately $3\frac{1}{2}$ times as frequent. The average intensity of client patches is less than $\frac{1}{3}$ that of network patches.

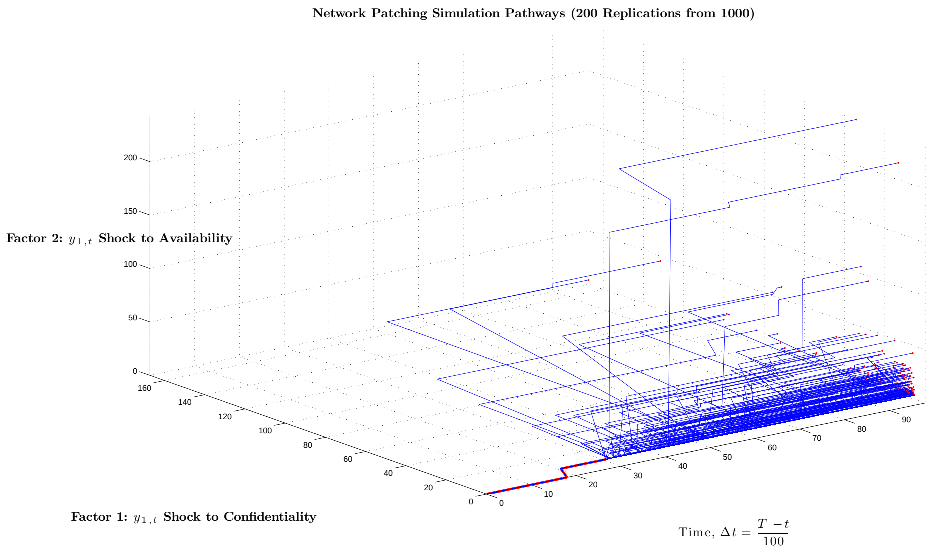

Fig. 2. Patch Arrivals (Network)

5.1 Simulation Results for Different Organizations

For chosen sets of parameter values, we examine the model's response to cost of the irregular patching using, as the appropriate metric, the υ/ν ratio. The adoption of this metric captures the importance of cost in determining the optimal frequency of patching for a given threat environment.

The subsequent analysis computes the change (here decreasing) in the expected numbers of out-of-cycle patches deployed against the changing cost ratio of in- and out-of-cycle patching, whilst holding the absolute cost of patching in-cycle constant.

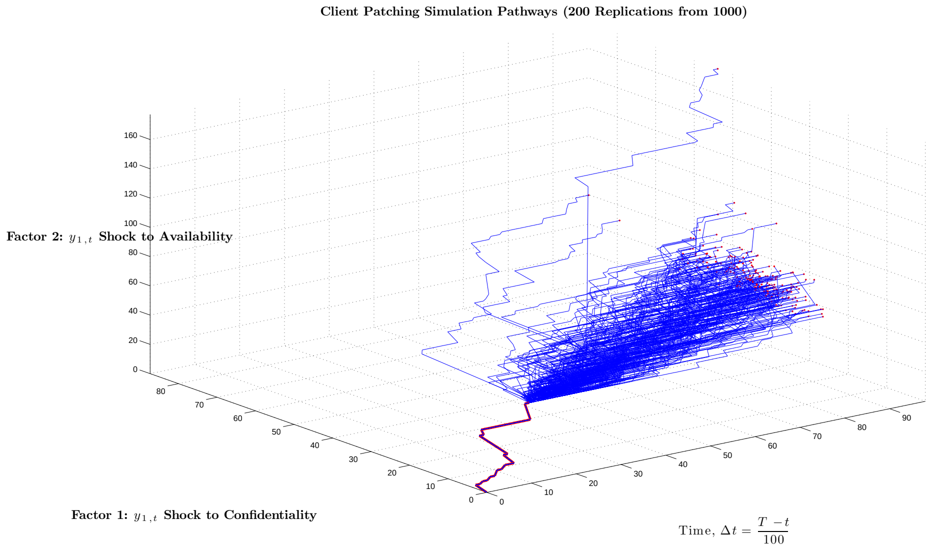

Fig. 3. Patch Arrivals (Client)

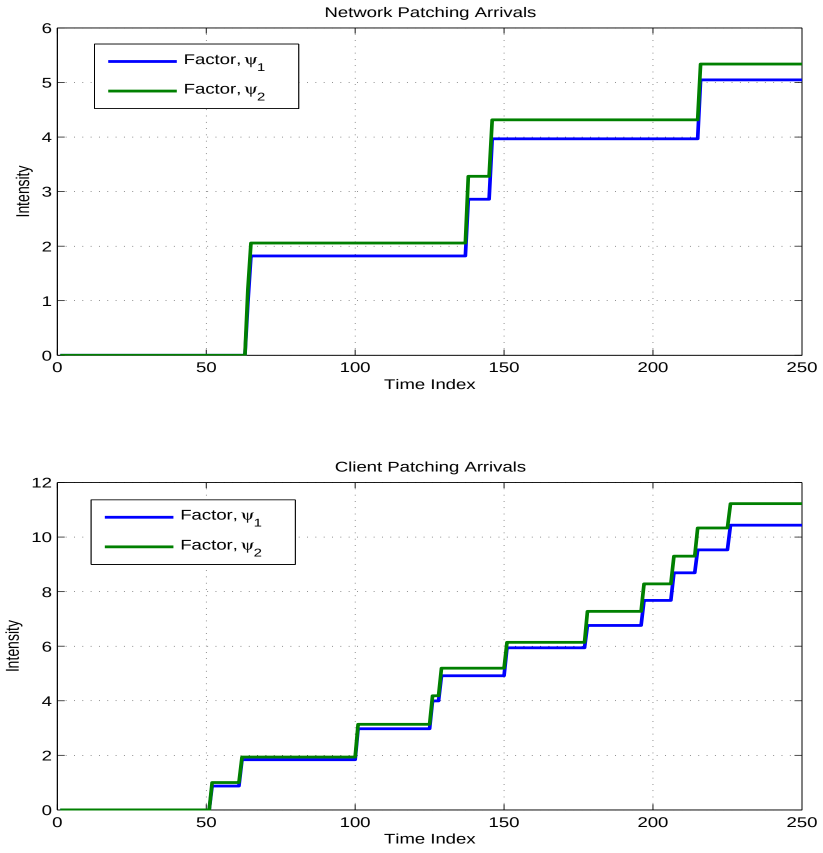

Fig. 4. Confidentiality (ψ_1)–Availability (ψ_2) Decompositions for Network and Client

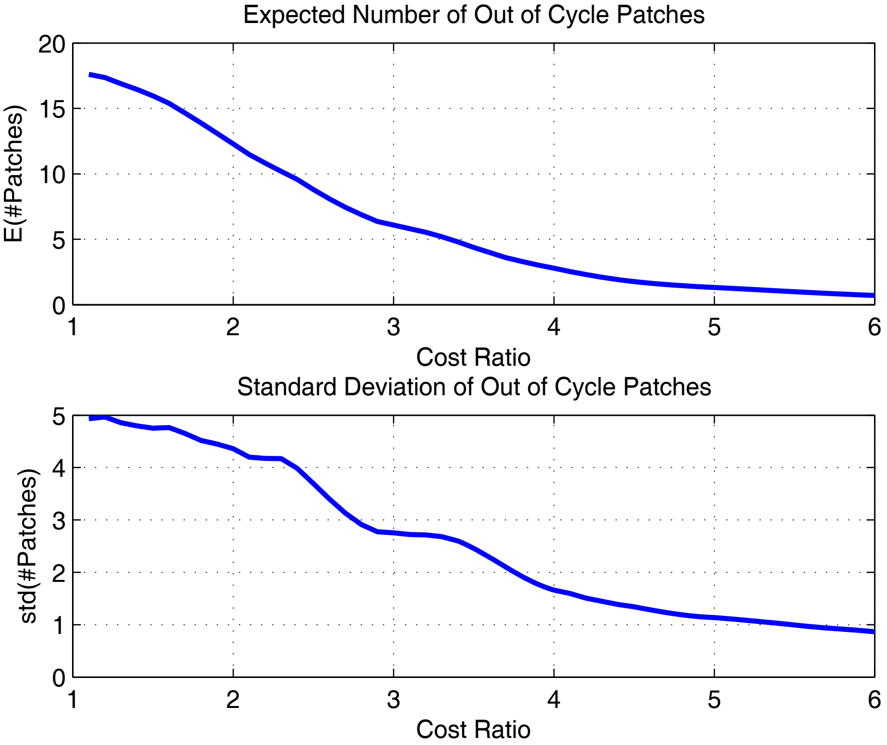

Fig. 5. Military Client Patching. Upper graph: Expected number of out-of-cycle patches. Lower graph: Standard deviation of out-of-cycle patches

Client Patching: Military versus Financial For client software patching, patches arrive often and have a low mean, albeit with a standard deviation that reflects the high variation in criticality between different client software $\frac{\mu}{\sigma} > 1$ for both C and A.

Figure 5 depicts against the cost ratio the expected number and standard deviation of out-of-cycle patches for a military organization, with parameter attributes shown in Table 1. We calibrate the cost of cycle patches such that the number of out-of-cycle patches is consistent with those observed in similar organizations with monthly patch management policies. Figure 6 presents the same information for the financial organization.

Several features are immediately apparent in this comparison. The difference between the preferences of the military and financial organizations results in a distinct profiles for out-of-cycle patching with respect to the same cost configurations. Military organizations *ceteris paribus* tend to apply more out-of-cycle patches than financial organizations, reflecting strong preference for confidentiality at the expense of availability.

The standard deviation of out-of-cycle patches for military organizations, albeit modest, suggests a tolerance for some low-level risks. The variation characterized by the standard deviation is indicative of the different applications of the system in, for example, battlefield and support functions. This is consistent with the profile of military organizations with a more flexible approach to taking systems off-line than financial firms with higher cost penalties (dominant C, weighting ω_1).

Figures 7 and 8 present, for military and financial organizations, the managers' total discounted loss, and average longest wait (in days, over a year) for a patch to be applied against cost ratio. We assume a discount rate of 6% per annum. Military organizations suffer higher losses through threats and tend to wait less time to patch over all cost ratios compared to financial firms. For example, given the same decomposition of the threat, and a cost ratio of 3, the maximum exposure time for client patching by financial firms is 10 days, whilst for the military organization it is less

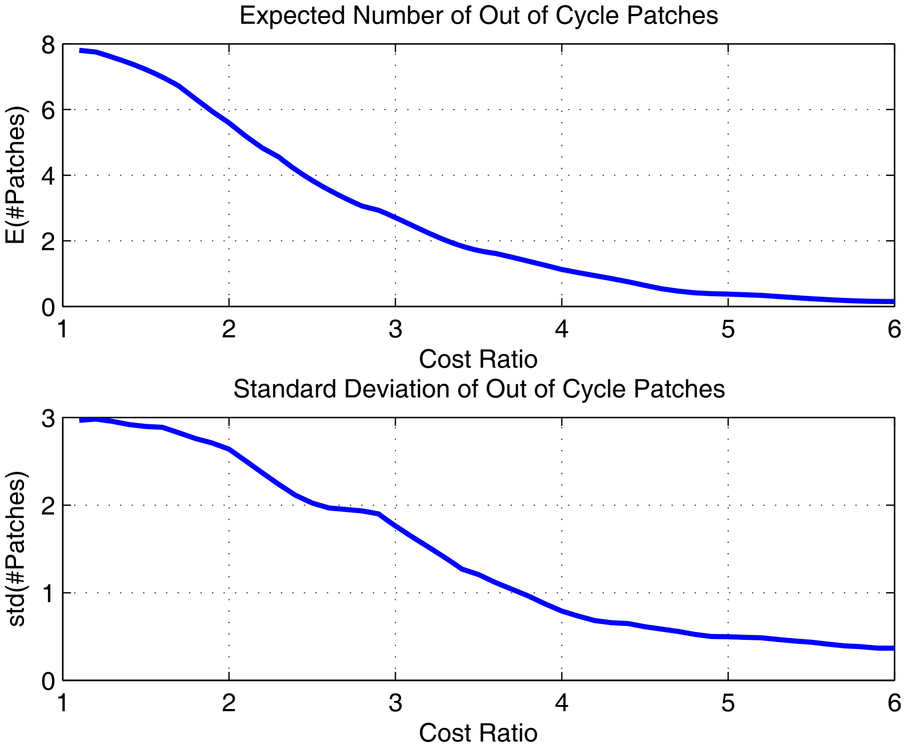

Fig. 6. Financial Client Patching. Upper graph: Expected number of out-of-cycle patches. Lower graph: Standard deviation of out-of-cycle patches

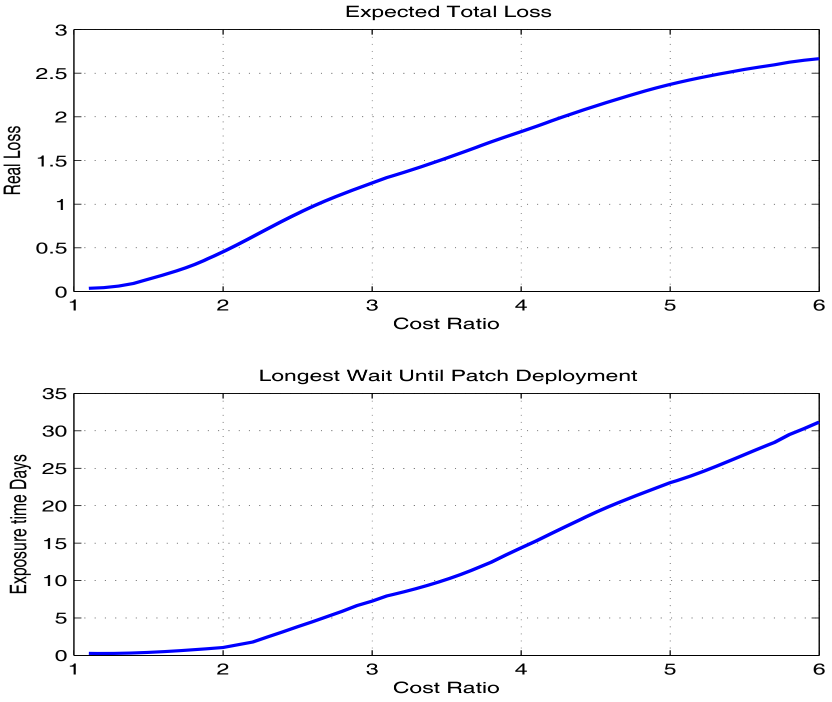

Fig. 7. Military Client Patching: Loss and longest wait

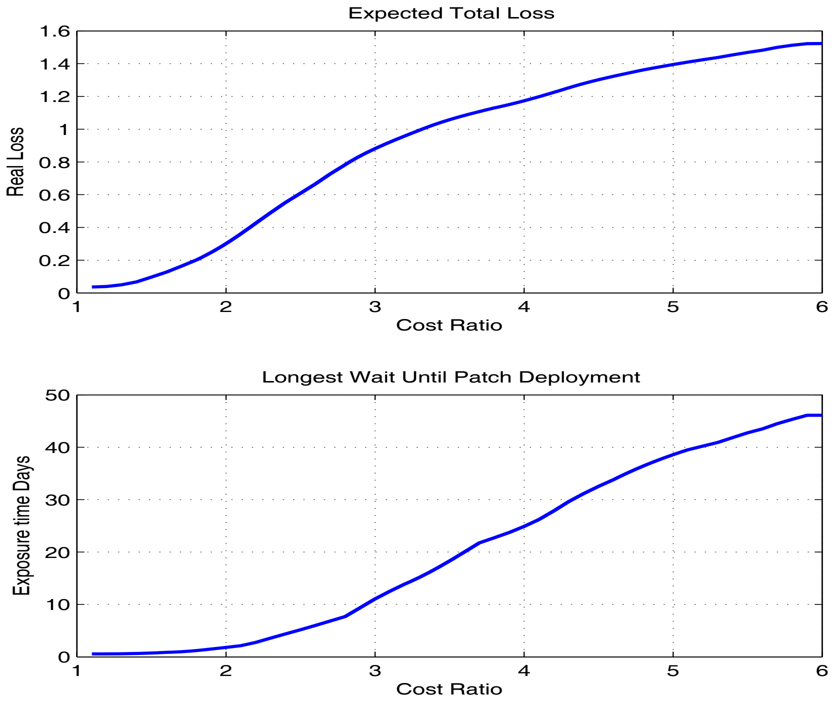

Fig. 8. Financial Client Patching: Loss and longest wait

than 6 days. Their losses stand at 0.9 and 1.5, respectively, reflecting the higher and unbalanced sensitivity of the military organization with respect to confidentiality. This sensitivity provides higher amplification to the factor of the threat that is attached to confidentiality in comparison with the factor assigned to availability. This profile is consistent throughout the range of simulated cost ratios.

Network Patching Military versus Financial Network patches are far less frequent than client patches, although the associated vulnerabilities are assumed to be uniformly more serious; that is, $\frac{\mu}{\sigma} < 1$, with a high average jump size μ for impacts on both C and A, and a higher level of correlation ρ between jump sizes. The arrival rate is calibrated to an average of 6 patches per year.

As in the client example, the upper graphs of Figures 9 and 10 present the expected number of out-of-cycle patches and the lower graphs of Figures 9 and 10 their respective standard deviations for military and financial organizations. Figures 11 and 12 present the total expected loss and the maximum tolerated exposure time for the two organizations.

The increased jump intensity of network (compared to client) vulnerabilities results in both organizations applying most patches as soon as they become available. In both cases, the standard deviation associated with the expected number of patches is very small, implying that there is no tolerance of the persistence of such vulnerabilities. Comparing the military and financial organizations, our simulations show that for the military organization the range of the economically meaningful cost ratio is about 3.[8] For this range, all patches are implemented upon arrival. Unlike

[8] For this parameter constellation, for a cost ratio beyond about 3, the loss rises so rapidly that the model gives only a corner solution. The simulation algorithm searches for a solution with the range of 1 to 7 days. Beyond a cost ratio of about 3, jumps to the 'extreme value' of about 7, and stays there. The change in early patch deployment is marked by a sudden switch away from instant patching at high cost

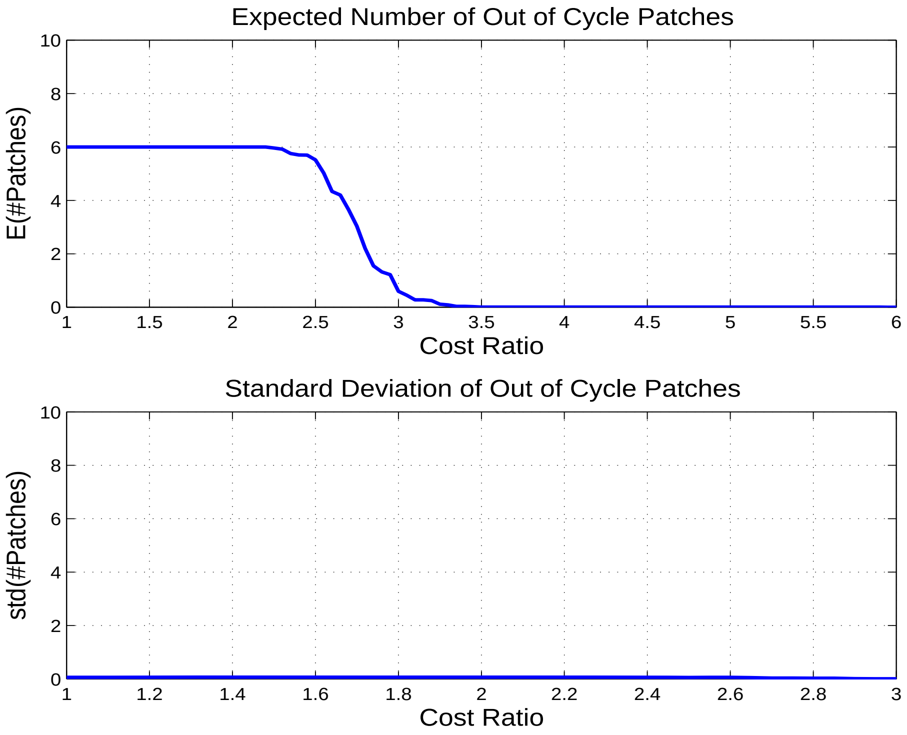

Fig. 9. Military Network Patching. Upper graph: Expected number of out-of-cycle patches. Lower graph: Standard deviation of out-of-cycle patches

the financial firm, this shows that patch implementation is sensitive to the cost ratio over a wider range. For the military organization, the longest wait until patch deployment is less than a day over the relevant range of the cost ratio, whilst, for the financial firm, the longest wait until patch deployment extends to 9 days over the full range of the cost ratio.

Comparing the patching policies for client and network in the financial organization that is sensitive to the cost ratio, the longest wait until patch deployment is 5 times higher for clients than for networks (45+ and 9, respectively). Similarly, for military organizations that exhibit a reduced sensitivity to the cost ratio, the difference, within the meaningful range, is between 3 days and 1 day (for a cost ratio of about 3).

Tables 3 and 4 summarize the main results from our simulations for the two organizations, for both client and network patching. The tables illustrate clearly the distinct patching profiles of the four situations.

In Table 3, the standard deviations correspond to the sensitivities of the same organizations to the relative severity of network and client threats.

	Client	Network
Military	5 (2.8)	6 (0)
Financial	2.2 (1.8)	0.75 (0.005)

Table 3. Number of irregular patches (standard deviation), for cost ratio 3

ratios, instead of the gradual decline observed in client patching. This is indicative of network patching being a far more acute policy problem, with very sharp discontinuities.

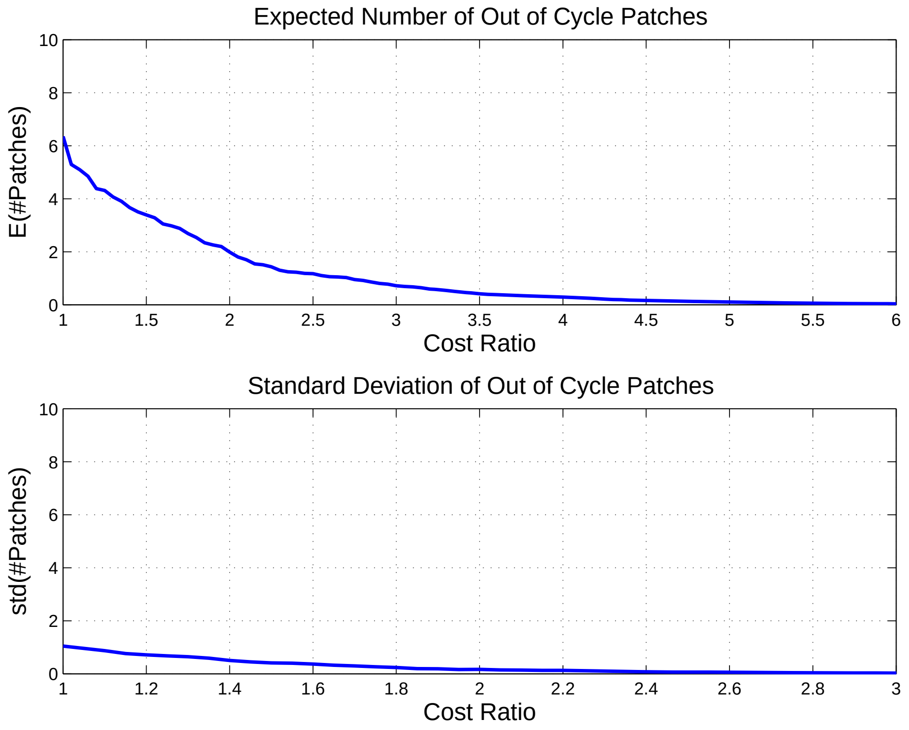

Fig. 10. Financial Network Patching. Upper graph: Expected number of out-of-cycle patches. Lower graph: Standard deviation of out-of-cycle patches

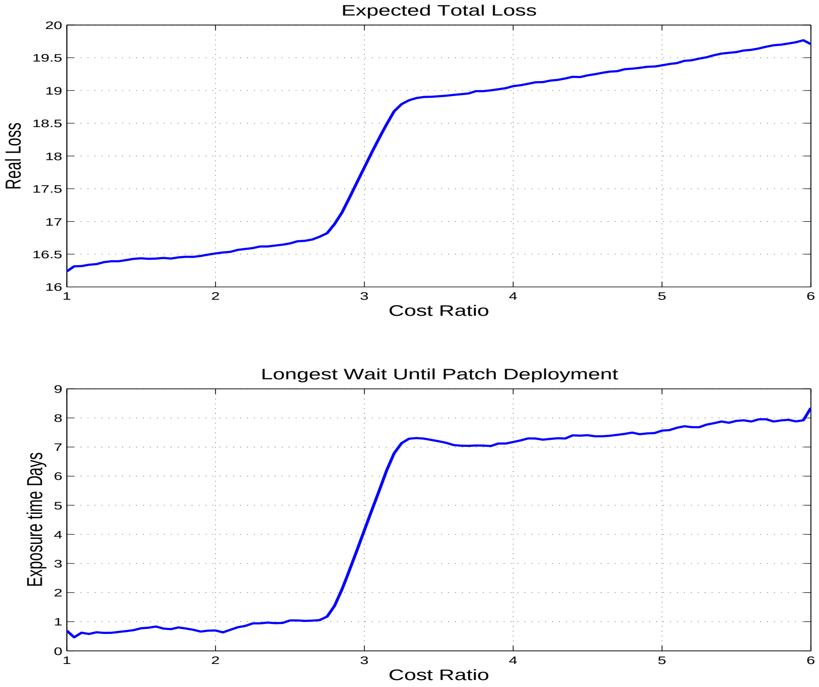

Fig. 11. Military Network Patching: Loss and longest wait

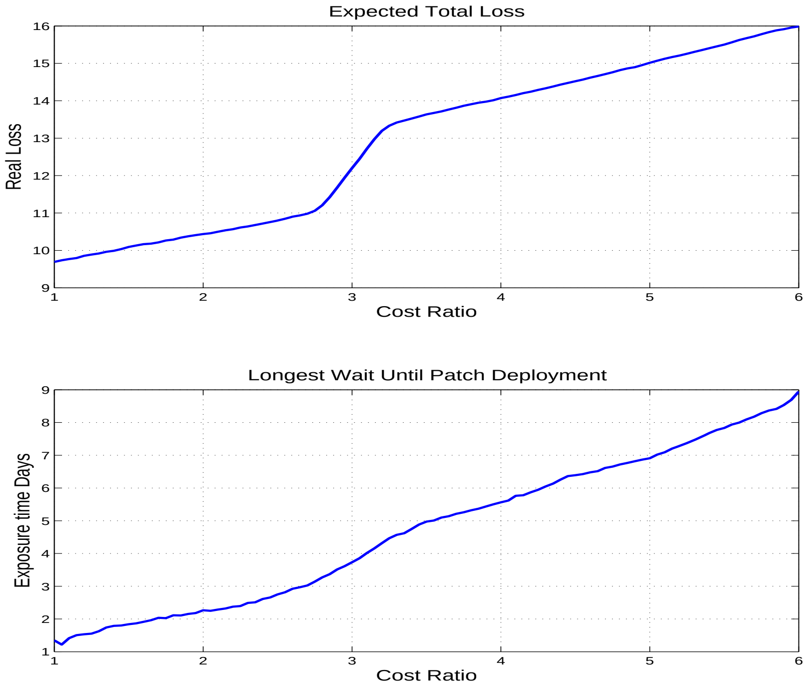

Fig. 12. Financial Network Patching: Loss and longest wait

Similarly, Table 4 indicates that network patches are mostly deployed upon arrival despite the associated cost.

	Client	Network
Military	6	0.5
Financial	10	3.5

Table 4. Longest wait to patch deployment, for cost ratio 3

6 Conclusions

We have introduced a model that develops a methodology for the determination of the optimal frequency for implementing patches in both network and client environments. The model recognises that patching is costly, but postponing deployment exposes the system to attacks by malware that may impair its performance and may result in the exposure of data as well as financial losses.

The methodology in this paper is based on the dynamics of the trade-offs between the attributes of the system that are of interest to the system's information security managers, using methods of optimization that are commonly deployed in economics. Operationalizing our approach will require more detailed models that more closely reflect the nature of specific organizations. Nevertheless, the approach developed in this paper constitutes a solid methodological base for addressing these problems.

Category	Parameters	Financial Network	Financial Client	Military Network	Military Client
Threat Environment	θ^{Ψ}	0.001	0.1	0.001	0.1
	θ_1^{Ψ}	0.1	0.001	0.1	0.001
	θ_2^{Ψ}	0.1	0.001	0.1	0.001
	θ_3^{Ψ}	0.2	0.001	0.2	0.001
	θ_4^{Ψ}	0.2	0.001	0.2	0.001
	θ_5^{Ψ}	0.2	0	0.2	0
System Properties	υ	$\in [1, 3]$	$\in [1, 3]$	$\in [1, 3]$	$\in [1, 3]$
	$\bar{\nu}$	50%	20%	35%	20%
	α'	50%	20%	35%	20%
	β'				
	$\pi_{1,1}, \ldots, \pi_{2,2}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Decision-maker's Preferences	ω_1	0.5	0.5	0.9	0.7
	ω_2	0.5	0.5	0.5	0.4
	ω_3	0.1	0.1	0.1	0.1
	$\bar{C}$	95%	90%	95%	90%
	$\bar{A}$	95%	90%	95%	90%
	$\bar{K}$	constant	constant	constant	constant

Table 5. Example parameter constellations for patching management based on representative industry calibrations

Following the approach developed in this paper, the system's operational status is characterized by availability and confidentiality that are exposed to serious risk of degradation owing to the potential for exploitation of vulnerabilities, as signalled by the arrival of patches. The model is based on the recognition that both IT managers and users appreciate the trade-off between the fundamental characteristics of information security considered here, namely confidentiality and availability. We have simplified our analysis by suppressing issues of integrity. The model's parameters can be clustered in a manner that allows us to categorize and compare the responses to shocks of various types of organizations.

We find that out-of-cycle patching is cost-sensitive and its deployment depends crucially on the preferences of the organization. In the case of client patching, we see that patching on arrival is not optimal, although tolerances to exposure are limited for military-type organizations. In the case of network patching, the military organization will patch on arrival for a wide range of cost differentials, whilst the financial organization exhibits a high degree of sensitivity to the same variate.

Finally, note that our approach does not rely for its predictions on the imputation of monetary losses associated with information degradation, but rather on metrics that are familiar to the information security community. We expect, therefore, to be able to validate the predications of our model empirically.

Acknowledgments

We are grateful to Yolanta Beres, Jonathan Griffin, and Simon Shiu (all at the Cloud and Security Lab, HP Labs, Bristol) for their advice on the state of the art in (modelling) patching strategies for both clients and networks. We are particularly grateful for their allowing us to reproduce their diagram of the 'Vulnerability Timeline' from Beres et al. (2008) [9].

References

1. R. Anderson. Why information security is hard: An economic perspective. In *Proc. 17th Annual Computer Security Applications Conference*, 2001.
2. R. Anderson, R. Böhme, R. Clayton, and T. Moore. Security economics and the internal market. Report to the European Network and Information Security Agency (ENISA), 2007, `http://www.enisa.europa.eu/doc/pdf/report_sec_econ_&_int_mark_20080131.pdf`.
3. R. Anderson and T. Moore. The economics of information security. *Science*, 314:610–613, 2006. Extended version available at `http://www.cl.cam.ac.uk/~rja14/Papers/toulouse-summary.pdf`.
4. A. Arbaugh, W.L. Fithem, and J. McHugh. Windows of vulnerability: A case study analysis. *IEEE Computer*, 2000.
5. A. Arora, R. Telang, and H. Xu. Optimal Policy for Software Vulnerability Disclosure. *Management Science*, 54(4):642–656, 2008.
6. T. August and T. Tunca. Network Software Security and User Incentives. *Management Science*, 52(11):1703–1720, 2006.
7. S. Beattie, S. Arnold, C. Cowans, P. Wagle, C. Wright, and A. Shostack. Timing the application of security patches for optimal uptime. In *LISA '02: 16th System Administration Conference*, 2002.
8. A. Beautement, R. Coles, J. Griffin, C. Ioannidis, B. Monahan, D. Pym, A. Sasse, and M. Wonham. Modelling the Human and Technological Costs and Benefits of USB Memory Stick Security. In M. Eric Johnson, editor, *Managing Information Risk and the Economics of Security*, pages 141–163. Springer, 2008.
9. Y. Beres, J. Griffin, S. Shiu, M. Heitman, D. Markle, and P. Ventura. Analysing the performance of security solutions to reduce vulnerability exposure window. In *Proceedings of the 2008 Annual Computer Security Applications Conference*, pages 33–42. IEEE Computer Society Conference Publishing Services (CPS), 2008.
10. Y. Beres, D. Pym, and S. Shiu. Decision support for systems security investment. In *Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP*, pages 118–125, 2010. doi: 10.1109/NOMSW.2010.5486590, ISBN: 978-1-4244-6037-3, INSPEC Accession Number: 11502735.
11. Catherine Bobtcheff and Stphane Villeneuve. Technology choice under several uncertainty sources. *European Journal of Operational Research*, 206(3):586–600, 2010.
12. H. Cavusoglu, H. Cavusoglu, and J. Zhang. Security Patch Management: Share the Burden or Share the Damage. *Management Science*, 54(4):657–670, 2008.
13. M. Collinson, B. Monahan, and D. Pym. Semantics for structured systems modelling and simulation. In *Proc. Simutools 2010*. ICST: ACM Digital Library and EU Digital Library, 2010. ISBN: 78-963-9799-87-5.
14. G. Coulouris, J. Dollimore, and T. Kindberg. *Distributed Systems: Concepts and Design.* Addison Wesley, 2005.
15. D.R. Cox and V. Isham. *Point Processes.* Monographs on Statistics and Applied Probability. Chapman and Hall, 1980.
16. Demos2k. `http://www.demos2k.org`.
17. S. Frei, M. May, U. Fiedler, and B. Plattner. Large-scale vulnerability analysis. In *Proc. of SIGCOMM'06 Workshop*. Association for Computing Machinery, 2006. Available at `www.techzoom.net/papers/sigcomm_lsad_large_scale_vulnerability_analysis_2006.pdf`.
18. M.P. Giannoni and M. Woodford. Optimal Interest-Rate Rules I: General Theory. Working Paper Series 9419, National Bureau of Economic Research, 2002. ISSU 9419, ISSN 0898-2937.
19. Gnosis. `http://www.hpl.hp.com/research/systems_security/gnosis.html`.
20. L. Gordon, M. Loeb, and W. Lucyshyn. Information security expenditures and real options: A wait-and-see approach. *Computer Security Journal*, 19(2):1–7, 2003.
21. L.A. Gordon and M.P. Loeb. The Economics of Information Security Investment. *ACM Transactions on Information and Systems Security*, 5(4):438–457, 2002.
22. L.A. Gordon and M.P. Loeb. *Managing Cybersecurity Resources: A Cost-Benefit Analysis.* McGraw Hill, 2006.
23. Kjell Hausken and Vicki M. Bier. Defending against multiple different attackers. *European Journal of Operational Research*, 211(2):370 – 384, 2011.
24. J.C. Hersey, H.C. Kunreuther, and P.J. Shoemaker. Sources of bias in assessment procedures for utility functions. *Management Science*, 28:936–953, 1982.

25. C. Ioannidis, D. Pym, and J. Williams. Investments and trade-offs in the economics of information security. In Roger Dingledine and Philippe Golle, editors, *Proceedings of Financial Cryptography and Data Security '09*, volume 5628 of *LNCS*, pages 148–166. Springer, 2009. Preprint available at `http://www.cs.bath.ac.uk/~pym/IoannidisPymWilliams-FC09.pdf`.
26. J.Y. Jaffrey. Some experimental findings on decision-making under risk and their implications. *European Journal of Operational Research*, 38:301–306, 1989.
27. A. Jimenéz, S. Ros-Insua, and A. Mateos. A decision support system for multi-attribute utility evaluation based on imprecise assignments. *Decision Support Systems*, 36:65–79, 2003.
28. E. Jonsson and A. Olovsson. Quantitative model of the security intrusion process based on attacker behaviour. *IEEE Transactions on Software Engineering*, 23(4):235–245, 1997.
29. R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. Wiley, 1976.
30. M. McCord and R. de Neufville. Lottery equivalents: reduction of the certainty effect problem in utility assessment. *Management Science*, 32:56–61, 1986.
31. R.A. Nobay and D.A. Peel. Optimal Discretionary Monetary Policy in a Model of Asymmetric Bank Preferences. *Economic Journal*, 113(489):657–665, 2003.
32. Enrico Pennings and Onno Lint. Market entry, phased rollout or abandonment? a real option approach. *European Journal of Operational Research*, 124(1):125 – 138, 2000.
33. P. Protter and D. Talay. The Euler Scheme for Levy Driven Stochastic Differential Equations. *The Annals of Probability*, 25(1):393–423, 1997.
34. B. Schneier. Managed security monitoring: Closing the window of exposure. Counterpane Internet Security. Manuscript available at: http://www.counterpane.com/window.pdf, 2000.