Enterprise IT service downtime cost and risk transfer in a supply chain

Shaun S. Wang^{1,2} · Ulrik Franke³

Received: 5 April 2019 / Revised: 20 September 2019 / Accepted: 9 January 2020 / Published online: 21 February 2020
© The Author(s) 2020

Abstract

In this paper we present an economic model for analyzing enterprise IT service downtime cost, first on a standalone basis and then in a supply chain setting. With a baseline probability model of Poisson arrival frequency with random downtime duration, we analyze optimal production of a firm's investments in reducing frequency and duration of downtime, and corresponding premiums for insuring against downtime cost. We also present a model for the spillover effect of downtime for interconnected firms in a supply chain, and discuss how third-party insurance coverage can help enterprises to internalize the externalities of spillover effects on the supply chain.

Keywords Enterprise IT service · Downtime cost · Supply chain

1 Introduction

An enterprise may face IT service downtime costs, due to a variety of causes including antagonistic IT attacks by hackers, non-antagonistic IT service outages, or natural catastrophes such as floods or solar storms. The potential losses from downtime can be highly significant, with hourly costs of IT service outages ranging from hundreds of thousands to even millions of US dollars (Rapoza 2014; Ponemon 2016), at least for large companies in certain industries. These large downtime costs are not unique to our present time and had been around since 20 years ago (IBM Global Services 1998). However, the sheer level of interdependence in modern IT service ecosystems is indeed a recent phenomenon. The advent of the internet over

the past two decades unlocked the potential of integrated supply chains for small and medium-sized enterprises (Stefansson 2002), where previously such supply chains had been the sole privilege of large companies. However, it is the ubiquity of cloud services in recent years that have increased today's complex dependencies – along with the fact that IT integration of supply chains does lead to more competitive business performance (Swafford et al. 2008). As a consequence, there is no shortage of research on the reliability, e.g., Dai et al. (2009) and Wood et al. (2010) and security, e.g., Takabi et al. (2010) and Subashini and Kavitha (2011) of cloud services.

In this paper, we analyze the cost of IT service downtime, first for a single firm, and then from the perspective of interdependencies between firms in a supply chain, where the downtime of one enterprise IT service may indirectly trigger further business interruptions at other firms. This is the case when a firm buys services such as credit-card payment, authentication, physical storage and associated inventory status information, customer relations data storage, bookkeeping, etc. from third parties. For a business process to execute, e.g. a customer successfully placing an order, the services of these other firms need to be available at the same time. From an economics perspective, such interdependency has interesting consequences as downtime at one firm may affect multiple other firms in a supply chain. Specifically, security investments and resource allocation by one firm may impose externalities on other firms connected to it in the supply chain (Zhao et al.

✉ Ulrik Franke
ulrik.franke@ri.se

Shaun S. Wang
wangsx@sustech.edu.cn

¹ Insurance Risk and Finance Research Centre, Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore, Singapore

² Present address: Department of Finance, Southern University of Science and Technology, Shenzhen, Guangdong, China

³ RISE Research Institutes of Sweden, P.O. Box 1263, SE-164 29 Kista, Sweden

2013). Whether there is a role for insurance to internalize this externality is also a question of both theoretical interest and practical implications. As a risk management tool, firms can purchase insurance coverage to manage the unforeseen consequence of downtime costs. A recent study by Lloyd's showed that a few days of downtime for a top cloud service provider would result in billions of US dollars in insured losses in the US alone (Lloyd's 2018). Pondering this, it should be remembered that even though the US is one of the most mature markets in terms of insurance, only a tiny fraction of the losses are insured (due to low take-up rates and low coverage limits in the policies offered). The uninsured losses, borne by the firms themselves, would be much greater in such a scenario.

It is this background of large costs, interdependencies, externalities and the role of insurance that motivates our work. Following a brief literature review in Section 2, we introduce a stochastic model of IT service downtime in Section 3. Section 4 then extends this model with the downtime costs to a firm, and Section 5 addresses the resulting resource allocation problem. Section 6 extends the model to the supply chain setting. In Section 7, we discuss how insurance products covering business disruption can help firms in a supply chain to manage the risk related to downtime costs. Finally, Section 8 concludes the paper.

2 Related work

As noted above, the general areas of digital supply chain and cloud computing security and reliability are well researched. However, the seemingly largest strand of research in cloud security primarily addresses antagonistic threats and data breaches (Ali et al. 2015; Ahmed and Hossain 2014). These are certainly pertinent issues, but a bit different from our focus, as we address downtime costs caused by both antagonistic and non-antagonistic factors. Furthermore, much of the technical literature delimits itself from economic aspects of availability, including the cost and incentive issues that we address. In the following, we briefly review some of the most relevant literature related to our contribution.

2.1 Economics of outages, security and supply chain resilience

From an economics perspective, supply chain security has been analyzed previously from several angles. Jenjarrusakul et al. (2013) consider sectoral and regional interdependence of Japanese firms under the influence of information security, in a model where sectoral security risk levels are represented by survey data. Their work is similar to ours in investigating the dependencies between firms, but as

a descriptive model, it differs from our more prescriptive model used to analyze how to minimize downtime cost and design insurance products covering business disruption.

Thomas et al. (2013) focus on the complexity of managing an incident in a supply chain, where many partners need to be coordinated. Their work is similar to ours in addressing the interplay between supply chain actors, but differs importantly by focusing on data breach rather than service outage.

Dynes et al. (2005) offer an interesting interview-based investigation of firms' self-assessment of their supply chain resilience in the face of internet outages up to a week. The results indicate overconfidence, and even though it is difficult to imagine that executives today maintain that their supply chains remain robust in the face of a week-long internet service outage, the relative overconfidence might still be similar.

Dynes et al. (2007) investigate the costs of internet outages to the US economy. Importantly, they estimate the multipliers by which the direct costs are related to the indirect ones, e.g. through supply chain ripple effects. While this is similar to our area of interest, Dynes et al. (2007) are more descriptive of fait accompli consequences, and do not focus on decision problems and incentives for individual firms to prevent outages.

Shahrad and Wentzlaff (2016) propose how the inefficiency of best-effort availability for cloud computing customers with varying demands and willingness to pay for availability can be improved upon by flexible availability offerings. They offer an OpenStack prototype and encouraging simulation results of improved profit margins. A main difference compared to our work is that Shahrad and Wentzlaff (2016) focus on the situation where a single service-provider uni-directionally serves many diverse customers, rather than the more general case of bi-directional dependencies that we address.

Aceto et al. (2018), in a recent survey of the literature on internet outages, note that this literature is rather scattered and difficult to get an overview of. While "internet outages" only partially overlap with our object of study, it is still worth observing that one of the open issues identified in the survey concerns lack of proper models for risk assessment. In particular, this is identified as a problem for insurers, and a barrier to properly priced insurance (ibid, section 7, p. 51). Our study makes a contribution to precisely this area.

In the literature there is abundant exposition of production of security investment (Gordon and Loeb 2002; Anderson and Moore 2006; Matsuura 2009), externalities of security investment (Camp and Wolfram 2004; Muermann and Kunreuther 2008; Herley 2009). However, the mainstream strand in this research field is externalities imposed by underinvestment in security against antagonistic threats. For example, Camp and Wolfram (2004) discuss

vulnerabilities that can be exploited by attackers. This differs from our focus on downtime, which can be either non-antagonistic or antagonistic in nature.

2.2 Cyber insurance

As we discuss implications for insurance, it is also relevant to consider the cyber insurance literature. Böhme and Kataria (2006) discuss correlations of cyber risks in insurance portfolios. This is similar to our interest in interdependencies between firms (second tier correlations, in the terminology of Böhme & Kataria), but different in focus: Böhme & Kataria model loss events using attack data obtained from honeypots, whereas we do not distinguish between downtime events caused by cyberattacks and other known or unknown, natural or human causes. Furthermore, as we analyze downtime, we are interested in durations as well as occurrences of events. Böhme & Kataria use a discrete (beta-binomial) distribution to model their loss events while our model is based on Poisson arrival frequency and random downtime duration.

From an empirical cyber insurance point of view, it should be noted that the downtime costs of business interruption are insurable. Franke (2017) notes that some kind of coverage for business interruption caused by antagonistic factors is covered by all insurance policies on offer in Sweden, whereas business interruptions caused by non-antagonistic factors are treated differently (some insurers exclude it, some include it, while some offer it as an option). A CCRS/RMS survey cited by the OECD (2017) found that business interruption resulting from third-party disruption was covered by only a third of the policies investigated.

2.3 Scope and contribution

This paper contributes to the literature by a cross-pollination of various academic fields, including information and system security, management information systems, economics of information security, insurance and actuarial science. This paper presents novel applications of popular probability models of IT service breakdown and duration of breakdown, as well as economic impacts and insurance premium calculations.

3 Probability model of downtime for an enterprise

For simplicity we consider a generic firm that offers a single IT service, with a target average availability (the fraction of total time that the service is working as it should) such as 99.95%. Traditionally, the average availability is often

defined as the ratio of Mean Time To Failure (MTTF) to the sum of Mean Time To Failure (MTTF) and the Mean Time To Restore (MTTR):

$$A = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}} \quad (1)$$

We note that the conventional measure of average availability A in Eq. 1 does not differentiate between the following two cases:

- i) 12 times of outage in a year, with 2 hours downtime each time
- ii) 1 single outage in a year, with downtime lasting 24 hours

Assuming 365 days per year, and 24 hours per day, conventional measure of average availability A in Eq. 1 would be the same for both cases, namely, $A = \frac{364}{365}$. However, these two cases may have different economic impacts. Suppose that the firm has an insurance policy that covers business interruption due to downtime, with a deductible of 8 hours of waiting time. Then, the insurance pay-out would be zero in case i) but non-zero in case ii).

We thus propose a probability-based model for IT service breakdowns.

3.1 Frequency of breakdown – poisson arrival model

Consider a single generic firm. We assume that the number of breakdowns in a given time interval follows a Poisson arrival process with intensity λ . The random number N of IT service breakdowns in any given one year time interval is

$$\Pr\{N = n\} = \exp(-\lambda) \cdot \frac{\lambda^n}{n!}, \quad \text{for } n = 0, 1, 2, 3, \dots$$

The Poisson arrival process can also be defined by stating that the time intervals between IT service breakdowns are exponential variables with mean $1/\lambda$ (e.g. Ross 1996).

Remark 1 The Poisson process is one of the most widely-used counting processes; it is used when the occurrences of certain events happen at a certain rate, but completely at random. For example, suppose that from historical data, we know that earthquakes occur in a certain area with a rate of 1 per five years. Other than this information, the exact timings of earthquakes seem to be completely random. Thus, we conclude that the Poisson process might be a good model for earthquakes. In practice, the Poisson process or its extensions have been used to model the number of car accidents at a location, the number of requests for individual documents on a web server in a given day, etc. Kingman (1992) and Haenggi (2012). More specifically, Poisson arrivals of outages are a common assumption in the literature on IT service reliability and availability,

e.g. Martinello et al. (2005); Jeske and Xuemei Zhang (2005); Franke (2012, 2016, 2019); Yamada (2014); Taylor and Ranganathan (2014).

3.2 Downtime duration model

Given that IT service breakdown occurs, the random duration of a breakdown T follows a probability distribution with raw moments

$$\mu^{(k)} = E[T^k], \quad \text{for } k = 1, 2, 3, \dots \quad (2)$$

Specifically, $\mu^{(1)}$ is the average downtime T per occurrence.

To facilitate later calculation of premium for insurance against downtime cost, we introduce a concept of deductible (waiting time) d , which is defined such that the downtime can be decomposed into two parts: $T = \min(T, d) + (T - d)_+$ where

$$\min(T, d) = \begin{cases} T, & T \leq d \\ d, & T > d \end{cases}; \quad (T - d)_+ = \begin{cases} 0, & T \leq d \\ T - d, & T > d \end{cases} \quad (3)$$

Note that the expected value of downtime in excess of the deductible d is:

$$E[(T - d)_+] = \mu^{(1)} - E[\min(T, d)]$$

For IT services, the lognormal distribution is often used to model outage durations (Schroeder and Gibson 2010; Franke et al. 2014). A lognormal distribution is defined by the probability density function:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - a)^2}{2\sigma^2}\right), \quad \text{for } x > 0. \quad (4A)$$

The k -th raw moment of the lognormal distribution is

$$\mu^{(k)} = \exp(ka + k^2\sigma^2/2), \quad \text{for } k = 1, 2, 3, \dots \quad (4B)$$

The lognormal distribution has a limited expected value (Bahnmann 2015):

$$E[\min(T, d)] = \mu^{(1)} \cdot \Phi\left(\frac{\ln(d) - a - \sigma^2}{\sigma}\right) + d\Phi\left(\frac{-\ln(d) + a}{\sigma}\right) \quad (4C)$$

where Φ is the standard normal cumulative distribution function.

Remark 2 The downtime duration can have other distributions and can take either discrete or tabular form.

4 Cost of downtime as function of duration

Consider the firm's downtime costs in a given year. Assume that given an occurrence of downtime, the firm's downtime cost $C(T)$ is a function of the duration T of downtime.

Generally, $C(T)$ should be a non-decreasing function of the downtime duration T . Here are three simple cost functions:

1. Linear: $C(T) = c \cdot T$, for some $c > 0$
2. Quadratic: $C(T) = c \cdot T^2$, for some $c > 0$
3. Constant: $C(T) = c$, for some $c > 0$

Remark 3 As empirical studies of IT service outage costs are rare, it is interesting to consider several different functional forms of the cost. Conceptually, it is worth distinguishing *revenue lost* from *repair costs* (Oppenheimer et al. 2003), and explore cost functions capable of capturing both. The *linear model* is in a sense the simplest one, corresponding to models such as Patterson's (2002), where the cost of an outage hour is always the same. The *quadratic model*, by contrast, corresponds to cases where costs snowball (Franke 2012; Vecchio 2016) so that short outages are barely noticeable, but longer ones can have dire consequences. An example is when customers stop using an ATM or credit card payment system even after it has come back online, because a long outage gave it a bad reputation. Such reputational impacts of service outages can be considerable, including plummeting stock prices (Bharadwaj et al. 2009), and be difficult to manage (Boritz and Mackler 1999; Manika et al. 2015). The *constant model* is a contrast to the other two, letting cost be independent of outage duration. It thus represents the case where fixed restart costs overshadow variable outage costs. Meland et al. (2017) cite data showing that *response costs* are on average three times as great as *loss of business income*. Though these categories are not exactly equivalent to fixed vs. variable costs, such data indicate that fixed, duration-independent, costs cannot be ignored. Physical industrial processes are good examples of IT dependent operations that have considerable fixed restart costs.

Note that the firm can incur a random number N of IT service breakdowns, where N has a Poisson distribution with mean λ . The n -th breakdown has a random duration T_n and incurs a downtime cost $C(T_n)$. Thus, in a given year, the firm's *aggregate downtime cost* S is a random sum of N independent random variables:

$$S = C(T_1) + C(T_2) + \dots + C(T_N) \quad (5A)$$

The aggregate downtime cost S in Eq. 5A has a compound Poisson distribution with the following mean and variance (Klugman et al. 2012):

$$E[S] = \lambda \cdot E[C(T)] \quad \text{Var}[S] = \lambda \cdot E[(C(T))^2] \quad (5B)$$

Now we analyze each of the three simple downtime cost functions.

4.1 Linear model

$C(T) = c \cdot T$ for some coefficient, $c > 0$.

The downtime cost per occurrence has the following moments:

$$E[C(T)] = c \cdot \mu^{(1)} \quad E[(C(T))^2] = c^2 \cdot \mu^{(2)}$$

The firm's aggregate downtime cost S in a given year has the following mean and variance:

$$E[S] = \lambda \cdot c \cdot \mu^{(1)} \quad \text{Var}[S] = \lambda \cdot c^2 \cdot \mu^{(2)}$$

4.2 Quadratic model

$C(T) = c \cdot T^2$ for some coefficient, $c > 0$.

As in Franke (2012), the coefficient c represents snowball effects.

The downtime cost per occurrence has the following moments:

$$E[C(T)] = c \cdot \mu^{(2)} \quad E[(C(T))^2] = c^2 \cdot \mu^{(4)}$$

The firm's aggregate downtime cost S in a given year has the following mean and variance:

$$E[S] = \lambda \cdot c \cdot \mu^{(2)} \quad \text{Var}[S] = \lambda \cdot c^2 \cdot \mu^{(4)}$$

4.3 Constant model

$C(T) = c$ for some coefficient, $c > 0$.

The downtime cost per occurrence has the following moments:

$$E[C(T)] = c \quad E[(C(T))^2] = c^2$$

The firm's aggregate downtime cost S in a given year has the following mean and variance:

$$E[S] = \lambda \cdot c \quad \text{Var}[S] = \lambda \cdot c^2$$

5 Resource allocations to cost reduction and risk transfer

It can be demonstrated that one effective way of reducing downtime is by building redundancy. Normally, the frequency of breakdown of any one component is small,

(λ is near zero). To illustrate the effect of redundancy, e.g. when two independent payment systems have been procured, the corresponding expected downtime frequency in the payment service is approximately reduced to λ^2 . Nevertheless, the sum of multiple independent sources of breakdown can increase the Poisson frequency λ . In general, more complex cases of architectural dependencies between IT services can be modeled using *fault trees* (Närman et al. 2014).

The concept of building redundancy to reduce breakdown frequency is not new. In the aviation industry, commercial aircrafts have almost everything at least in duplicate, for instance, multiple engines, auxiliary fuel pumps, dual spark plugs, dual electrical displays and circuitry. In addition to this redundancy, there are strict maintenance and training requirements for all commercial aircrafts. This redundancy leads to an extremely small likelihood of an airliner completely "breaking down".

Now we return to discussions of enterprise IT services. We assume that a firm has already deployed *baseline investments* to assure some level of redundancy which corresponds to a baseline frequency-duration model:

- The *baseline* breakdown frequency N_0 has a Poisson distribution with mean λ_0 , and
- The baseline downtime duration T_0 has raw moments $\mu_0^{(k)} = E[T_0^k]$, for $k = 1, 2, 3, \dots$

5.1 Uptime production and resource allocation

To facilitate the firm's decision problem in resource allocation, assume that the firm can invest Capital K to reduce frequency, and invest Labor L in reducing duration of downtime. We assume a Cobb-Douglas (1928) type production function for IT service availability:

1. Investment K in reducing the frequency (e.g., investing in prevention and redundancy):

$$\lambda(K) = \lambda_0 \cdot K^{-\alpha}, \quad \text{for some } \alpha \geq 0 \quad (6A)$$

2. Investment L in reducing the duration (e.g., investing in detection and response) by a scaling factor:

$$\Pr\{T(L) \leq t\} = \Pr\{T_0 \cdot L^{-\beta} \leq t\}, \quad \text{for } \beta \geq 0 \quad (6B)$$

Thus the k -th raw moment of $T(L)$ satisfies

$$\mu^{(k)}(L) = \mu_0^{(k)} \cdot L^{-k\beta}, \quad \text{for } k = 1, 2, 3, \dots \quad (6C)$$

Remark 4 Franke (2014) considered a similar Cobb-Douglas type production model without using a probabilistic model. The intuition behind the model is that Capital K can buy better hardware (or similar hardware, to build redundancy), reducing the frequency of downtime (i.e.

increasing the MTTF), while Labor L can be used to monitor the system and take swift action when it fails, which also reduces the average duration (decreasing the MTTR). Of course, Labor and Capital need not be taken literally, but can rather be seen as stylized descriptions of two different kinds of available investments.

With the Cobb-Douglas type production model in place, the firm faces a total cost (which is the sum of investments K and L made to maintain availability, and the residual aggregate downtime cost S) as follows:

$$E[\text{Total Cost}] = K + L + E[S] = K + L + (\lambda_0 \cdot K^{-\alpha}) \cdot E[C(T(L))] \quad (7A)$$

The variance of the firm's aggregate downtime cost S is

$$\text{Var}[S] = (\lambda_0 \cdot K^{-\alpha}) \cdot E[C(T(L))^2] \quad (7B)$$

Proposition 1 Consider the Linear Model: $C(T) = c \cdot T$, for some coefficient $c > 0$.

a) For a given fixed budget $K + L = M$, the optimal allocations of resources to minimize the expected total cost (7A) are:

$$K^* = \frac{\alpha}{\alpha + \beta} M, \quad L^* = \frac{\beta}{\alpha + \beta} M \quad \text{and} \quad \frac{K^*}{L^*} = \frac{\alpha}{\beta} \quad (8A)$$

The optimal level M that minimizes the total cost (7A) has a closed-form formula:

$$M^* = (\alpha + \beta) \left(\alpha^{-\alpha} \beta^{-\beta} \lambda_0 c \mu_0^{(1)} \right)^{\frac{1}{\alpha + \beta + 1}} \quad (8B)$$

With the resource allocations in Eqs. 8A and 8B, the minimal expected total cost is

$$E[\text{Total Cost}] = \frac{\alpha + \beta + 1}{\alpha + \beta} M^* \quad (8C)$$

b) The resource allocations that minimize the variance of aggregate downtime cost (7B) are:

$$K^{**} = \frac{\alpha}{\alpha + 2\beta} M, \quad L^{**} = \frac{2\beta}{\alpha + 2\beta} M \quad \text{and} \quad \frac{K^{**}}{L^{**}} = \frac{\alpha}{2\beta}$$

The level M that minimizes the variance of aggregate downtime cost (7B) is:

$$M^{**} = (\alpha + 2\beta) \left(\alpha^{-\alpha} (2\beta)^{-2\beta} \lambda_0 c^2 \mu_0^{(2)} \right)^{\frac{1}{\alpha + 2\beta + 1}}$$

Proof in [Appendix](#).

Remark 5 From Eq. 8B, the optimal level of investment M^* is an increasing function of both the baseline expected frequency λ_0 and the baseline expected downtime cost $c\mu_0^{(1)}$. Minimizing the expected total cost and decreasing the variance of the total cost, respectively, are the two availability management strategies discussed by Franke (2012, especially Section VI.A). When empirically investigating how IT professionals act in procuring availability Service Level Agreements (SLAs), however, most did not minimize the expected total cost, and many exhibited decision-making patterns not easily explained (Franke and Buschle 2016).

Proposition 2 Consider the Quadratic Model: $C(T) = c \cdot T^2$, for some coefficient $c > 0$.

a) For a given fixed budget $K + L = M$, the optimal allocations of resources to minimize the expected total cost (7A) are:

$$K^* = \frac{\alpha}{\alpha + 2\beta} M, \quad L^* = \frac{\beta}{\alpha + 2\beta} M \quad \text{and} \quad \frac{K^*}{L^*} = \frac{\alpha}{2\beta} \quad (9A)$$

The optimal level M that minimizes the total cost (7A) has a closed-form formula:

$$M^* = (\alpha + 2\beta) \left(\alpha^{-\alpha} (2\beta)^{-2\beta} \lambda_0 c \mu_0^{(2)} \right)^{\frac{1}{\alpha + 2\beta + 1}} \quad (9B)$$

With the resource allocations in Eqs. 9A and 9B, the minimal expected total cost is

$$E[\text{Total Cost}] = \frac{\alpha + 2\beta + 1}{\alpha + 2\beta} M^* \quad (9C)$$

b) The resource allocations that minimize the variance of aggregate downtime cost (7B) are:

$$K^{**} = \frac{\alpha}{\alpha + 4\beta} M, \quad L^{**} = \frac{4\beta}{\alpha + 4\beta} M \quad \text{and} \quad \frac{K^{**}}{L^{**}} = \frac{\alpha}{4\beta}$$

The level M that minimizes the variance of aggregate downtime cost (7B) is:

$$M^{**} = (\alpha + 4\beta) \left(\alpha^{-\alpha} (4\beta)^{-4\beta} \lambda_0 c^2 \mu_0^{(4)} \right)^{\frac{1}{\alpha + 4\beta + 1}}$$

Proof in [Appendix](#).

Remark 6 From Proposition 2, large snowball effect of the cost of downtime would indicate relatively more investment to shorten the duration of downtime.

Proposition 3 Consider the Constant Model: $C(T) = c$. For a given fixed budget $K + L = M$, the optimal allocations are $K = M$ and $L = 0$. Specifically, we have:

$$M^* = (\alpha \cdot \lambda_0 \cdot c)^{\frac{1}{\alpha+1}}$$

Proof in [Appendix](#).

Remark 7 A large fixed cost drives investment into minimizing the number of breakdowns. The result from the degenerate constant model is extreme, but a more realistic model of the allocation problem might be a weighted sum of the constant, linear, and quadratic models.

5.2 Risk transfer of downtime cost by insurance

Suppose that the firm can transfer the downtime cost to an insurer by purchasing insurance that reimburse the firm's cost of breakdown after a deductible (waiting time) d . The actuarially fair insurance premium can be calculated as

$$\text{Premium} = \lambda \cdot E[C((T - d)_+)]$$

It is noted that the level of a firm's investments K and L can directly affect the insurance premium.

Example 1 Assume that the baseline frequency follows a Poisson arrival with $\lambda_0 = 0.5$, and the baseline downtime duration T_0 follows a lognormal distribution (4A) with $a = 0.7$ and $\sigma = 1.2$. From Eq. 4B we have $\mu_0^{(1)} = E[T_0] = \exp(0.7 + 1.2^2/2)$.

Assume that the production parameters in Eqs. 6A and 6B are $\alpha = 0.4$ and $\beta = 0.6$, with

$$\lambda(K) = 0.5K^{-0.4} \quad \text{and} \quad \mu^{(1)}(L) = \exp(0.7 + 1.2^2/2) \cdot L^{-0.6}$$

Let the downtime cost function be a linear model: $C(T) = c \cdot T$ with $c = 1$. According to Proposition 1, from Eq. 8B, the optimal level of resource is

$$M^* = (\alpha + \beta) \left(\alpha^{-\alpha} \beta^{-\beta} \lambda_0 c \mu_0^{(1)} \right)^{\frac{1}{\alpha+\beta+1}} = 2.01.$$

From Eq. 8A we have $\frac{K^*}{L^*} = \frac{\alpha}{\beta} = \frac{0.4}{0.6}$, with $K^* = 0.804$ and $L^* = 1.206$. From Eq. 8C, the minimal Expected Total Cost is $E[\text{Total Cost}] = 2M^* = 4.03$. Figure 1 illustrates how resource allocation $\{K, L\}$ affect the firm's total cost (the sum of resources allocated and expected aggregate downtime cost), where an optimal pair $\{K^*, L^*\}$ exists.

Figure 2 illustrates that, as expected, the premium for insuring against downtime cost (with a waiting time deductible $d = 8$) decreases as the allocated resources $\{K, L\}$ increases.

Fig. 1 Optimization of Production of Resource Allocation

Remark 8 The total optimal level of spending M^* remains the same whether it is used for producing uptime or purchasing insurance. This follows from the fact that we assume an actuarially fair premium and enterprises without risk-aversion.

Example 2 As a slight variation from Example 1, instead of using a linear cost function, we now consider a quadratic cost function $C(T) = c \cdot T^2$ with $c = 0.05727$ such that $E[C(T)]$ remains the same as in Example 1. From Eq. 9B, the new optimal level of resource allocation is

$$M^* = (\alpha + 2\beta) \left(\alpha^{-\alpha} (2\beta)^{-2\beta} \lambda_0 c \mu_0^{(2)} \right)^{\frac{1}{\alpha+2\beta+1}} = 2.24.$$

From Eq. 9A we have $\frac{K^*}{L^*} = \frac{\alpha}{2\beta} = \frac{0.4}{1.2}$ and thus $K^* = 0.56$ and $L^* = 1.68$. Thus, relatively more resources are allocated to reducing duration of downtime as compared to Example 1.

5.3 Insurance with Aggregate Limits

Recall that the firm's aggregate downtime cost S in Eq. 5A in a given year is a random sum of independent random cost

Fig. 2 Calculated Insurance Premium under various resource allocations

variables. It is common for an insurance contract to specify not only a deductible d (waiting time) per occurrence but also impose a cap (annual aggregate limit) on the total insurance payment in a given policy year. Our frequency-duration model can facilitate calculation of the probability distribution of the aggregate downtime cost for a firm, and thus enables calculating the cost of insurance (with deductible and aggregate limit).

For the firm, assume a Poisson frequency of breakdown with mean λ and the following tabular *per occurrence* downtime cost (after applying waiting time deductible):

Downtime cost x	0	1	2	3	...
Probability $f(x)$	$f(0)$	$f(1)$	$f(2)$	$f(3)$	...

The aggregate downtime cost S in Eq. 5A, before applying aggregate limit, has a compound Poisson distribution which can be computed using the Panjer recursion (Panjer 1981):

$$g(x) = \Pr\{S = x\} = \sum_{y=1}^x \frac{\lambda}{x} \cdot y \cdot f(y) \cdot g(x-y), \quad x = 1, 2, \dots \quad (10)$$

with a starting value $g(0) = \exp(-(1 - f(0))\lambda)$. Once the aggregate loss distribution is computed, the aggregate limit can be applied to the aggregate loss distribution for computing the insurance cost.

6 Model of downtime cost for enterprises in a supply chain

As in Wang (2017), we consider a supply chain ecosystem of m firms, indexed by $j = 1, 2, \dots, m$, which are interconnected through business relations (vendors, contractors, suppliers, service providers, etc.). The execution of a business process at firm j depends on available IT services from a set of other firms, i.e. an outage at any one of these vendors can potentially halt the business process of firm j .

We label a “tag” j to the notations to refer to a quantity that is specific to firm j . For firm j , we let N_j represent the random number of IT service breakdowns in a year, and T_j represent the random duration of breakdown given one occurrence. Assume that N_j have a Poisson distribution with mean λ_j . Let T_j have raw moments $\mu_j^{(k)} = E[T_j^k]$ for $k = 1, 2, \dots$

In a supply chain setting, IT service downtime in one firm can propagate through the supply chain and cause business disruption to another firm. Let the parameter $\theta_{i,j} \leq 1$ represent the propagation coefficient, i.e., the likelihood that a downtime of firm i 's IT service will affect another firm j 's business operation. Note the directional propagation:

- 1) Outbound: Firm j exports downtime to another firm i with propagation coefficient $\theta_{j,i}$
- 2) Inbound: Firm j imports downtime from another firm i , with propagation coefficient $\theta_{i,j}$

We get a matrix of propagation coefficients with diagonal values $\theta_{j,j} = 1$ but not necessarily symmetric:

$$\begin{pmatrix} 1 & \theta_{1,2} & \dots & \theta_{1,m} \\ \theta_{2,1} & 1 & \dots & \theta_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{m,1} & \theta_{m,2} & \dots & 1 \end{pmatrix} \quad (11)$$

Remark 9 The propagation matrix may be relatively sparse, given that flowchart like IT service architectures are accurate. However, it may be less sparse than expected, as is sometimes discovered when consequences of outages reveal unexpected dependencies, see for example Swedish Civil Contingencies Agency et al. (MSB 2014). Böhme and Schwartz (2010) note that fully-connected graphs allow for modelling of network externalities.

When firm j allocates resources $\{K_j, L_j\}$ in IT services maintenance, the corresponding number of breakdowns for firm j is a random variable $N_j(K_j)$ with a Poisson mean $\lambda_j(K_j)$, and the downtime for each breakdown occurrence, $T_j(L_j)$, has a mean $\mu_j^{(1)}(L_j)$.

Let $K = (K_1, K_2, \dots, K_m)$ and $L = (L_1, L_2, \dots, L_m)$ be the vectors of resource allocations by the m firms. We consider only 1-step propagation from the origin firm, and ignore further steps of propagation. For simplicity, we delimit ourselves to the 1-step case only.

6.1 Independent sources of breakdown

The frequency of breakdown for a firm is the sum of direct breakdown and indirect breakdown, both of which follow Poisson arrivals:

- 1) direct breakdown N_j follows a Poisson arrival process with intensity $\lambda_j(K_j)$
- 2) indirect breakdown from firm i follows independent Poisson arrival process with intensity $\lambda_i(K_i)$

The combined frequency of breakdown due to direct and indirect causes follows a Poisson distribution with the mean $\sum_{i=1}^m \theta_{i,j} \cdot \lambda_i(K_i)$.

Remark 10 Though the model does not allow for statistically dependent breakdowns, it can nevertheless capture operational dependencies such as multiple firms being dependent on a single cloud service provider n by setting the corresponding $\theta_{n,j}$ parameters close to unity for all j .

For simplicity, we assume a linear model $C_j(T_j) = c_j \cdot T_j$ for firm j 's cost of downtime. Though not perfect, it can be considered a reasonable first approximation of a more advanced model.

The expected cost to firm j from all possible breakdowns in the supply chain is

$$E[\text{Downtime Cost}_{all \rightarrow j}] = \sum_{i=1}^m \theta_{i,j} \cdot \lambda_i(K_i) \cdot c_j \cdot \mu^{(1)}(L_i) \quad (12A)$$

Note the indices: Cost coefficients c_j belong to firm j , whose cost we are assessing, but the number and durations of outages are summed over all the firms in the ecosystem.

Remark 11 As noted by Franke (2017), different insurance companies take different approaches to which parts of Eq. 12A they cover. One of the companies investigated offers to cover outages at all external service providers with an increase in the premium of some 20–25% and the indemnity limit is cut in half. This corresponds to breaking (12A) into two terms, one internal and one external:

$$E[\text{Downtime Cost}_{all \rightarrow j}] = \underbrace{\lambda_j(K_j) \cdot c_j \cdot \mu_j^{(1)}(L_j)}_{\text{internal}} + \underbrace{\sum_{i \neq j}^m \theta_{i,j} \cdot \lambda_i(K_i) \cdot c_j \cdot \mu_i^{(1)}(L_i)}_{\text{external}} \quad (12B)$$

The increase in premium and decrease of indemnity limit reflects the insurer's assessment of the relative magnitudes of the first (internal) and the second (external) terms in Eq. 12B. Alternatively, the same insurer offers coverage of outages at a specific list of some 3–5 named providers. This corresponds to breaking (12A) into two different terms:

$$E[\text{Downtime Cost}_{all \rightarrow j}] = \sum_{i \in \text{Covered}} \theta_{i,j} \lambda_i(K_i) c_j \mu_i^{(1)}(L_i) + \sum_{i \notin \text{Covered}} \theta_{i,j} \lambda_i(K_i) c_j \mu_i^{(1)}(L_i) \quad (12C)$$

The index j , of course, is in the Covered set. Other insurers investigated by Franke (2017), make no distinction between internal outages and outages at external service providers, i.e. they cover all of Eq. 12A. Indeed, one of the insurers interviewed remarks that most companies are actually better off from the business continuity perspective by trusting service providers like Microsoft, Google, or Amazon rather than trying to build equally reliable in-house IT service.

A firm's own IT services breakdown causes a direct business interruption cost, as well as spillover effect (business interruption to other firms) in a supply chain, with a total expected downtime cost:

$$E[\text{Downtime Cost}_{j \rightarrow all}] = \lambda_j(K_j) \cdot \left(\sum_{i=1}^m \theta_{j,i} \cdot c_i \cdot \mu_j^{(1)}(L_j) \right) \quad (13A)$$

$$E[\text{Downtime Cost}_{j \rightarrow all}] = \underbrace{\lambda_j(K_j) \cdot c_j \cdot \mu_j^{(1)}(L_j)}_{\text{first-party cost}} + \underbrace{\lambda_j(K_j) \sum_{i \neq j} \theta_{j,i} \cdot (c_i \cdot \mu_j^{(1)}(L_j))}_{\text{third-party liability}} \quad (13B)$$

In Eq. 13B, the second term represents, equivalently, (i) the expected externality cost imposed on the others in the supply-chain by the breakdown of firm j in the absence of liability, or (ii) the expected liability that has to be paid in the presence of such liability, or (iii) the actuarially fair cost of a third-party liability insurance which covers the other firms' costs of the downtime caused.

Proposition 4 Everything else equal, to minimize the total expected downtime cost (12A), the optimal investment K_j for firm j in reducing the breakdown frequency exceeds the optimal investment for the same firm on a standalone basis.

Remark 12 However, firm j does not have any incentives to purchase insurance to cover third parties, unless it is required to do so by external pressure.

It is illuminating to rewrite (12B) in yet another form:

$$E[\text{Downtime Cost}_{all \rightarrow j}] = \Theta_j \cdot \left(\lambda_j(K_j) \cdot c_j \cdot \mu_j^{(1)}(L_j) \right), \quad \text{where the term} \quad \Theta_j = 1 + \frac{\left(\sum_{i \neq j}^m \theta_{i,j} \cdot \lambda_i(K_i) \cdot \mu_i^{(1)}(L_i) \right)}{\lambda_j(K_j) \cdot \mu_j^{(1)}(L_j)} > 1 \quad (14A)$$

is a supply chain dependency multiplier (analogous to the concept of multiplier used by Dynes et al. (2007)). In the special case that all firms in the supply chain have independent and identically distributed downtime frequency N_j and duration T_j , Eq. 14A can be simplified to

$$\Theta_j = 1 + \sum_{i \neq j}^m \theta_{i,j} \quad (14B)$$

If firm j has no expected cost from outages in the rest of the supply chain, $\Theta_j = 1$, i.e. there is no multiplier effect.

However, Θ_j is typically greater than one and the expected cost to firm j is directly proportional to the multiplier Θ_j .

Assuming independent drivers for the origins of breakdowns among firms, due to the propagation of downtime in a supply chain, the cost to all firms in the supply chain from all possible breakdowns by all firms in the supply chain is

$$E[\text{Downtime Cost}_{all \rightarrow all}] = \sum_{j=1}^m \sum_{i=1}^m \theta_{i,j} \cdot \lambda_i(K_i) \cdot (c_j \cdot \mu_i^{(1)}(L_i)) \quad (15)$$

Equation 15 can be used to calculate risk aggregation for an entire supply chain. An insurer may cover many firms in the same supply chain, and an insurer may purchase reinsurance protection for its own insurance portfolio as a way of diversifying concentration risk. Sometimes capital market solutions can be used to transfer the aggregated supply chain risk to investors.

Example 3 Consider four firms ($j = 1, 2, 3, 4$) which all depend upon one IT Service Provider firm 5 (see Fig. 3). Assume that the propagation coefficients are $\theta_{5,j} = 0.8$ but $\theta_{j,5} = 0$, for $j = 1, 2, 3, 4$. The propagation coefficients matrix thus looks as follows:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0.8 & 0.8 & 0.8 & 0.8 & 1 \end{pmatrix}$$

- a) Effects on the four firms: For each of the four firms ($j = 1, 2, 3, 4$), according to Eq. 12A, we have

$$E[\text{Downtime Cost}_{all \rightarrow j}] = c_j [\lambda_j(K_j) \cdot \mu^{(1)}(L_j) + 0.8\lambda_5(K_5) \cdot \mu^{(1)}(L_5)]$$

Thus, for each of the four firms ($j = 1, 2, 3, 4$), the downtime experienced is a sum of one term that can be controlled by investments K_j and L_j and one

Fig. 3 A simple supply chain of four firms and their common IT Service Provider

term beyond the control of firm j . While each firm j can optimize balance between K_j and L_j according to Proposition 1, they cannot do anything about the second term, except try to keep c_j as small as possible.

- b) Effects on the IT Service Provider: According to Eq. 12B, we have

$$E[\text{Downtime Cost}_{5 \rightarrow all}] = \lambda_5(K_5) \cdot c_5 \cdot \mu^{(1)}(L_5) + \sum_{j=1}^4 c_j \cdot 0.8 \cdot \lambda_5(K_5) \cdot \mu^{(1)}(L_5)$$

In the absence of 3rd party liability, the IT Service Provider carries only its first-party cost, the first term, whereas the second term represents externalities carried by the firms. Thus, there is little incentive to reduce downtime as much as would be socially optimal. Even if the four firms ($j = 1, 2, 3, 4$) take out insurance policies covering outages at their provider, incentives will still be misaligned and their actuarially fair premiums would cost them as much as their part of the expected outage cost. If, on the other hand, the IT Service Provider (firm 5) were liable for the downtime costs it causes the firms, it would have incentives to increase its investment $K_5 + L_5$ to the socially optimal level.

- c) Insurance on the ecosystem: It can be shown that the total loss to the ecosystem can be optimized by minimizing (15). An insurer is generally concerned about risk accumulation in its portfolios, as it may cover multiple firms in the same supply chain, and thus may be liable for much of the total downtime cost within a supply chain, including both first-party and third-party insurance. An insurer (or a group of insurers and reinsurers) that insures the four, or even five, firms ($j = 1, 2, 3, 4, 5$) thus has an incentive to control the total supply chain downtime cost (15) by encouraging investments at those insured customer firms where it is the most efficient to do so. Wang (2019) proposed an integrated approach to information security investment and cyber insurance, with more focus on risk advisory services and partnership. Similar mechanisms could be proposed for managing downtime costs in a supply chain.

7 Characterization of insurance policies of downtime cost

We note that the models introduced in the previous sections describe existing cyber insurance offerings on the market quite well. Such offerings can be broadly described as

comprising of first and third-party insurance coverage (e.g., Gordon et al. 2003; Biener et al. 2015; Franke 2017).

First-party insurance coverage for the direct costs of a firm's IT service outages are modeled in Example 1. Simple assumptions about the production of security give insight into the pricing of these products. The use of deductibles by insurers to discourage under-investment from the insured are readily understood from this framework, and can be incorporated into the model in a way that resembles call options. Based on the supply chain model, where each firm is both an importer and an exporter of downtime, the demand for first-party coverage of costs caused by other firms can be understood. If these firms were strictly liable for these costs, there would be no externalities. However, as noted by Franke (2017, 2018), many large IT service providers instead have standard SLAs that severely limit their liability. Cyber insurance (covering first-party downtime loss) can then be used to cover the residual gap between the full loss of the firm (client) and whatever liability the service provider accepts.

Third-party liability coverage can also be understood based on the supply chain model. Given some liability (full or partial), a firm has an interest to insure itself against the claims of other firms, affected by outages originating at the first firm. Since these other firms can be much larger in terms of turnover and profit than the origin firm, and their claims can be similarly larger, such insurance can be a prudent measure. Again, insurers have an incentive to adjust the premiums offered based on the security investments of the insured. Furthermore, the insurer has an incentive to reward investments that are in line with the loss function of the potentially affected third parties: if most third parties have large snowball effects, the insurer would encourage investments to shorten the duration of outages. Additionally, insurers might encourage customers to engage in SLA negotiations with others in the supply chain in order to learn more about their preferences, particularly their downtime cost functions.

Risk Aggregation Across the Supply Chain: As noted in conjunction with Example 3, and insurer is generally concerned about risk accumulation in its portfolios, and could thus be instrumental in aligning incentives and securing investment where it is most needed. Wang (2019) proposed an integrated approach to information security investment and cyber insurance. Based on the insights from the model, an improved cyber insurance design is proposed with more focus on risk advisory services and partnership. Franke and Draeger (2019) proposed a similar scheme, where insurers could act to enable collective action on behalf of their insureds facilitating collective funding of additional IT service incident managers in the face of limited capacity.

8 Conclusions and outlook

This paper presents a probability model (Poisson arrival frequency with lognormal downtime duration) for an enterprise, and a propagation model for the supply chain, building on literature from multiple disciplines including IT service, supply chain, actuarial science, insurance and economics of information security. We used the model to study interactions of a firm's security investment (resource allocation) and externality of spillover effects, as well as implications in first-party versus third-party insurance.

8.1 Single-firm case

We show how total enterprise resources in the single firm-case can be optimally allocated depending on the relative effectiveness of reducing (i) the frequency of outages and (ii) the duration of outages. When the cost of downtime increases more than linearly with time, it is optimal to allocate more resources to reduce the duration of outages. Conversely, When the cost of downtime increases slowly, or not at all, with time, it is optimal to allocate more resources to reduce the frequency of outages. Though the results are theoretical, they still offer useful rules of thumb for management practice.

8.2 Interconnected firms case

The supply chain-model with several connected firms offers interesting conceptual insights. We highlight three:

First, the *supply chain dependency multiplier* Θ_j in Eqs. 14A and 14B is a measure of how much the interdependence in the supply chain network increases the downtime costs of a firm over and above those caused by internal factors. Conceptually, it points to a way of reducing the expected cost to firm j by reducing the correlation parameters $\theta_{i,j}$ (including bringing some down to 0, via cutting the dependency entirely). In a practical setting, one way of reducing dependency is to build a kind of "supply chain redundancy" across firms. This is non-trivial, and quite different from redundancy in a firm's own internal operation. For example, while everyone has redundant hard drives in data warehouses, and it might be feasible to reroute a credit card transaction to a redundant payment system, it is significantly more cumbersome to maintain dual independent but fully synchronized inventory systems that keep track of physical products in a storage area somewhere at an off-shore location. Furthermore, in some common cases like IT services that depend on Google accounts for authentication of users, it is virtually impossible to maintain a redundant alternative authentication system. However, this is not necessarily always the case. Fostering more

redundancy at the supply chain – or ecosystem – level will require collective investment and coordination.

Second, though the correlation parameters $\theta_{i,j}$ are easy to define, in practice it can be difficult to attribute the causes of downtime in practice as it propagates through the supply chain. This is true for several reasons. First of all, in the case of antagonistically caused downtime, all the usual difficulties in attribution apply (see e.g. Nicholson et al. (2013), Rid and Buchanan (2015) for different perspectives on this). However, there is a second complication specific to downtime: since an outage is not a discrete event, its actual duration can have multiple causes. One root cause (e.g. a cooling system breaking down) might have brought the IT service down in the first place (as servers in one place overheat), but another cause (misconfiguration in the alternative site) might prevent it from coming back online, thus prolonging the outage. Without going into more details, we note that conceptually, the difficulty of attribution might make it easier to offer more extensive first-party insurance (which does not require attribution) rather than third-party liability insurance (which does require attribution). As such, this insight helps us understand why existing cyber insurance policies are a mix of first and third-party coverage. Different combinations will be necessary in different situations and will have to be customized to suit different industries and firms.

Third, the parameters $\theta_{i,j}$ are hard to estimate statistically before a major incidence of IT service breakdown. Firms may not even realize that there are dependencies until it is too late (see for example Swedish Civil Contingencies Agency et al. MSB 2014). Thus, similar to penetration tests to counter antagonistic cyber security threats, there is a need for more “fire drill” type testing, to document consequences and learn how to increase resilience. This, however, is difficult for a number of reasons, the most obvious being that systems in production cannot be easily used for testing purposes, while development systems designed for testing purposes typically lack precisely the kind of third-party dependencies such exercises would seek to discover and mitigate. Nevertheless, the consequences of not conducting such drills beforehand might be even costlier.

8.3 Managerial implications

IT service downtime is a problem faced by all enterprises today. With increasing dependency on IT services comes increasing sensitivity to outages. However, all firms are not IT service dependent in the same way, so it is important for managers to find the measures most appropriate for their operations.

For any firm allocating finite resources in managing IT downtime risk, it is important to look at the overall picture, by evaluating the relative effectiveness of reducing

the frequency of downtime occurrence ($\lambda(K)$) or downtime duration ($\mu^{(1)}(L)$), as well as how the business cost of downtime increases with the duration ($c(T)$). Even though it can be difficult in practice to determine exact numbers, the exercise of trying to do so can be rewarding. In particular, it is valuable to try and estimate whether downtime cost increases more than linearly with the duration of downtime, in which case efforts should aim to minimize *duration*, or if downtime cost is mostly fixed response costs, in which case efforts should aim to minimize *frequency of occurrence*.

However, in an increasingly inter-connected world, a firm not only needs to look at its own IT system, but also at the propagation of downtime in its complex supply-chain. As noted above, it can be challenging to try and map out dependencies and estimate correlation parameters $\theta_{i,j}$. Nevertheless, there might also some low-hanging fruits. One recent case study indicates that some firms building novel 5G-enabled services around would benefit from formalizing their service level agreements with each other, thus enabling a more mature risk sharing in the ecosystem (Olsson and Franke 2019). Another recent study, albeit a small one, indicates that although companies are willing to share IT vulnerability information with each other, proactive sharing of vulnerability information is relatively rare, and customers do not require such information from their providers (Olsson et al. 2019). Such passivity about information sharing stands in stark contrast to the recommendations about active disclosure and sharing found in contemporary cyber security guidelines. As these examples suggest, there are relatively straightforward actions that could be taken to mitigate IT service supply-chain risk.

For insurers, the results highlight both the need for active management of portfolio accumulation risk the opportunity to re-innovate business models. Starting with accumulation risk, insurers should work to improve current practices for mapping dependencies among their insureds. Whereas much underwriting today is based on (self-assessment) forms being filled out and interviews conducted (Woods et al. 2017), it is fully conceivable for insurers today to proactively and continuously scan their customers to produce an up-to-date situational awareness of their vulnerabilities and inter-dependencies. Though this comes at a cost, this cost-benefit analysis is worth conducting. This leads to the business opportunity: insurers are in a position where they could take the lead in not only pricing cyber risk, but more proactively working to reduce it. Possible examples already mentioned include facilitating collective action on behalf of customers, thus enabling beneficial security investments that not otherwise be realized (Wang 2019; Franke and Draeger 2019). However, the future role of insurance in cyber security governance is by no means clear, and it is up to insurance managers to show leadership

in finding novel ways to reduce systemic risks (Woods and Moore 2019).

8.4 Future work

The results also suggest some avenues for future work. First, it would be interesting to further explore the potential impact of the complexity and structure of the supply chain. While in general, the sum of *independent* Poisson processes (from various nodes in a supply-chain network) produces another Poisson process, if the Poisson processes are *correlated*, the combined effect would be more complicated.

Another avenue of future research is more empirical studies of the frequency and duration of IT service breakdown (testing the Poisson frequency model and the lognormal model for downtime duration), as well as empirical tests of the effectiveness of security spending in reducing the frequency and duration of IT service breakdown. Such future work would require well designed field studies and gathering of survey data.

Acknowledgments This research is in part supported by the Cyber Risk Management Project (CyRiM), sponsored by the Monetary Authority of Singapore, the Cyber Security Agency of Singapore, Aon, Lloyd's, MSIG, SCOR, and TransRe. U. Franke is partially supported by the Swedish Civil Contingencies Agency, MSB, agreement no. 2015-6986.

Funding Information Open access funding provided by RISE Research Institutes of Sweden.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Proofs of propositions

Proof (Proposition 1)

- a) Let M be a fixed budget and $L = M - K$. From Eq. 7A we have

$$E[\text{Total Cost}] = M + \left(\lambda_0 c \mu_0^{(1)} \right) \left\{ K^{-\alpha} (M - K)^{-\beta} \right\}$$

To minimize the Expected Total Cost we get an optimal solution $K^* = \frac{\alpha}{\alpha+\beta} M$ and $L^* = \frac{\beta}{\alpha+\beta} M$. Express the Expected Total Cost in terms of M :

$$E[\text{Total Cost}] = M + \left(\lambda_0 c \mu_0^{(1)} \right) \left\{ \left(\frac{\alpha}{\alpha+\beta} M \right)^{-\alpha} \left(\frac{\beta}{\alpha+\beta} M \right)^{-\beta} \right\}$$

From this equation we can derive formula (8B). Plugging (8A) and (8B) into the Expected Total Cost we obtain (8C).

- b) Follow steps of a), using 2β to replace β , c^2 to replace c , and $\mu_0^{(2)}$ to replace $\mu_0^{(1)}$. $\square$

Proposition 2

- a) Follow the steps of proof of proposition 1, using 2β to replace β , and $\mu_0^{(2)}$ to replace $\mu_0^{(1)}$
 b) Follow the steps of proof of proposition 1, using 4β to replace 2β , and $\mu_0^{(4)}$ to replace $\mu_0^{(2)}$. $\square$

Proposition 3 This can be derived from Eq. 7A by letting $E[C(T)] = c$. $\square$

References

- Aceto G, Botta A, Marchetta P, Persico V, Pescapé A (2018) A comprehensive survey on internet outages. *J Netw Comput Appl* 113:36–63
- Ahmed M, Hossain MA (2014) Cloud computing and security issues in the cloud. *Int J Netw Secur Appl* 6(1):25
- Ali M, Khan SU, Vasilakos AV (2015) Security in cloud computing: Opportunities and challenges. *Inform Sci* 305:357–383
- Anderson R, Moore T (2006) The economics of information security. *Science* 314(5799):610–613. <https://doi.org/10.1126/science.1130992>
- Bahnemann D (2015) Distributions for Actuaries. CAS Monograph Series, Casualty Actuarial Society. <https://www.casact.org/pubs/monographs/papers/02-Bahnemann.pdf>
- Bharadwaj A, Keil M, Mähring M (2009) Effects of information technology failures on the market value of firms. *J Strategic Inf Syst* 18(2):66–79. <https://doi.org/10.1016/j.jsis.2009.04.001>
- Biener C, Eling M, Wirfs JH (2015) Insurability of cyber risk: an empirical analysis. *Geneva Paper Risk Insur Issues Pract* 40(1):131–158. <https://doi.org/10.1057/gpp.2014.19>
- Böhme R, Kataria G (2006) Models and measures for correlation in cyber-insurance. In: Workshop on Economics of Information Security – WEIS
- Böhme R, Schwartz G (2010) Modeling Cyber-Insurance: Towards a Unifying Framework. In: Workshop on Economics of Information Security – WEIS
- Boritz E, Mackler E (1999) Reporting on systems reliability. *J Account* 188(5):75–87
- Camp LJ, Wolfram CD (2004) Pricing security: Vulnerabilities as externalities Available at SSRN: <https://ssrn.com/abstract=894966>
- Cobb CW, Douglas PH (1928) A theory of production. *Amer Econ Rev* 18(1):139–165. <http://www.jstor.org/stable/1811556>

- Dai YS, Yang B, Dongarra J, Zhang G (2009) Cloud service reliability: Modeling and analysis. In: 15th IEEE Pacific Rim International Symposium on Dependable Computing. Citeseer, pp 1–17
- Dynes S, Brechbuhl H, Johnson ME (2005) Information security in the extended enterprise: Some initial results from a field study of an industrial firm. In: Workshop on Economics of Information Security (WEIS)
- Dynes S, Johnson ME, Andrijeic E, Horowitz B (2007) Economic costs of firm-level information infrastructure failures: Estimates from field studies in manufacturing supply chains. *Int J Logist Manag* 18(3):420–442
- Franke U (2012) Optimal IT Service Availability: Shorter Outages, or Fewer? *IEEE Trans Netw Serv Manag* 9(1):22–33
- Franke U (2014) Enterprise architecture analysis with production functions. In: 2014 IEEE 18th International Enterprise Distributed Object Computing Conference (EDOC). IEEE, pp 52–60
- Franke U, Holm H, König J (2014). The distribution of time to recovery of enterprise IT services. *IEEE Trans Reliab* 63(4):858–867. <https://doi.org/10.1109/TR.2014.2336051>
- Franke U, Buschle M (2016) Experimental evidence on decision-making in availability service level agreements. *IEEE Trans Netw Serv Manag* 13(1):58–70. <https://doi.org/10.1109/TNSM.2015.2510080>
- Franke U (2017) The cyber insurance market in Sweden. *Comput Secur* 68:130–144. <https://doi.org/10.1016/j.cose.2017.04.010>
- Franke U, Katsikas SK, Alcaraz C (2018) Cyber insurance against electronic payment service outages. In: 14th International Workshop on Security and Trust Management. Springer International Publishing, Cham, pp 73–84. https://doi.org/10.1007/978-3-030-01141-3_5
- Franke U, Draeger J (2019) Two simple models of business interruption accumulation risk in cyber insurance. In: 2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA). <https://doi.org/10.1109/CyberSA.2019.8899678>
- Gordon L, Loeb M (2002) The economics of information security investment. *ACM Trans Inf Syst Secur (TISSEC)* 5(4):438–457
- Gordon LA, Loeb MP, Sohail T (2003) A framework for using insurance for cyber-risk management. *Commun ACM* 46(3):81–85
- Haenggi M (2012) Stochastic geometry for wireless networks. Cambridge University Press, Cambridge
- Herley C (2009) So long, and no thanks for the externalities: the rational rejection of security advice by users. In: Proceedings of the 2009 workshop on New security paradigms workshop. ACM, pp 133–144
- IBM Global Services (1998) Improving systems availability. Technical report, IBM Global Services
- Jenjarrussakul B, Tanaka H, Matsuura K (2013) Sectoral and Regional Interdependency of Japanese Firms Under the Influence of Information Security Risks. In: The Economics of Information Security and Privacy. Springer, pp 115–134
- Jeske DR, Xuemei Zhang PL (2005) Adjusting software failure rates that are estimated from test data. *IEEE Trans Reliab* 54(1):107–114. <https://doi.org/10.1109/TR.2004.842531>
- Kingman J (1992) Poisson Processes. Clarendon Press
- Klugman SA, Panjer HH, Willmot GE (2012) Loss models: from data to decisions, vol 715. Wiley, New York. <https://doi.org/10.1002/9780470391341>
- Lloyd's (2018) Cloud down. impacts on the us economy. Technical report, Lloyd's and AIR. <https://www.lloyds.com/news-and-risk-insight/risk-reports/lib-rary/technology/cloud-down>, accessed on February 23, 2018
- Manika D, Papagiannidis S, Bourlakis M (2015) Can a ceo's youtube apology following a service failure win customers' hearts? *Technol Forecast Soc Chang* 95:87–95. <https://doi.org/10.1016/j.techfore.2013.12.021>
- Martinello M, Kaâniche M, Kanoun K (2005) Web service availability—impact of error recovery and traffic model. *Reliab Eng Syst Saf* 89(1):6–16. <https://doi.org/10.1016/j.res.2004.08.003>, safety, Reliability and Security of Industrial Computer Systems
- Matsuura K (2009) Productivity space of information security in an extension of the Gordon-Loeb's Investment Model. In: Managing Information Risk and the Economics of security. Springer, pp 99–119
- Meland PH, Tøndel IA, Moe M, Seehusen F (2017) Facing uncertainty in cyber insurance policies. In: International Workshop on Security and Trust Management. Springer, pp 89–100
- MSB (2014) International case report on cyber security incidents – Reflections on three cyber incidents in the Netherlands, Germany and Sweden. Swedish Civil Contingencies Agency (MSB), National Cyber Security Centre (NCSC), the Netherlands, and Bundesamt für Sicherheit in der Informationstechnik, Germany, publication number MSB782
- Muermann A, Kunreuther H (2008) Self-protection and insurance with interdependencies. *J Risk Uncert* 36(2):103–123
- Närman P, Franke U, König J, Buschle M, Ekstedt M (2014) Enterprise architecture availability analysis using fault trees and stakeholder interviews. *Enterprise Inf Syst* 8(1):1–25
- Nicholson A, Janicke H, Watson T (2013) An initial investigation into attribution in SCADA systems. In: 1st International Symposium for ICS & SCADA Cyber Security Research (ICS-CSR)
- OECD (2017) Enhancing the Role of Insurance in Cyber Risk Management. <https://doi.org/10.1787/9789264282148-en>
- Olsson T, Franke U (2019) Risks and Assets: A Qualitative Study of a Software Ecosystem in the Mining Industry. In: Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, ESEC/FSE 2019, pp 895–904. <https://doi.org/10.1145/3338906.3340443>
- Olsson T, Hell M, Höst M, Franke U, Borg M (2019) Sharing of vulnerability information among companies – a survey of Swedish companies. In: 45th EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA), in press
- Oppenheimer D, Ganapathi A, Patterson DA (2003) Why do internet services fail, and what can be done about it? In: USENIX Symposium on internet technologies and systems, Seattle, vol 67
- Panjer HH (1981) Recursive evaluation of a family of compound distributions. *ASTIN Bullet J IAA* 12(1):22–26
- Patterson D (2002) A simple way to estimate the cost of downtime. In: Proc. 16th Systems Administration Conf.—LISA, pp 185–8
- Ponemon (2016) 2016 Cost of data center outages. Technical report, Ponemon Institute and Emerson Network Power
- Rapoza J (2014) Preventing virtual application downtime. Technical report, Aberdeen Group
- Rid T, Buchanan B (2015) Attributing cyber attacks. *J Strateg Stud* 38(1-2):4–37
- Ross SM (1996) Stochastic processes, pp 59–60
- Schroeder B, Gibson G (2010) A large-scale study of failures in high-performance computing systems. *IEEE Trans Depend Sec Comput* 7(4):337–350
- Shahrad M, Wentzlaff D (2016) Availability knob: Flexible user-defined availability in the cloud. In: Proceedings of the Seventh ACM Symposium on Cloud Computing. ACM, pp 42–56
- Stefansson G (2002) Business-to-business data sharing: a source for integration of supply chains. *Int J Prod Econ* 75(1-2):135–146

- Subashini S, Kavitha V (2011) A survey on security issues in service delivery models of cloud computing. *J Netw Comput Appl* 34(1):1–11
- Swafford PM, Ghosh S, Murthy N (2008) Achieving supply chain agility through IT integration and flexibility. *Int J Prod Econ* 116(2):288–297
- Takabi H, Joshi JB, Ahn GJ (2010) Security and privacy challenges in cloud computing environments. *IEEE Security & Privacy* 8(6):24–31. <https://doi.org/10.1109/MSP.2010.186>
- Taylor Z, Ranganathan S (2014) Designing High Availability Systems: DFSS and Classical Reliability Techniques with Practical Real Life Examples. IEEE. <https://doi.org/10.1002/9781118739853>
- Thomas RC, Antkiewicz M, Florer P, Widup S, Woodyard M (2013) How bad is it?—a branching activity model to estimate the impact of information security breaches. A Branching Activity Model to Estimate the Impact of Information Security Breaches
- Vecchio D (2016) How to Derive Business Value From DevOps. Technical report, Gartner, Inc., g00317166
- Wang S (2017) Knowledge Set of Attack Surface and Cybersecurity Rating for Firms in a Supply Chain. Available at SSRN 3064533. <https://doi.org/10.2139/ssrn.3064533>
- Wang SS (2019) Integrated framework for information security investment and cyber insurance. *Pac Basin Financ J* 101173:57. <https://doi.org/10.1016/j.pacfin.2019.101173>
- Wood T, Cecchet E, Ramakrishnan KK, Shenoy PJ, van der Merwe JE, Venkataramani A (2010) Disaster recovery as a cloud service: Economic benefits & deployment challenges. *HotCloud* 10:8–15
- Woods D, Agrafiotis I, Nurse JR, Creese S (2017) Mapping the coverage of security controls in cyber insurance proposal forms. *J Internet Serv Appl* 8(1):8
- Woods D, Moore T (2019) Does insurance have a future in governing cybersecurity? *IEEE Security and Privacy Magazine*
- Yamada S (2014) Software reliability modeling: fundamentals and applications, vol 5. Springer. <https://doi.org/10.1007/978-4-431-54565-1>
- Zhao X, Xue L, Whinston AB (2013) Managing interdependent information security risks: cyberinsurance, managed security services, and risk pooling arrangements. *J Manag Inf Syst* 30(1):123–152

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.