Math. Model. Nat. Phenom. 17 (2022) 40
https://doi.org/10.1051/mmnp/2022041

Mathematical Modelling of Natural Phenomena
www.mmnp-journal.org

EXACT INSURANCE PREMIUMS FOR CYBER RISK OF SMALL AND MEDIUM-SIZED ENTERPRISES

Stefano Chiaradonna* and Nicolas Lanchier

Abstract. As cyber attacks have become more frequent, cyber insurance premiums have increased, resulting in the need for better modeling of cyber risk. Toward this direction, Jevtić and Lanchier [*Insur. Math. Econ.* **91** (2020) 209–223] proposed a dynamic structural model of aggregate loss distribution for cyber risk of small and medium-sized enterprises under the assumption of a tree-based local-area-network topology that consists of the combination of a Poisson process, homogeneous random trees, bond percolation processes, and cost topology. Their model assumes that the contagion spreads through the edges of the network with the same fixed probability in both directions, thus overlooking a dynamic cyber security environment implemented in most networks, and their results give an exact expression for the mean of the aggregate loss but only a rough upper bound for the variance. In this paper, we consider a bidirectional version of their percolation model in which the contagion spreads through the edges of the network with a certain probability of moving toward the lower level assets of the network but with another probability of moving toward the higher level assets of the network, which results in a more realistic cyber security environment. In addition, our mathematical approach is quite different and leads to exact expressions for both the mean and the variance of the aggregate loss, and therefore an exact expression for the insurance premiums.

Mathematics Subject Classification. 60K35, 91G05.

Received February 21, 2022. Accepted September 1, 2022.

1. Introduction

1.1. Cyber risk

According to The Institute of Risk Management, *cyber risk* is “any risk of financial loss, disruption or damage to the reputation of an organization from some sort of failure of its information technology (IT) systems” [33]. Following Jevtić and Lanchier [19], we define cyber risk due to a data breach as “the risk of a financial loss caused by a breach of an institution’s IT infrastructure by unauthorized parties, resulting in exploitation, taking possession of, or disclosure of data assets”. Cyber risk has attracted considerable attention within the past years due to the rapidly growing number of cyber attacks [44]. In the first six months of 2021, there were 2.5 billion malware attacks and 2.5 trillion intrusion attempts [32] in which an intruder gains or attempts to

Keywords and phrases: Cyber risk, aggregate loss, insurance premium, bond percolation, Galton-Watson tree.

School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287, USA.

* Corresponding author: schiarad@asu.edu

gain unauthorized access to a system or its network [26]. This is particularly concerning for small and medium-sized enterprises (SMEs),[1] which constitute a major part of most economies [17, 19], since they face more disadvantages than larger enterprises. According to a recent cybersecurity survey by Cynet in 2022 [7], some dominating cybersecurity disadvantages for SMEs include smaller cybersecurity budgets and staff. Therefore, it is unsurprising that an SME can have an average loss of $178,000 from just a single data breach [30]. Since economies increasingly rely on the interconnections between SMEs and large enterprises, the ever-growing vulnerabilities of SMEs pose great concerns to global supply chains [42]. Thus, protecting SMEs and providing cybersecurity resources are of paramount importance.

1.2. Cybersecurity practices and drawbacks

Organizations may use fundamental cybersecurity techniques such as multi-factor authentication (MFA) [28] or firewalls to decrease the damage of a cyberattack [24]. The purpose of these techniques, if properly conducted, is to segment the network into smaller blocks so that the contagion can be isolated or at least cause minimal damage by impacting only a fragment of the network. The strategy of *network segmentation*, which is the practice of dividing an organization's network into sub-networks of devices or accounts that share similar security requirements by separating access to the most sensitive and vulnerable services on the network *via* the principle of least privilege[2], is an important factor in determining the potential financial loss to an organization from a cyber attack [6, 8, 29]. In network segmentation, privileges are actions that an organization's employee known as a user is permitted to perform *via* an account or computer in the organization's network [28]. One example of network segmentation is the use of administrative privileges. Some users, especially network administrators, have the highest administrative privileges that give them the greatest access to the network because they can bypass critical security settings [4, 16]. These network administrators can assign different administrative privilege levels to other users' credentials based on the principle of least privilege [6, 8]. In other words, some users can traverse the network more easily than others due to their level of administrative privilege. In practice, there are many different ways to segment the network and the possibilities only increase as the organization expands in size [22, 39]. Because of this, there is a great lack of formal approaches for network administrators to optimally segment networks [24]. And so, they segment the network based on their own experiences, preferences, and available resources of the organization [22, 24]. Despite this, proper network segmentation is still typically an effective method to deter contagions from infecting other components of the organization's network and so reducing the impact of the cyber attack [9].

However, poorly segmented networks, such as improperly authorized or widely granted administrative privileges, may allow intruders to traverse the network and gain access to sensitive data by exploiting these vulnerable privileges [6, 8, 9, 28]. Incidents have shown that a hacker can exploit vulnerable privileges and gain access to confidential accounts of senior executives [35] as well as the ability to access thousands of servers [9]. Because some users have greater access to the network than others due to varying levels of administrative privileges, the credentials of a user remain one of the most sought-after data assets by attackers [16, 28, 32]. According to the Verizon 2018 Data Breach report, credentials and associated privileges abuse are three of the top five malicious actions in breaches in 2018 [37]. Furthermore, out of the 3,841 incidents analyzed in Verizon's 2021 Data Breach report, of which 1,767 had confirmed data disclosure, the data compromised were credentials of a user at 85% [38]. Because of this, administrative privileges (and other cybersecurity practices for network segmentation) are an important factor in quantifying the cyber risk of an organization, especially by not having all users with the same level of access [15]. To mitigate some of the drawbacks of particular cybersecurity practices, SMEs are turning to cyber insurance as a solution.

[1] Definitions of SMEs vary across legal jurisdictions and industries. In the U.S., medium-sized manufacturers are considered to have fewer than 500 employees [19, 36].

[2] The National Institute of Standards and Technology defines least privilege as "the principle that a security architecture should be designed so that each entity is granted the minimum system resources and authorizations that the entity needs to perform its function" [27].

1.3. Cyber insurance

Due to the increase in cyber risks, there has been more demand for cyber insurance [34] as companies have been underestimating the financial impact of cyber risk, so they purchase more coverage or higher limits [23]. According to the U.S. Government Accountability Office, there was a 60% increase in cyber insurance policies from 2016 to 2019 and approximately a 50% increase in the amount of total direct written premiums from $2.1 billion to $3.1 billion [34]. Similarly, in the U.K., the number of cyber insurance claims doubled between 2019 and 2020 [23]. For all companies with more than $500 million in annual revenue in 2020, the average cyber insurance limits rose by 2% [23]. Also in 2020, there was an increase of about 15% on average in cyber insurance pricing [23]. Furthermore, more than half of the brokers surveyed by [34] reported that their clients saw a 10–30% price increase in their cyber insurance premiums. However, more companies are purchasing cyber insurance largely due to the growing number of cyber attacks [23] despite seeing higher insurance prices from the increased severity and frequency of the attacks [34]. Therefore, many SMEs are seeking coverage against data loss, revenue loss due to data breach, legal expenses, and other costs [5, 18, 44] as a second line of defense to control and mitigate cyber attacks [2]. However, pricing cyber risk insurance products is still very new due to its unique characteristics, such as the lack of a standard scoring system or actuarial tables for rate making [44].

1.4. Literature review

The effective implementation of cyber insurance requires the proper understanding of an organization's cybersecurity risks and the fair pricing of insurance products. In this setting, there have been many actuarial works that conducted empirical analysis of cyber insurance policies [31, 43] and provided pricing for cyber insurance [11, 21, 46], particularly using Copulas [10, 18]. Moreover, many actuarial works consider publicly available datasets such as the Advisen dataset [1], the Privacy Rights Clearinghouse dataset [10, 14], and the SAS operational risk global database [13] for pricing cyber insurance premiums. However, these models consider cyber risk at the macro level rather than at the micro level, *i.e.*, an individual organization's network. Due to the inherent nature of technological advancements, cyber risk is dynamic, so reliance on historical cyber risk data could be misleading [2, 12]. Therefore, traditional actuarial modeling is insufficient for cyber insurance on the micro level. Instead, cyber risk should be modeled based on a network structure due to the organization's cyber security environment, which is the resilience of the organization's network [2, 3].

1.5. Network resilience

It describes the network's ability to function in the presence of adverse conditions [25]. Using the organization's network resilience, one can have a better understanding of an optimal risk management strategy [2]. There have been efforts in studying cyber risk from the micro level perspective such as using epidemic models [3, 44], Bayesian models [2, 45], and attack graphs [40, 41]. Despite these efforts, the above papers do not fully utilize the aspect of cyber resilience, which is very much connected to bond percolation [19, 25].

The dynamical model designed in Jevtić and Lanchier [19] for SMEs consists of several components, including a bond percolation process, to calculate the aggregate loss resulting from consecutive cyber attacks on a tree based local area network (LAN). Exclusively tree based LAN topologies are not typically observable in large enterprises [19] since larger enterprises have more available resources and personnel, thus creating more complex networks, which may result in large networks with cycles. Therefore, their results are primarily applicable to SMEs, which are the dominant form of organizations in an economy [19]. Their model has received some rapidly growing interest from the actuarial science community, reaching recently the status of United States patent (see [20]), and the goal of this paper is to extend their model and improve their results. Furthermore, the percolation process of their model assumes that the contagion spreads through the edges of the network with a fixed probability, thus modeling a static cyber security environment that cannot account for various network segmentation strategies. We extend their model by assuming that the contagion spreads through the edges with a certain probability of moving toward the lower levels of the network but with another (typically smaller) probability of moving toward the upper levels of the network, thus resulting in a bidirectional version of the percolation component in [19] that accounts for network segmentation strategies an SME may implement. Therefore, our bidirectional version considers a more realistic and dynamic cyber security environment of an SME's network. Furthermore, while [19] only obtained a rough upper bound for the variance of the aggregate loss, our analysis relies on other techniques that lead to an exact expression of the variance, and therefore an exact expression for the insurance premiums.

2. Model description

As previously explained, the model we consider in this paper is a bidirectional version of the model introduced in Jevtić and Lanchier [19] for SMEs. In this section, we give a rigorous description of the five components of the model. To begin with, we assume that the cyber attacks occur in continuous time at a constant rate:

1. we let (N_t) be a Poisson process with intensity λ,

and assume that the ith attack occurs at time

$$T_i = \inf\{t : N_t = i\} \quad \text{for} \quad i = 1, 2, \ldots$$

In other words, the times between consecutive cyberattacks are independent exponential random variables with the same parameter λ, meaning that

$$P(T_1 > s) = P(T_{i+1} - T_i > s) = e^{-\lambda s} \quad \text{for all} \quad s > 0 \quad \text{and} \quad i = 1, 2, \ldots$$

At the times of the attacks, the LAN consists of a random tree, which is more observable in SMEs than larger enterprises [19], so the network topology does not contain any cycle. Depending on whether the network is static or dynamic, the network can be fixed once and for all or consist of a sequence of independent realizations of the random tree, but we point out that our results are not sensitive to this distinction. Moreover, the lack of publicly available information about the size and distribution of the nodes on a LAN (see [22]) requires us to take into account the stochastic nature of real-world networks. More precisely, in the dynamical context,

2. we let $\mathbb{T}_i = (V_i, E_i)$ be independent realizations of the Galton-Watson tree to account for the lack of cycles in the network, the stochastic nature of an organization's growth, and the variability of segmenting a network. Furthermore, the Galton-Watson tree has radius R, *i.e.*, there are k vertices connected to the root with probability p_k, then k additional vertices connected to each of those vertices with probability p_k, and so on, up to generation R. To ensure that the tree has radius R and avoid trivialities, we assume $p_0 = 0$.

In the terminology of branching processes, the probabilities p_k are referred to as the offspring distribution, and two vertices connected by an edge are called the parent and the offspring, with the parent being the vertex closer to the root. Next, to fix the source of the attack,

3. we let $X_i \in V_i$ be a vertex chosen at random.

The actual distribution of X_i is unimportant for this work since our objective is to compute the insurance premium as a function of X_i. Now, to model the contagion itself (how the infection spreads through the network from the source), we use bidirectional bond percolation:

4. we let $p, q \in (0, 1)$, and assume that each edge of the tree is independently open

$$\begin{array}{ll} \text{with probability } p & \text{in the direction parent} \to \text{offspring} \\ \text{with probability } q & \text{in the direction offspring} \to \text{parent.} \end{array}$$

In other words, each edge is identified to two arrows. The arrow going away from the root is open with probability p whereas the arrow going toward the root is open with probability q. As previously mentioned, the distinction between p and q is motivated by the presence of administrative privileges suggesting that the infection spreads more easily moving away from the root than toward the root, meaning that $p > q$. The model in [19] corresponds to the particular case $p = q$. Then, the set of infected vertices is the open percolation cluster starting from the source:

$$\mathscr{C}_i = \{y \in V_i : \text{there is a directed open path } X_i \to x\}.$$

To define the aggregate loss of multiple cyberattacks on the network, the last step is to assign a cost to the percolation cluster:

5. we let $c_i(x)$ for all $i > 0$ and $x \in V_i$ be independent and identically distributed

and think of this random variable as the cost of vertex x. Then, the total loss resulting from the ith cyber attack and the aggregate loss up to time t are given respectively by

$$C_i = \sum_{x \in \mathscr{C}_i} c_i(x) \quad \text{and} \quad L_t = \sum_{i=1}^{N_t} C_i = \sum_{i=1}^{N_t} \sum_{x \in \mathscr{C}_i} c_i(x).$$

In other words, the loss resulting from the ith cyber attack is the total cost of all the vertices that have been infected during the attack, and the aggregate loss is the cumulative loss resulting from all the cyber attacks that occurred by time t.

3. Main results

Our main objective is to compute the mean and the variance of L_t as insurance premiums are calculated from these two quantities. To state our results and express the mean and the variance of the aggregate loss, we need some key quantities. First, we let μ and σ^2 be respectively the mean and the variance of the offspring distribution (the random number of edges starting from each vertex), which we assume to be finite:

$$\mu = \sum_{k=1}^{\infty} k p_k < \infty \quad \text{and} \quad \sigma^2 = \sum_{k=1}^{\infty} (k - \mu)^2 p_k < \infty.$$

Recalling that the local costs $c_i(x)$ are identically distributed, to simplify the notation, we let c denote their common distribution. Similarly, because the consecutive Galton-Watson trees, percolation processes, sources of infection, and local costs are identically distributed, the numbers of infected vertices $S_i = \text{card}\,(\mathscr{C}_i)$ are also identically distributed, and we let S denote the common distribution of the size of the consecutive percolation clusters. The model parameters (offspring distribution, distribution of the source, percolation parameters p and q, and distribution of the local costs) vary from one company to another, but the goal of this paper is not to estimate these parameters. Instead, our main objective is to compute explicitly the mean and variance of the aggregate loss, and therefore the insurance premiums, as a function of these parameters.

3.1. Aggregate loss

The aggregate loss can be expressed using the loss resulting from a single attack by conditioning on the number of attacks, which is a critical component in the frequency-severity approach as the industry standard for pricing insurance risks [19, 44, 47]. Similarly, the loss resulting from a single attack can be expressed using the local cost c by conditioning on the cluster size S. Using also that the number of attacks is Poisson distributed, we obtain the following result.

Theorem 3.1. *The mean and variance of the aggregate loss are given by*

$$E(L_t) = \lambda t E(S) E(c) \quad \text{and} \quad \operatorname{Var}(L_t) = \lambda t E(S) \operatorname{Var}(c) + \lambda t E(S^2)(E(c))^2.$$

The theorem shows that the mean and variance of the aggregate loss depends on the first and second moments of the size of the infected cluster from a single contagion so we can capture the aggregate loss from the size of a single contagion. Therefore, in order to compute the insurance premiums, the next step is to compute the first and second moments of the cluster size.

3.2. First moment

To begin with, we look at the first moment of the cluster size. This quantity has been computed explicitly in [19] in the symmetric case $p = q$ using combinatorial techniques. To extend their result to the more general asymmetric case, we partition the set of infected vertices into subtrees and use linearity of the expectation, which gives the following theorem.

Theorem 3.2. *The conditional first moment on the tree with radius R given that the infection starts at distance r from the root is equal to*

$$E_r(S) = \frac{1}{1-\mu p}\left(1 + q\left(\frac{1-q^r}{1-q}\right)(1-p) - (\mu p)^{R-r+1}\left(\frac{1 - pq(1+(\mu-1)(\mu pq)^r)}{1-\mu pq}\right)\right)$$

for all $\mu p, \mu pq, q \neq 1$.

Our theorem excludes the cases $\mu p = 1$, $\mu pq = 1$ and $q = 1$ for simplicity but the first moment in these special cases can be easily deduced from the fact that it is continuous with respect to the parameters μ, p and q. The same holds for the second moment in the next theorem. Note also that setting $p = q$ in the theorem gives

$$E_r(S) = \frac{1}{1-\mu p}\left(1 + p(1-p^r) - (\mu p)^{R-r+1}\left(\frac{1-p^2(1+(\mu-1)(\mu p^2)^r)}{1-\mu p^2}\right)\right)$$

for all $\mu p, \mu p^2 \neq 1$, which is exactly the expression found in Theorem 4 of [19], but we point out that, even though our result extends the result in [19] to the asymmetric case, our approach leads to a much shorter and more elegant proof. More precisely, the proof in [19] relies on a tedious combinatorial argument that consists in counting the number of open paths of a given length starting from the source of the infection whereas our proof consists in finding the infected vertices along the path going from the source of the infection to the root of the tree and then partitioning the cluster of infected vertices into (disjoint) subtrees starting from each of these vertices.

Natural and more realistic extensions of the model can be obtained by turning p and q into functions that increase with the height, modeling the fact that edges closer to the root are less vulnerable due to a reinforcement of security. Similarly, instead of assuming that the costs are identically distributed across the network, one can assume more realistically that the cost decreases stochastically with the distance from the root, modeling the fact that components far from the root contain less sensitive information than the ones closer to the root. These more general assumptions, however, lead to messy algebra and the lack of an explicit expression for the first moment. For instance, replacing the parameter q which is constant across the network by a collection

$$q_1 < q_2 < \cdots < q_{R-1} < q_R$$

where q_i represents the probability that an arrow going from distance i to distance $i-1$ from the root is open, and repeating the proof of Theorem 3.2 below, show that the conditional first moment of the size of the infection

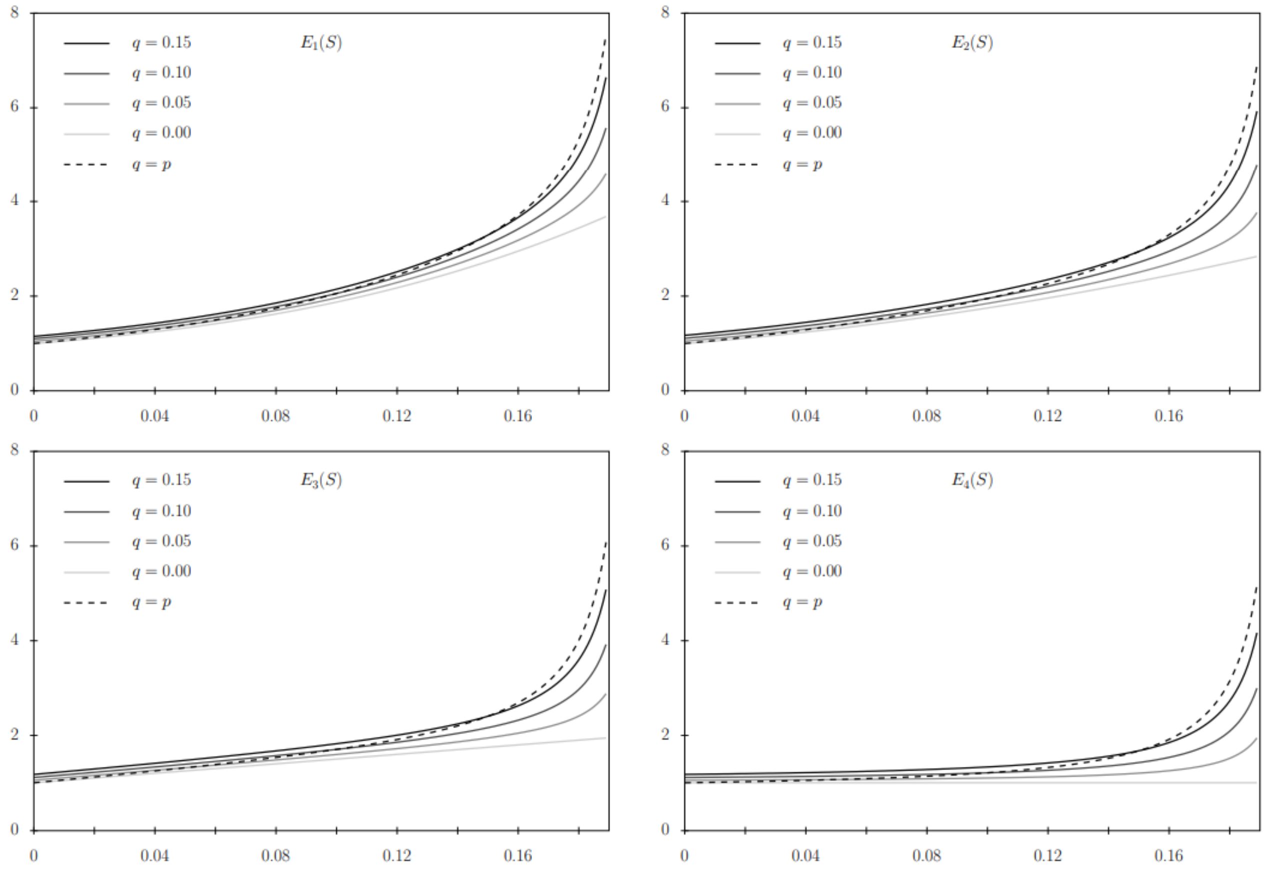

Figure 1. First moment of the cluster size as a function of p for various values of q and r. In each picture, the mean and variance of the offspring distribution are $\mu = \sigma^2 = 5$ and the radius of the tree is $R = 4$.

becomes

$$\begin{aligned} E_r(S) \;&=\; \frac{1-(\mu p)^{R-r+1}}{1-\mu p} + \sum_{k=0}^{r-1} k \left(\prod_{j=0}^{k-1} q_{r-j}\right)(1-q_{r-k}) \\ &\quad + r\left(\prod_{j=0}^{r-1} q_{r-j}\right) + (\mu-1)p \sum_{i=0}^{r-1}\left(\prod_{j=0}^{i} q_{r-j}\right) \frac{1-(\mu p)^{R-r+i+1}}{1-\mu p} \end{aligned}$$

for all $\mu p \neq 1$, which cannot be further simplified. Assuming instead that the costs are not identically distributed does not change the size of the infection. However, the mean and variance of the aggregate loss can no longer be deduced from Theorem 3.1 and one would have to compute separately the moments of the number of infected vertices at a given distance from the root, which again leads to messy algebra and the lack of explicit expressions. In particular, we assume some global homogeneity across the network to make our model mathematically tractable.

3.3. Second moment

Our approach to compute the second moment is similar but relies in addition on independence. More precisely, writing again the cluster of infected vertices as a disjoint union of subtrees, the second moment can be computed using that the sizes of these subtrees are independent random variables. To express the second moment of the cluster size, we let

$$\mu_+ = \mu p \quad \text{and} \quad \mu_- = (\mu - 1)p, \tag{3.1}$$

quantities that will be interpreted later as the mean number of infected offspring in certain subtrees of the LAN, and

$$\sigma_+^2 = p(1-p)\mu + p^2\sigma^2 \quad \text{and} \quad \sigma_-^2 = p(1-p)(\mu - 1) + p^2\sigma^2, \tag{3.2}$$

quantities that will be interpreted later as the variance of the number of infected offspring in certain subtrees of the LAN. For all $j = 0, 1, \ldots, r$, we also define

$$\begin{aligned} \mu_{1,j} &= \frac{1 - \mu_+^{R-j+1}}{1 - \mu_+} \\ \mu_{2,j} &= \frac{\sigma_+^2}{(1-\mu_+)^2} \left(\frac{1 - \mu_+^{2(R-j)+1}}{1 - \mu_+} - (2(R-j)+1)\mu_+^{R-j} \right) + \left(\frac{1 - \mu_+^{R-j+1}}{1 - \mu_+} \right)^2 \end{aligned} \tag{3.3}$$

for all $\mu_+ \neq 1$.

Theorem 3.3. *The conditional second moment on the tree with radius R given that the infection starts at distance r from the root is equal to*

$$\begin{aligned} E_r(S^2) = \sum_{k=0}^{r} \Bigg(\mu_{2,r} &+ 2\mu_{1,r} \sum_{i=1}^{k} (1 + \mu_- \mu_{1,r-i+1}) \\ &+ \sum_{i=1}^{k} (1 + 2\mu_-\mu_{1,r-i+1} + \mu_-\mu_{2,r-i+1} + (\sigma_-^2 + \mu_-^2 - \mu_-)(\mu_{1,r-i+1})^2) \\ &+ \sum_{i,j \in \{1,2,\ldots,k\}, i \neq j} (1 + \mu_-\mu_{1,r-i+1})(1 + \mu_-\mu_{1,r-j+1}) \Bigg) q_k. \end{aligned}$$

where $q_k = q^k(1-q)$ for $k = 0, 1, \ldots, r-1$, and $q_r = q^r$.

Theorem 3.3 is the main contribution of this work as it gives an exact expression of the second moment, which leads to an exact pricing of the standard deviation principle, whereas ([19], Thm. 5) only derived a rough upper bound for this pricing. Although the expression is implicit, it can be computed explicitly using a computer program (see Fig. 2) for each set of parameters.

We also point out that the expressions in both Theorems 3.2 and 3.3 can be simplified when the LAN consists of the infinite Galton-Watson tree. In this case, the percolation process is supercritical when $\mu p > 1$, meaning that the cluster of infected vertices is infinite with positive probability, so the first and second moments are both infinite. In the subcritical phase $\mu p < 1$, it follows from the monotone convergence theorem that the first and second moments of the cluster size on the infinite tree can be obtained by taking the limit as $R \to \infty$ in both theorems. In particular, in the infinite case, the first moment reduces to

$$E_r(S) = \frac{1}{1-\mu p}\left(1 + q\left(\frac{1-q^r}{1-q}\right)(1-p)\right)$$

for all $\mu p, q \neq 1$ while, using some algebra, we get

$$\begin{aligned} E_r(S^2) \ = \ & \frac{1}{(1-\mu p)^2}\left(1 + \frac{p(1-p)\mu + p^2\sigma^2}{1-\mu p}\right) \\ & + \left(1 + \frac{2(1+(\mu-1)p)}{1-\mu p} + \frac{2(\mu-1)p + p(1-p)(\mu-1) + p^2\sigma^2 + (\mu-1)^2p^2}{(1-\mu p)^2}\right. \\ & \qquad\qquad \left. + \ \frac{(p(1-p)\mu + p^2\sigma^2)(\mu-1)p}{(1-\mu p)^3}\right) q\left(\frac{1-q^r}{1-q}\right) \\ & + \left(1 + \frac{(\mu-1)p}{1-\mu p}\right)^2 \frac{2q^2(1 - rq^{r-1} + (r-1)q^r)}{(1-q)^2} \end{aligned} \tag{3.4}$$

for all $\mu p, q \neq 1$ for the second moment on the infinite tree (see Sect. 7 for a proof).

3.4. Exponential decay of the diameter

Another quantity of interest that also accounts for the geometry of the set of infected vertices is the diameter of the cluster $\mathscr{C}$ defined as the maximum graph distance (number of edges) between any two infected vertices:

$$\operatorname{diam}(\mathscr{C}) = \max\{d(x,y) : x, y \in \mathscr{C}\}.$$

In this case, studying the spread of the infection from a dynamical point of view starting from the highest infected vertex and moving one generation down the tree at each time step, we can prove an exponential decay. More precisely, we have the following theorem.

Theorem 3.4. *Let $\mu p < 1$. Then, the conditional probability that the diameter is larger than $2n$ given that the infection starts at distance r from the root is*

$$P_r(\operatorname{diam}(\mathscr{C}) \geq 2n) \leq \frac{1-(q/\mu p)^{r+1}}{1-(q/\mu p)}\ (\mu p)^n \quad \textit{for all} \quad n > r \ \textit{and}\ \mu p \neq 1.$$

The theorem indeed implies that, in the subcritical phase $\mu p < 1$, the tail distribution of the diameter of the cluster of infected vertices decays exponentially.

3.5. Insurance premiums

Combining Theorems 3.1–3.3, we can now perform an exact pricing of cyber risk insurance. Following [19], we consider the three pricing principles

$$\begin{aligned} \text{actuarial fair premium:} \quad & P = E(L_1) \\ \text{expectation principle:} \quad & P = E(L_1) + \delta E(L_1) \\ \text{standard deviation principle:} \quad & P = E(L_1) + \delta\sqrt{\operatorname{Var}(L_1)} \end{aligned}$$

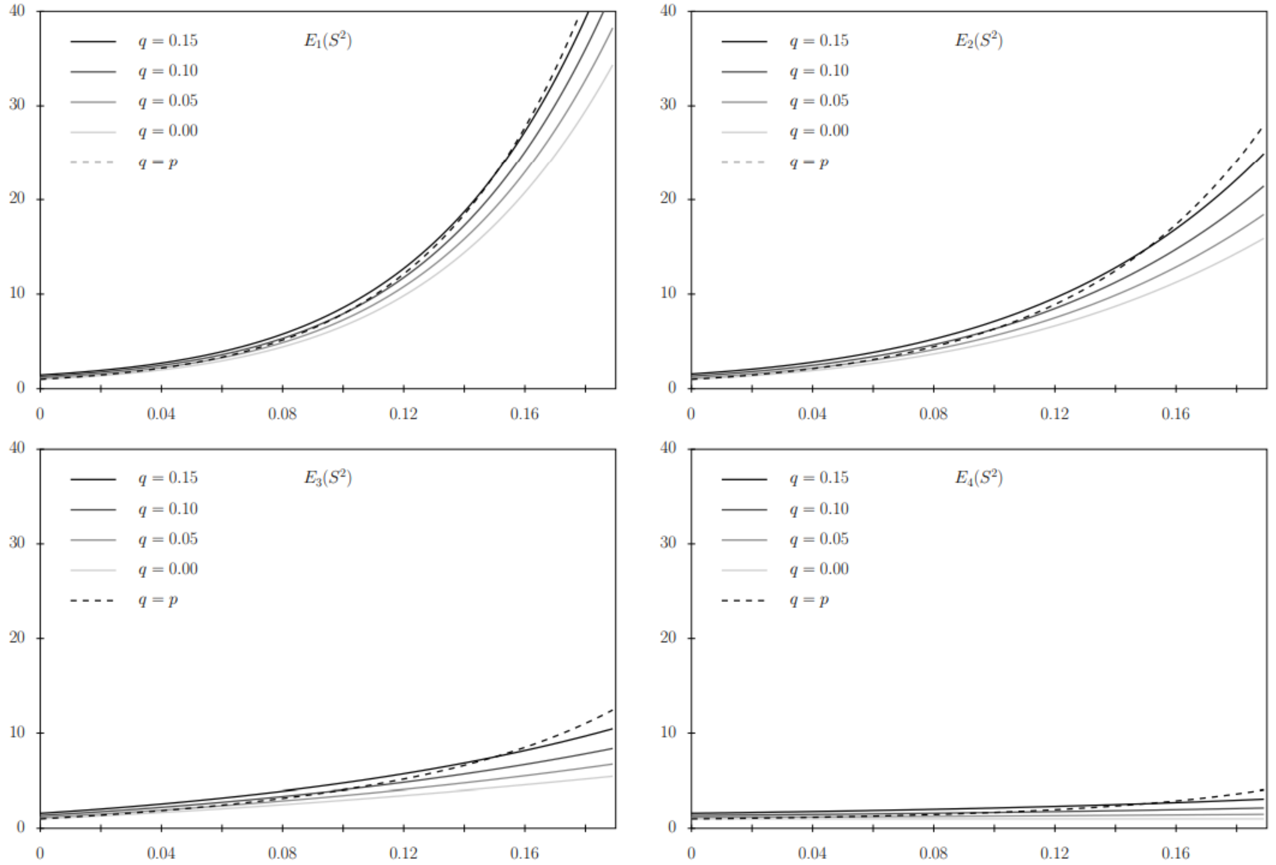

FIGURE 2. Second moment of the cluster size as a function of p for various values of q and r. In each picture, the mean and variance of the offspring distribution are $\mu = \sigma^2 = 5$ and the radius of the tree is $R = 4$.

where L_1 represents the aggregate loss per unit of time. According to Theorem 3.1,

$$E(L_1) = \lambda E(S)E(c) \quad \text{and} \quad \sqrt{\mathrm{Var}(L_1)} = \sqrt{\lambda E(S)\,\mathrm{Var}(c) + \lambda E(S^2)(E(c))^2}, \tag{3.5}$$

showing that all three insurance premiums are functions of the system parameters and the first and second moments of the cluster size. In particular, for each set of parameters (the rate of occurrence of the cyber attacks, the mean and variance of the offspring distribution, the local costs, etc.), all three premiums can be computed exactly using Theorems 3.2 and 3.3. Because the expressions of the first two moments in the theorems are quite complicated, we refer the reader to Figures 1 and 2 for pictures of the first and second moments as functions of the parameters p, q and r, for a fixed value of the mean and variance of the offspring distribution. The dashed curves in the pictures correspond to the symmetric case $p = q$ considered in [19] where the first moment was computed exactly while only a very rough upper bound for the second moment was derived. In contrast, the approach used in this paper leads to exact insurance premiums for all three principles.

Not surprisingly, it follows from (3.5) that all three premiums are nondecreasing with respect to the rate λ and the mean and variance of the local costs. The figures also show that both the first and the second moments of the cluster size are nondecreasing with respect to p and q and so are the three premiums. This result is intuitively clear and can be proved rigorously using a popular technique in probability theory called coupling, which consists in this case in constructing bond percolation processes with different parameters on the same probability space. We note however that the cluster size and the insurance premiums are not always increasing or always decreasing with respect to r, the distance from the root to the source of the infection. Indeed,

- When $p = 1$ and $q = 0$, the set of infected vertices consists of the subtree starting from the source of the infection going away from the root therefore, in this case, the cluster size and the premiums are decreasing with respect to r.
- When $p = 0$ and $q = 1$, the set of infected vertices consists of the unique path going from the source of the infection to the root of the tree therefore, in this case, the cluster size and the premiums are increasing with respect to r.

The rest of this paper is devoted to the proof of the four theorems.

4. Proof of Theorem 3.1 (aggregate loss)

To prove Theorem 3.1, we first observe that, because the Galton-Watson trees, percolation processes and local costs are independent and identically distributed across time, the consecutive costs C_i are independent and identically distributed as well. In particular, letting C be the common distribution of the random variables C_i and conditioning on the number of cyber attacks N_t, which is a Poisson process, we get

$$\begin{aligned} E(L_t \mid N_t = n) &= E(C_1 + \cdots + C_{N_t} \mid N_t = n) = nE(C) \\ \operatorname{Var}(L_t \mid N_t = n) &= \operatorname{Var}(C_1 + \cdots + C_{N_t} \mid N_t = n) = n \operatorname{Var}(C). \end{aligned} \tag{4.1}$$

The first equation in (4.1) implies that

$$E(L_t) = E(E(L_t \mid N_t)) = E(N_t E(C)) = E(N_t)E(C) \tag{4.2}$$

while using also the second equation in (4.1) and the law of total variance,

$$\begin{aligned} \operatorname{Var}(L_t) &= E(\operatorname{Var}(L_t \mid N_t)) + \operatorname{Var}(E(L_t \mid N_t)) = E(N_t \operatorname{Var}(C)) + \operatorname{Var}(N_t E(C)) \\ &= E(N_t) \operatorname{Var}(C) + \operatorname{Var}(N_t)(E(C))^2. \end{aligned} \tag{4.3}$$

Using that the local costs are independent and identically distributed across the LAN, and conditioning on the size S of a single cyber attack, we also have

$$E(C \mid S = s) = sE(c) \quad \text{and} \quad \operatorname{Var}(C \mid S = s) = s \operatorname{Var}(c). \tag{4.4}$$

The first equation in (4.4) implies that

$$E(C) = E(E(C \mid S)) = E(SE(c)) = E(S)E(c) \tag{4.5}$$

while using also the second equation in (4.4) and the law of total variance,

$$\begin{aligned} \operatorname{Var}(C) &= E(\operatorname{Var}(C \mid S)) + \operatorname{Var}(E(C \mid S)) = E(S \operatorname{Var}(c)) + \operatorname{Var}(SE(c)) \\ &= E(S) \operatorname{Var}(c) + \operatorname{Var}(S)(E(c))^2. \end{aligned} \tag{4.6}$$

Finally, using that $E(N_t) = \operatorname{Var}(N_t) = \lambda t$, and combining (4.2) and (4.5), we get

$$E(L_t) = \lambda t E(C) = \lambda t E(S)E(c).$$

Combining (4.3), (4.5) and (4.6), and using that $\operatorname{Var}(S) + (E(S))^2 = E(S^2)$, we get

$$\begin{aligned} \operatorname{Var}(L_t) &= \lambda t \operatorname{Var}(C) + \lambda t (E(C))^2 \\ &= \lambda t E(S) \operatorname{Var}(c) + \lambda t \operatorname{Var}(S)(E(c))^2 + \lambda t (E(S)E(c))^2 \\ &= \lambda t E(S) \operatorname{Var}(c) + \lambda t E(S^2)(E(c))^2. \end{aligned}$$

This completes the proof of the theorem.

5. Partition into disjoint subtrees

To get ready for the proofs of Theorems 3.2 and 3.3 in the next two sections, we first describe the partition of the set of infected vertices into disjoint subtrees and collect several useful preliminary results that explain the parameters introduced in (3.1)–(3.3). More precisely, we study the distribution of the random number of subtrees and compute the first and second moment of the size of these subtrees. From now on, we assume that the infection starts at a vertex x with $d(0, x) = r$. By spherical symmetry, the specific choice of x is unimportant as long as the vertex is at distance r from the root. There is a unique directed path

$$x_0 = 0 \to x_1 \to x_2 \to \cdots \to x_{r-1} \to x_r = x$$

of length r going from the root to vertex x and we let

$$D = \max \{i = 0, 1, \ldots, r : x_{r-i} \text{ is infected}\}.$$

This is the distance between the source of the infection and the highest infected vertex, and we refer the reader to Figure 3 for a picture. The following lemma gives preliminary results about the random variable D that will be useful later to prove the theorems.

Lemma 5.1. *For all $q \neq 1$,*

$$E(D) = q \left(\frac{1 - q^r}{1 - q} \right) \quad \text{and} \quad E(D(D-1)) = \frac{2q^2(1 - rq^{r-1} + (r-1)q^r)}{(1-q)^2}.$$

Proof. Because the infection spreads toward the root of the random tree with probability q independently through each of the edges, we have

$$P(D = k) = q^k(1-q) \quad \text{for } k = 0, 1, \ldots, r-1, \quad \text{and} \quad P(D = r) = q^r. \tag{5.1}$$

In particular, using that, for all $x \neq 1$,

$$\sum_{k=1}^{r-1} k x^{k-1} = \frac{\partial}{\partial x} \left(\sum_{k=0}^{r-1} x^k \right) = \frac{\partial}{\partial x} \left(\frac{1 - x^r}{1 - x} \right) = \frac{1 - r x^{r-1} + (r-1)x^r}{(1-x)^2},$$

we deduce that the first moment is given by

$$\begin{aligned} E(D) &= \sum_{k=0}^{r-1} k q^k (1-q) + r q^r = q(1-q) \sum_{k=1}^{r-1} k q^{k-1} + r q^r \\ &= q(1-q) \left(\frac{1 - r q^{r-1} + (r-1) q^r}{(1-q)^2} \right) + r q^r = q \left(\frac{1 - q^r}{1 - q} \right) \end{aligned}$$

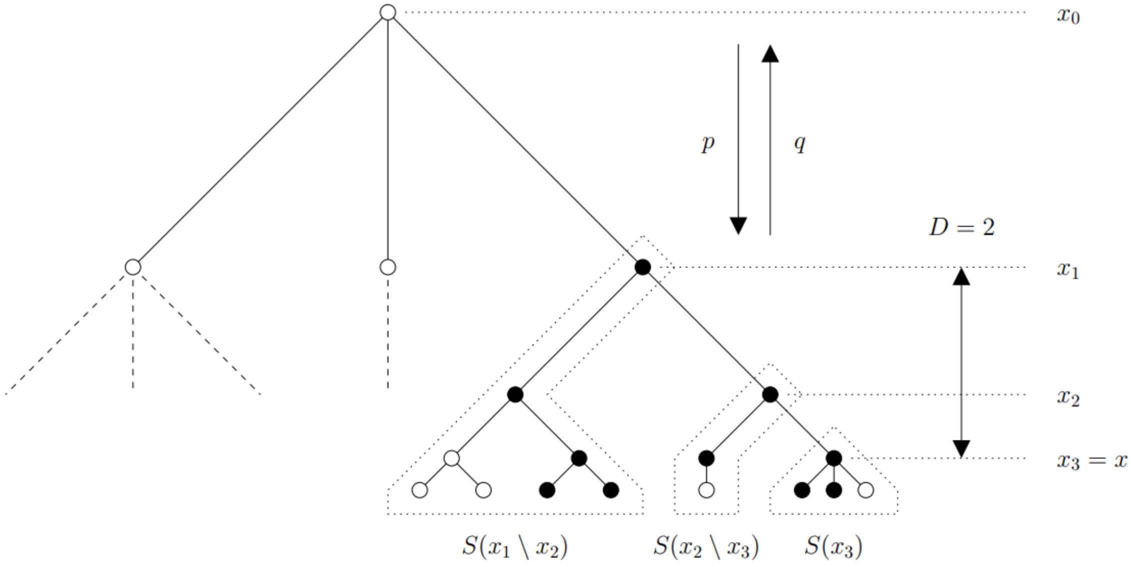

FIGURE 3. Picture of the partition into disjoint subtrees used to prove Theorems 3.2 and 3.3. The black vertices represent the set of infected vertices while the white vertices are not infected. In our example, the infection starts from x_3 and spreads up to x_1, which results in a partition of the cluster of infected vertices into three disjoint subtrees. Starting from the source of the infection, the numbers of infected vertices in the subtrees are 3, 2 and 5, respectively.

for all $q \neq 1$. Similarly, we have

$$\begin{aligned}\sum_{k=2}^{r-1} k(k-1)x^{k-2} &= \frac{\partial^2}{\partial x^2}\left(\sum_{k=0}^{r-1} x^k\right) = \frac{\partial}{\partial x}\left(\frac{1-rx^{r-1}+(r-1)x^r}{(1-x)^2}\right) \\ &= \frac{2-r(r-1)x^{r-2}+2r(r-2)x^{r-1}-(r-1)(r-2)x^r}{(1-x)^3}\end{aligned}$$

for all $x \neq 1$, from which it follows that

$$\begin{aligned}E(D(D-1)) &= \sum_{k=0}^{r-1} k(k-1)q^k(1-q)+r(r-1)q^r \\ &= q^2(1-q)\ \frac{2-r(r-1)q^{r-2}+2r(r-2)q^{r-1}-(r-1)(r-2)q^r}{(1-q)^3} \\ &\qquad\qquad + q^2\ \frac{r(r-1)q^{r-2}-2r(r-1)q^{r-1}+r(r-1)q^r}{(1-q)^2} \\ &= \frac{2q^2(1-rq^{r-1}+(r-1)q^r)}{(1-q)^2}\end{aligned}$$

for all $q \neq 1$. This completes the proof. □

Next, for $j = 0, 1, \ldots, r$, we define the random variables

$$\begin{array}{rcl} S(x_j) & = & \text{number of infected vertices in the subtree starting at } x_j \\ S(x_j \setminus x_{j+1}) & = & \text{number of infected vertices in the subtree starting at } x_j \\ & & \text{but excluding the subtree starting at } x_{j+1}. \end{array} \tag{5.2}$$

For instance, in the realization shown in Figure 3, we have

$$S(x_3) = 3, \quad S(x_2 \setminus x_3) = 2, \quad S(x_1 \setminus x_2) = 5, \quad S(x_0 \setminus x_1) = 0.$$

To estimate the first and second moments of the size of the infected cluster later, we now compute the first and second moments of the random variables in (5.2). To do this, let $\xi_i = \text{Bernoulli}\,(p)$ be independent, and let Y be the offspring distribution, *i.e.*, the random variable describing the number of edges starting from a given vertex and going away from the root. In particular,

$$X_+ = \xi_1 + \xi_2 + \cdots + \xi_Y \quad \text{and} \quad X_- = \xi_1 + \xi_2 + \cdots + \xi_{Y-1}$$

are the random variables describing the number of infected offspring of a given vertex and the number of infected offspring of a given vertex excluding a given offspring, respectively. The next two lemmas give the mean and the variance of these two random variables.

Lemma 5.2. *We have $E(X_+) = \mu_+$ and $E(X_-) = \mu_-$ as defined in* (3.1).

Proof. Conditioning on Y, we get

$$E(X_+) = E(E(X_+ \,|\, Y)) = E(E(\xi_1 + \cdots + \xi_Y \,|\, Y)) = E(Y)E(\xi_n) = \mu p = \mu_+.$$

Similarly, for the mean of X_-,

$$E(X_-) = E(E(X_- \,|\, Y)) = E(E(\xi_1 + \cdots + \xi_{Y-1} \,|\, Y)) = E(Y-1)E(\xi_n) = (\mu - 1)p = \mu_-.$$

This completes the proof. □

Lemma 5.3. *We have $\text{Var}(X_+) = \sigma_+^2$ and $\text{Var}(X_-) = \sigma_-^2$ as defined in* (3.2).

Proof. Conditioning on Y and using the law of total variance, we get

$$\begin{array}{rcl} \text{Var}(X_+) & = & E(\text{Var}(X_+ \,|\, Y)) + \text{Var}(E(X_+ \,|\, Y)) \\ & = & E(p(1-p)Y) + \text{Var}(pY) = p(1-p)\mu + p^2\sigma^2 = \sigma_+^2. \end{array}$$

Similarly, for the variance of X_-,

$$\begin{array}{rcl} \text{Var}(X_-) & = & E(\text{Var}(X_- \,|\, Y)) + \text{Var}(E(X_- \,|\, Y)) \\ & = & E(p(1-p)(Y-1)) + \text{Var}(p(Y-1)) = p(1-p)(\mu - 1) + p^2\sigma^2 = \sigma_-^2. \end{array}$$

This completes the proof. □

Now, applying Lemmas 8 and 9 from [19] implies that the first and second moments of the first set of random variables in (5.2) are given respectively by

$$\begin{aligned} E(S(x_j)) &= \frac{1-\mu_+^{R-j+1}}{1-\mu_+} \\ E(S(x_j)^2) &= \frac{\sigma_+^2}{(1-\mu_+)^2}\left(\frac{1-\mu_+^{2(R-j)+1}}{1-\mu_+} - (2(R-j)+1)\mu_+^{R-j}\right) + \left(\frac{1-\mu_+^{R-j+1}}{1-\mu_+}\right)^2 \end{aligned}$$

for all $\mu_+ \neq 1$, which are the expressions in (3.3). In particular, we have the following lemma.

Lemma 5.4. *For all $j = 0, 1, \ldots, r$, we have $E(S(x_j)) = \mu_{1,j}$ and $E(S(x_j)^2) = \mu_{2,j}$.*

We now look at the second set of random variables in (5.2).

Lemma 5.5. *For all $j = 0, 1, \ldots, r$, we have $E(S(x_j \setminus x_{j+1})) = 1 + \mu_-\mu_{1,j+1}$ and*

$$E(S(x_j \setminus x_{j+1})^2) = 1 + 2\mu_-\mu_{1,j+1} + \mu_-\mu_{2,j+1} + (\sigma_-^2 + \mu_-^2 - \mu_-)(\mu_{1,j+1})^2.$$

Proof. Letting $y_1, y_2, \ldots, y_X$ be the infected offspring of x_j other than x_{j+1},

$$S(x_j \setminus x_{j+1}) = 1 + S(y_1) + \cdots + S(y_X) \quad \text{and} \quad X = X_- \text{ in distribution.} \tag{5.3}$$

In particular, conditioning on X and using Lemmas 5.2 and 5.4, we obtain

$$\begin{aligned} E(S(x_j \setminus x_{j+1})) &= E(E(1 + S(y_1) + \cdots + S(y_X) \,|\, X)) \\ &= 1 + E(X)E(S(y_1)) = 1 + \mu_- E(S(x_{j+1})) = 1 + \mu_-\mu_{1,j+1}. \end{aligned}$$

Taking the square in (5.3),

$$\begin{aligned} S(x_j \setminus x_{j+1})^2 &= 1 + 2\sum_{n=1}^{X} S(y_n) + \left(\sum_{n=1}^{X} S(y_n)\right)^2 \\ &= 1 + 2\sum_{n=1}^{X} S(y_n) + \sum_{n=1}^{X} S(y_n)^2 + \sum_{n,m\in\{1,2,\ldots,X\}, n\neq m} S(y_n)S(y_m), \end{aligned}$$

then conditioning on X and using independence as well as Lemmas 5.2–5.4,

$$\begin{aligned} E(S(x_j \setminus x_{j+1})^2) &= E(E(S(x_j \setminus x_{j+1})^2 \,|\, X)) \\ &= 1 + 2E(X)E(S(y_n)) + E(X)E(S(y_n)^2) + E(X(X-1))E(S(y_n))^2 \\ &= 1 + 2\mu_- E(S(x_{j+1})) + \mu_- E(S(x_{j+1})^2) + (\sigma_-^2 + \mu_-^2 - \mu_-)E(S(x_{j+1}))^2 \\ &= 1 + 2\mu_-\mu_{1,j+1} + \mu_-\mu_{2,j+1} + (\sigma_-^2 + \mu_-^2 - \mu_-)(\mu_{1,j+1})^2. \end{aligned}$$

This completes the proof. □

6. Proof of Theorem 3.2 (first moment)

Using the previous lemmas, we are now ready to prove Theorem 3.2. To begin with, observe that, on the event $D = k$, the total number of infected vertices can be written as

$$S = S(x_r) + \sum_{i=r-k}^{r-1} S(x_i \setminus x_{i+1}) = S(x_r) + \sum_{i=1}^{k} S(x_{r-i} \setminus x_{r-i+1}). \tag{6.1}$$

Then, conditioning on D and using Lemma 5.5, we get

$$\begin{aligned} E_r(S) &= \sum_{k=0}^{r} E(S \,|\, D = k)\, P(D = k) \\ &= \sum_{k=0}^{r} \Bigg(E(S(x_r)) + \sum_{i=1}^{k} E(S(x_{r-i} \setminus x_{r-i+1})) \Bigg) P(D = k) \\ &= \sum_{k=0}^{r} \Bigg(E(S(x_r)) + \sum_{i=0}^{k-1} (1 + \mu_- E(S(x_{r-i})) \Bigg) P(D = k) \\ &= E(S(x_r)) + E(D) + \mu_- \sum_{k=1}^{r} \sum_{i=0}^{k-1} E(S(x_{r-i}))\, P(D = k). \end{aligned} \tag{6.2}$$

Exchanging the two sums, and using Lemma 5.4, we obtain

$$\begin{aligned} \sum_{k=1}^{r} \sum_{i=0}^{k-1} E(S(x_{r-i}))\, P(D = k) &= \sum_{i=0}^{r-1} \sum_{k=i+1}^{r} E(S(x_{r-i}))\, P(D = k) \\ &= \sum_{i=0}^{r-1} E(S(x_{r-i}))\, P(D > i) = \sum_{i=0}^{r-1} q^{i+1}\, E(S(x_{r-i})) = \sum_{i=0}^{r-1} q^{i+1}\, \mu_{1,r-i} \\ &= \sum_{i=0}^{r-1} q^{i+1} \left(\frac{1 - \mu_+^{R-r+i+1}}{1 - \mu_+} \right) = \frac{1}{1 - \mu_+} \left(q \sum_{i=0}^{r-1} q^i - q\, \mu_+^{R-r+1} \sum_{i=0}^{r-1} (q\mu_+)^i \right) \\ &= \frac{q}{1 - \mu_+} \left(\left(\frac{1 - q^r}{1 - q} \right) - \mu_+^{R-r+1} \left(\frac{1 - (q\mu_+)^r}{1 - q\mu_+} \right) \right) \end{aligned} \tag{6.3}$$

for all $\mu_+, q, q\mu_+ \neq 1$. Combining (6.2) and (6.3), and using Lemmas 5.1 and 5.4, we deduce that

$$\begin{aligned} E_r(S) &= \frac{1 - \mu_+^{R-r+1}}{1 - \mu_+} + q \left(\frac{1 - q^r}{1 - q} \right) + \frac{q\mu_-}{1 - \mu_+} \left(\left(\frac{1 - q^r}{1 - q} \right) - \mu_+^{R-r+1} \left(\frac{1 - (q\mu_+)^r}{1 - q\mu_+} \right) \right) \\ &= \frac{1 - \mu_+^{R-r+1}}{1 - \mu_+} + q \left(\frac{1 - q^r}{1 - q} \right) \left(\frac{1 - \mu_+ + \mu_-}{1 - \mu_+} \right) + \frac{q\mu_-}{1 - \mu_+} \left(-\mu_+^{R-r+1} \left(\frac{1 - (q\mu_+)^r}{1 - q\mu_+} \right) \right) \\ &= \frac{1}{1 - \mu_+} \left(1 + q \left(\frac{1 - q^r}{1 - q} \right) (1 - \mu_+ + \mu_-) - \mu_+^{R-r+1} \left(1 + q\mu_- \left(\frac{1 - (q\mu_+)^r}{1 - q\mu_+} \right) \right) \right) \end{aligned}$$

for all $\mu_+, q, q\mu_+ \neq 1$. Using Lemma 5.2 also gives $1 - \mu_+ + \mu_- = 1 - p$ and

$$1 + q\mu_- \left(\frac{1 - (q\mu_+)^r}{1 - q\mu_+} \right) = \frac{1 - pq(1 + (\mu - 1)(\mu pq)^r)}{1 - \mu pq}$$

for all $\mu pq \neq 1$. In conclusion,

$$E_r(S) = \frac{1}{1 - \mu p} \left(1 + q \left(\frac{1 - q^r}{1 - q} \right) (1 - p) - (\mu p)^{R - r + 1} \left(\frac{1 - pq(1 + (\mu - 1)(\mu pq)^r)}{1 - \mu pq} \right) \right)$$

for all $\mu p, \mu pq, q \neq 1$, which completes the proof of Theorem 3.2.

7. Proof of Theorem 3.3 and (3.4) (second moment)

To prove Theorem 3.3, we follow the same strategy as for Theorem 3.2 using also that the numbers of infected vertices in disjoint subtrees are independent. Taking the square in (6.1),

$$\begin{aligned} S^2 &= S(x_r)^2 + 2S(x_r) \sum_{i=1}^{k} S(x_{r-i} \setminus x_{r-i+1}) + \left(\sum_{i=1}^{k} S(x_{r-i} \setminus x_{r-i+1}) \right)^2 \\ &= S(x_r)^2 + 2S(x_r) \sum_{i=1}^{k} S(x_{r-i} \setminus x_{r-i+1}) \\ &\qquad + \sum_{i=1}^{k} S(x_{r-i} \setminus x_{r-i+1})^2 + \sum_{i,j \in \{1,2,\ldots,k\}, i \neq j} S(x_{r-i} \setminus x_{r-i+1}) S(x_{r-j} \setminus x_{r-j+1}), \end{aligned}$$

then conditioning on the event $D = k$ and using independence of the random variables in (5.2) (because they represent the number of infected vertices in disjoint subtrees), we obtain

$$\begin{aligned} E_r(S^2) = \sum_{k=0}^{r} \Bigg(& E(S(x_r)^2) + 2E(S(x_r)) \sum_{i=1}^{k} E(S(x_{r-i} \setminus x_{r-i+1})) \\ & + \sum_{i=1}^{k} E(S(x_{r-i} \setminus x_{r-i+1})^2) \\ & + \sum_{i,j \in \{1,2,\ldots,k\}, i \neq j} E(S(x_{r-i} \setminus x_{r-i+1}))\, E(S(x_{r-j} \setminus x_{r-j+1})) \Bigg) P(D = k). \end{aligned} \tag{7.1}$$

Then, using Lemmas 5.4 and 5.5, and recalling from (5.1) that

$$P(D = k) = q_k \quad \text{for all} \quad k = 0, 1, \ldots, r,$$

the right-hand side of (7.1) becomes

$$\begin{aligned} E_r(S^2) = \sum_{k=0}^{r} \Bigg(& \mu_{2,r} + 2\mu_{1,r} \ \sum_{i=1}^{k} (1 + \mu_- \mu_{1,r-i+1}) \\ & + \sum_{i=1}^{k} (1 + 2\mu_- \mu_{1,r-i+1} + \mu_- \mu_{2,r-i+1} + (\sigma_-^2 + \mu_-^2 - \mu_-)(\mu_{1,r-i+1})^2) \\ & + \sum_{i,j \in \{1,2,\ldots,k\}, i \neq j} (1 + \mu_- \mu_{1,r-i+1})(1 + \mu_- \mu_{1,r-j+1}) \Bigg) q_k. \end{aligned}$$

This completes the proof of the theorem.

To simplify the previous expression for the second moment when the percolation process is subcritical $\mu_+ = \mu p < 1$ and the LAN is infinite, and prove (3.4), we first observe that, by the monotone convergence theorem, the second moment on the infinite tree is equal to the limit of the second moment on the finite tree as the radius $R \to \infty$. The reason why the expression simplifies in the infinite tree limit is because the terms $\mu_{1,j}$ and $\mu_{2,j}$ no longer depend on the index j, which is due to the fact that they now represent the first and second moments of the number of infected vertices on infinite subtrees that are identically distributed. More precisely, taking the limit as $R \to \infty$ in (3.3) and using Lemma 5.4, we get

$$E(S(x_r)) = \mu_{1,r} = \frac{1}{1 - \mu_+} \quad \text{and} \quad E(S(x_r)^2) = \mu_{2,r} = \frac{1}{(1 - \mu_+)^2} \left(1 + \frac{\sigma_+^2}{1 - \mu_+}\right) \tag{7.2}$$

for all $\mu_+ \neq 1$. This, together with Lemma 5.5, implies that

$$\begin{aligned} E(S(x_{r-i} \setminus x_{r-i+1})) &= 1 + \mu_- \mu_{1,r-i+1} = 1 + \frac{\mu_-}{1 - \mu_+} \\ E(S(x_{r-i} \setminus x_{r-i+1})^2) &= 1 + 2\mu_- \mu_{1,r-i+1} + \mu_- \mu_{2,r-i+1} + (\sigma_-^2 + \mu_-^2 - \mu_-)(\mu_{1,r-i+1})^2 \\ &= 1 + \frac{2\mu_-}{1 - \mu_+} + \frac{\mu_-}{(1 - \mu_+)^2} \left(1 + \frac{\sigma_+^2}{1 - \mu_+}\right) + \frac{\sigma_-^2 + \mu_-^2 - \mu_-}{(1 - \mu_+)^2} \\ &= 1 + \frac{2\mu_-}{1 - \mu_+} + \frac{\sigma_-^2 + \mu_-^2}{(1 - \mu_+)^2} + \frac{\sigma_+^2 \mu_-}{(1 - \mu_+)^3} \end{aligned} \tag{7.3}$$

for all $\mu_+ \neq 1$. Using that, in the limit as $R \to \infty$, the terms in the two sums over i and the terms in the sum over $i \neq j$ in equation (7.1) are constant given by (7.2) and (7.3), we obtain

$$\begin{aligned} E_r(S^2) \ &= \ \frac{1}{(1 - \mu_+)^2} \left(1 + \frac{\sigma_+^2}{1 - \mu_+}\right) + \frac{2}{1 - \mu_+} \left(1 + \frac{\mu_-}{1 - \mu_+}\right) E(D) \\ & \quad + \left(1 + \frac{2\mu_-}{1 - \mu_+} + \frac{\sigma_-^2 + \mu_-^2}{(1 - \mu_+)^2} + \frac{\sigma_+^2 \mu_-}{(1 - \mu_+)^3}\right) E(D) + \left(1 + \frac{\mu_-}{1 - \mu_+}\right)^2 E(D(D-1)) \\ &= \ \frac{1}{(1 - \mu_+)^2} \left(1 + \frac{\sigma_+^2}{1 - \mu_+}\right) + \left(1 + \frac{2(1 + \mu_-)}{1 - \mu_+} + \frac{2\mu_- + \sigma_-^2 + \mu_-^2}{(1 - \mu_+)^2} + \frac{\sigma_+^2 \mu_-}{(1 - \mu_+)^3}\right) E(D) \\ & \quad + \left(1 + \frac{\mu_-}{1 - \mu_+}\right)^2 E(D(D-1)) \end{aligned}$$

for all $\mu_+, q \neq 1$. Then, using Lemma 5.1, we get

$$\begin{aligned} E_r(S^2) \ &= \ \frac{1}{(1-\mu_+)^2} \ \bigg(1 + \frac{\sigma_+^2}{1-\mu_+}\bigg) \\ &+ \ \bigg(1 + \frac{2(1+\mu_-)}{1-\mu_+} + \frac{2\mu_- + \sigma_-^2 + \mu_-^2}{(1-\mu_+)^2} + \frac{\sigma_+^2 \mu_-}{(1-\mu_+)^3}\bigg) \ q \bigg(\frac{1-q^r}{1-q}\bigg) \\ &+ \ \bigg(1 + \frac{\mu_-}{1-\mu_+}\bigg)^2 \ \frac{2q^2(1 - rq^{r-1} + (r-1)q^r)}{(1-q)^2} \end{aligned}$$

for all $\mu_+, q \neq 1$, and finally Lemmas 5.2 and 5.3,

$$\begin{aligned} E_r(S^2) \ &= \ \frac{1}{(1-\mu p)^2} \ \bigg(1 + \frac{p(1-p)\mu + p^2\sigma^2}{1-\mu p}\bigg) \\ &+ \ \bigg(1 + \frac{2(1+(\mu-1)p)}{1-\mu p} + \frac{2(\mu-1)p + p(1-p)(\mu-1) + p^2\sigma^2 + (\mu-1)^2p^2}{(1-\mu p)^2} \\ &\qquad\qquad + \ \frac{(p(1-p)\mu + p^2\sigma^2)(\mu-1)p}{(1-\mu p)^3}\bigg) \ q \bigg(\frac{1-q^r}{1-q}\bigg) \\ &+ \ \bigg(1 + \frac{(\mu-1)p}{1-\mu p}\bigg)^2 \ \frac{2q^2(1 - rq^{r-1} + (r-1)q^r)}{(1-q)^2} \end{aligned}$$

for all $\mu p, q \neq 1$, which completes the proof of (3.4).

8. Proof of Theorem 3.4 (exponential decay)

To prove exponential decay of the diameter of the cluster of infected vertices, the idea is to study the process that keeps track of the random number of infected vertices at distance n from the highest infected vertex in the tree. More precisely, on the event $D = k$, we let

$$X_n = \operatorname{card}(\mathscr{C}_n) \quad \text{where} \quad \mathscr{C}_n = \{x \in \mathscr{C} : d(x, x_{r-k}) = n\}.$$

Recall that x_{r-k} is the unique vertex along the path connecting the source of the infection and the root of the tree that is at distance k from the source of the infection. The next lemma shows that, in the subcritical phase $\mu p < 1$, the expected value of the process decays exponentially.

Lemma 8.1. *Given that the infection starts at distance r from the root,*

$$E_r(X_{n+1}) = \begin{cases} \mu p \, E(X_n) + (1-p) \, q^{n+1} & \textit{for} \quad n < r \\ \mu p \, E(X_n) & \textit{for} \quad n \geq r. \end{cases}$$

Proof. Recall that each vertex produces μ offspring on average and that each of the offspring of an infected vertex is infected with probability p, from which it follows that each infected vertex has μp infected offspring on average. Now, given $D = k$, the process is conditioned so that

$$x_{r-k} \in \mathscr{C}_0, \quad x_{r-k+1} \in \mathscr{C}_1, \quad \ldots \quad x_{r-1} \in \mathscr{C}_{k-1} \quad \text{and} \quad x_r \in \mathscr{C}_k$$

are infected so, until generation $k-1$, the vertices in $\mathscr{C}_n$ have μp infected offspring on average except for vertex x_{r-k+n} that has $(\mu-1)p+1$ infected offspring on average. This implies that

$$E_r(X_{n+1}\,|\,X_n, D=k) = \left\{\begin{array}{cll} \mu p(X_n-1)+(\mu-1)p+1 & \text{for} & n<k \\ \mu p X_n & \text{for} & n \geq k. \end{array}\right.$$

Conditioning on the random variable D, we deduce that

$$\begin{aligned} E_r(X_{n+1}\,|\,X_n) &= \sum_{k=0}^{\infty} E_r(X_{n+1}\,|\,X_n, D=k) P_r(D=k) \\ &= \sum_{k=0}^{n} (\mu p X_n)\, P_r(D=k) + \sum_{k=n+1}^{\infty} (\mu p(X_n-1)+(\mu-1)p+1)\, P_r(D=k) \\ &= \sum_{k=0}^{n} (\mu p X_n)\, P_r(D=k) + \sum_{k=n+1}^{\infty} (\mu p X_n + 1 - p)\, P_r(D=k) \\ &= \sum_{k=0}^{\infty} (\mu p X_n)\, P_r(D=k) + (1-p) \sum_{k=n+1}^{\infty} P_r(D=k) = \mu p X_n + (1-p) P_r(D>n). \end{aligned}$$

Recalling the probability mass function of D, we deduce that

$$E_r(X_{n+1}) = E(E(X_{n+1}\,|\,X_n)) = \left\{\begin{array}{cll} \mu p\, E(X_n) + (1-p)\, q^{n+1} & \text{for} & n<r \\ \mu p\, E(X_n) & \text{for} & n \geq r. \end{array}\right.$$

This completes the proof. □

It follows from the lemma that, for all $n \leq r$,

$$\begin{aligned} E_r(X_n) &= (\mu p)\, E_r(X_{n-1}) + (1-p)\, q^n \\ &= (\mu p)^2 E_r(X_{n-2}) + (\mu p)(1-p)\, q^{n-1} + (1-p)\, q^n \\ &= (\mu p)^3 E_r(X_{n-3}) + (\mu p)^2 (1-p)\, q^{n-2} + (\mu p)(1-p)\, q^{n-1} + (1-p)\, q^n \\ &= (\mu p)^n E_r(X_0) + (1-p)((\mu p)^{n-1}\, q + (\mu p)^{n-2}\, q^2 + \cdots + (\mu p)\, q^{n-1} + q^n) \\ &\leq (\mu p)^n E_r(X_0) + (\mu p)^{n-1}\, q + (\mu p)^{n-2}\, q^2 + \cdots + (\mu p)\, q^{n-1} + q^n. \end{aligned}$$

Then, using that $E_r(X_0)=1$, we get

$$E_r(X_n) \leq \sum_{k=0}^{n} (\mu p)^{n-k} q^k = (\mu p)^n \sum_{k=0}^{n} \left(\frac{q}{\mu p}\right)^k = \frac{1-(q/\mu p)^{n+1}}{1-(q/\mu p)}\ (\mu p)^n$$

for all $n \leq r$ and $\mu q \neq q$. Observing also that, for all $n > r$,

$$E_r(X_n) \leq (\mu p)\, E_r(X_{n-1}) \leq (\mu p)^2 E_r(X_{n-2}) \leq \cdots \leq (\mu p)^{n-r} E_r(X_r)$$

we conclude that, for all $n > r$, the diameter exceeds $2n$ with probability

$$\begin{aligned} P_r(\operatorname{diam}(\mathscr{C}) \geq 2n) \ &\leq \ P_r(X_n > 0) = \sum_{k=1}^{\infty} P_r(X_n = k) \\ &\leq \sum_{k=1}^{\infty} k\, P_r(X_n = k) = E_r(X_n) \leq (\mu p)^{n-r} E_r(X_r) \\ &\leq (\mu p)^{n-r} \ \frac{1-(q/\mu p)^{r+1}}{1-(q/\mu p)} \ (\mu p)^r = \frac{1-(q/\mu p)^{r+1}}{1-(q/\mu p)} \ (\mu p)^n \end{aligned}$$

for all $\mu q \neq q$. This completes the proof of Theorem 3.4.

Acknowledgements. This work is partially supported by National Science Foundation under grant CNS-2000792. The authors would like to thank two anonymous referees for their careful reading and comments that helped improve the preliminary version of this work.

References

[1] I. Aldasoro, L. Gambacorta, P. Giudici and T. Leach, The drivers of cyber risk (2020). Available at https://www.bbc.com/news/technology-59612917 (accessed 06 December 2021).
[2] Z. Amin, A practical road map for assessing cyber risk. *J. Risk Res.* **22** (2019) 32–43.
[3] Y. Antonio and S. Indratno, Cyber insurance rate making based on markov model for regular networks topology. *J. Phys.* **1752** (2021) 012002.
[4] Australian Cyber Security Centre, Restricting Administrative Privileges (2021). Available at https://www.cyber.gov.au/acsc/view-all-content/publications/restricting-administrative-privileges (accessed 16 December 2021).
[5] R. Betterley, Cyber privacy insurance market survey: a tough market for larger insureds, but smaller insureds finding eager insurers (2016). Available at http://betterley.com/samples/cpims16_nt.pdf (accessed 12 December 2021).
[6] Cybersecurity and Infrastructure Security Agency, Securing network infrastructure devices (2018). Available at https://www.cisa.gov/uscert/ncas/tips/ST18-001 (accessed 12 December 2021).
[7] Cynet, 2022 Survey of CISOs with small cyber security teams (2022). Available at https://go.cynet.com/hubfs/2022%20CISO%20Survey%20of%20Small%20Cyber%20Security%20Teams.pdf (accessed 08 August 2022).
[8] Department of Homeland Security, The increasing threat to network infrastructure devices and recommended mitigations (2016). Available at https://cyber.dhs.gov/assets/report/ar-16-20173.pdf (accessed: 16 November 2021).
[9] Department of Justice: Southern District of New York, California man pleads guilty to hacking websites for the Combating Terrorism Center at West Point and the New York City Comptroller (2018). Available at https://www.justice.gov/usao-sdny/pr/california-man-pleads-guilty-hacking-websites-combating-terrorism-center-west-point-and (accessed: 21 November 2021).
[10] M. Eling and K. Jung, Copula approaches for modeling cross-sectional dependence of data breach losses. *Insur. Math. Econ.* **82** (2018) 167–180.
[11] M. Eling, K. Jung and J. Shim, Unraveling heterogeneity in cyber risks using quantile regressions. *Insur. Math. Econ.* **104** (2022) 222–242.
[12] M. Eling and J. Wirfs, Modelling and management of cyber risk. *Int. Actuar. Assoc. Life Section* (2015).
[13] M. Eling and J. Wirfs, What are the actual costs of cyber risk events? *Eur. J. Oper. Res.* **272** (2019) 1109–1119.
[14] S. Farkas, O. Lopez and M. Thomas, Cyber claim analysis using generalized Pareto regression trees with applications to insurance. *Insur. Math. Econ.* **98** (2021) 92–105.
[15] Federal Bureau of Investigation, Indicators of compromised associated with Diavol (2022). Available at https://www.ic3.gov/Media/News/2022/220120.pdf (accessed: 03 December 2021).
[16] H. Ferraiolo, D.A. Cooper, A.R. Regenscheid, K. Scarfone and M.P. Souppaya, Best practices for privileged user PIV authentication (2016). Available at https://www.nist.gov/publications/best-practices-privileged-user-piv-authentication?pub_id=920826 (accessed 25 August 2021).
[17] P. Georgi, L. Morrow and T. Highfill, Updated and expanded small business statistics: Wages, employment, and gross output by industry and enterprise size, 2012–2017 (2021). Available at https://apps.bea.gov/scb/2021/11-november/pdf/1121-small-business.pdf (accessed 16 December 2021).
[18] H. Herath and T. Herath, Copula-based actuarial model for pricing cyber-insurance policies, *Insur. Mark. Compan.* **2** (2011) 7–20.
[19] P. Jevtić and N. Lanchier, Dynamic structural percolation model of loss distribution for cyber risk of small and medium-sized enterprises for tree-based LAN topology. *Insur. Math. Econ.* **91** (2020) 209–223.
[20] P. Jevtić and N. Lanchier, Systems and methods for a simulation program of a percolation model for the loss distribution caused by a cyber attack. uS Patent No. 11,354,752 (2022).

[21] K. Jung, Extreme data breach losses: an alternative approach to estimating probable maximum loss for data breach risk. *North Am. Actuar. J.* **25** (2021) 580–603.

[22] I. Kovačević, S. Groš and A. Derek, Automatically generating models of IT systems. *IEEE Access* **10** (2022) 13536–13554.

[23] Marsh, U.K. cyber insurance trends 2020 (2021). Available at https://www.marsh.com/uk/services/cyber-risk/insights/uk-cyber-insurance-trends-2020.html (accessed 16 December 2021).

[24] N. Mhaskar, M. Alabbad and R. Khedri, A formal approach to network segmentation. *Comput. Secur.* **103** (2021) 102162.

[25] T.J. Moore and J.-H. Cho, Applying percolation theory, in Cyber Resilience of Systems and Networks Springer (2019), pp. 107–133.

[26] National Institute of Standards and Technology, Intrusion (2021). Available at https://csrc.nist.gov/glossary/term/intrusion (accessed 16 December 2021).

[27] National Institute of Standards and Technology, Least privilege (2021). Available at https://csrc.nist.gov/glossary/term/least_privilege (accessed 04 December 2021).

[28] National Security Agency, Defend Privileges and Accounts (2019). Available at https://media.defense.gov/2019/Sep/09/2002180330/-1/-1/0/Defend%20Privileges%20and%20Accounts%20-%20Copy.pdf (accessed: 26 August 2021).

[29] National Security Agency, Segment networks and deploy application-aware defenses (2019). Available at https://media.defense.gov/2019/Sep/09/2002180325/-1/-1/0/Segment%20Networks%20and%20Deploy%20Application%20Aware%20Defenses%20-%20Copy.pdf (accessed 09 December 2021).

[30] NetDiligence, Cyber Claims Study (2019). Available at https://dev.networkstandard.com/wp-content/uploads/2020/05/2019_NetD_Claims_Study_Report_1.2.pdf (accessed: 10 December 2021).

[31] S. Romanosky, L. Ablon, A. Kuehn and T. Jones, Content analysis of cyber insurance policies: how do carriers price cyber risk? *J. Cybersecur.* **5** (2019) 1–19.

[32] SonicWall, Mid-Year Update: SonicWall Cyber Threat Report (2021). Available at https://www.sonicwall.com/2021-cyber-threat-report/ (accessed 18 December 2021).

[33] The Institute of Risk Management, Cyber risk and risk management (2018). Available at https://www.theirm.org/what-we-say/thought-leadership/cyber-risk/ (accessed 11 December 2021).

[34] U.S. Government Accountability Office, Cyber Insurance: insurers and policyholders face challenges in an evolving market (2021). Available at https://www.gao.gov/products/gao-21-477 (accessed 14 December 2021).

[35] U.S. Securities and Exchange Commission, IT specialist settles charges of insider trading on hacked nonpublic information (2016). Available at https://www.sec.gov/news/pressrelease/2016-256.html (accessed 04 December 2021).

[36] U.S. Small Business Administration, Table of small business size standards matched to North American industry classification system codes (2019). Available at https://www.sba.gov/sites/default/files/2019-08/SBA%20Table%20of%20Size%20Standards_Effective%20Aug%2019%2C%202019_Rev.pdf (accessed: 03 December 2021).

[37] Verizon, 2018 Verizon Data Breach Investigations Report (2018). Available at https://www.verizon.com/business/resources/reports/dbir/ (accessed 16 December 2021).

[38] Verizon, 2021 Verizon Data Breach Investigations Report (2021). Available at https://www.verizon.com/business/resources/reports/dbir/ (accessed 15 December 2021).

[39] N. Wagner, C.Ş. Şahin, M. Winterrose, J. Riordan, J. Pena, D. Hanson and W.W. Streilein, Towards automated cyber decision support: a case study on network segmentation for security, in *2016 IEEE Symposium Series on Computational Intelligence.* IEEE (2016) 1–10.

[40] H. Wang, Z. Chen, J. Zhao, X. Di and D. Liu, A vulnerability assessment method in industrial internet of things based on attack graph and maximum flow. *IEEE Access* **6** (2018) 8599–8609.

[41] S. Wang, Z. Zhang and Y. Kadobayashi, Exploring attack graph for cost-benefit security hardening: a probabilistic approach. *Comput. Secur.* **32** (2013) 158–169.

[42] World Economic Forum, Global cybersecurity outlook 2022 (2022). Available at https://www3.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2022.pdf (accessed 16 August 2022).

[43] X. Xie, C. Lee and M. Eling, Cyber insurance offering and performance: an analysis of the U.S. cyber insurance market. *Geneva Papers on Risk and Insurance- Issues and Practice* **45** (2020) 690–736.

[44] M. Xu and L. Hua, Cybersecurity insurance: modeling and pricing. *North Am. Actuar. J.* **23** (2019) 220–249.

[45] P. Żebrowski, A. Couce-Vieira and A. Mancuso, A Bayesian framework for the analysis and optimal mitigation of cyber threats to cyber-physical systems. *Risk Anal* (2022). https://doi.org/10.1111/risa.13900

[46] G. Zeller and M. Scherer, A comprehensive model for cyber risk based on marked point processes and its application to insurance. *Eur. Actuar. J.* **12** (2022) 33–85.

[47] X. Zhang, M. Xu, J. Su and P. Zhao, Structural models for fog computing based internet of things architectures with insurance and risk management applications. *Eur. J. Oper. Res.* (2022). https://doi.org/10.1016/j.ejor.2022.07.033