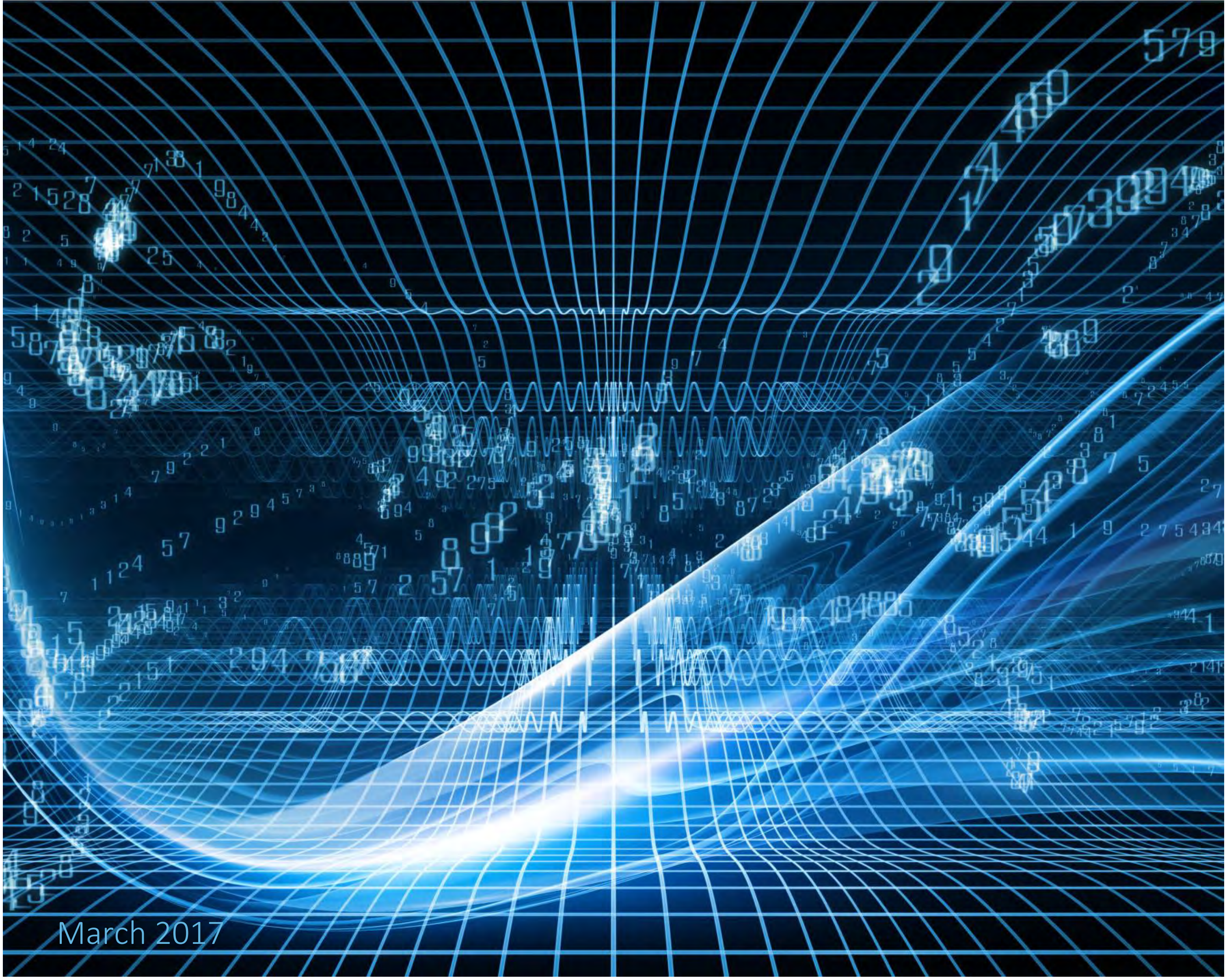
SOCIETY OF
ACTUARIES

Cybersecurity Insurance: Modeling and Pricing



March 2017

Cybersecurity Insurance: Modeling and Pricing

Maochao Xu
Department of Mathematics
Illinois State University, USA

Lei Hua, ASA
Division of Statistics
Northern Illinois University, USA

Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information.

Contents

Cybersecurity Insurance: Modeling and Pricing

Cybersecurity risk has attracted considerable attention in recent decades. However, modeling the cybersecurity risk is still in its infancy, mainly because of its unique characteristics. In this paper, we develop a framework for modeling and pricing cybersecurity risk. The proposed model consists of three components: epidemic models, loss functions and premium strategies. We study the dynamic upper bounds for the infection probabilities based on both Markov and non-Markov models. A simulation approach is proposed to compute the premium for the cybersecurity risk for practical use. The effects of different infection distributions and dependence among infection processes on the losses are studied as well.

Acknowledgments

The authors are very grateful to the members of the Project Oversight Group for their insightful and constructive comments, which led to this improved version of the paper. This research is supported by the Society of Actuaries (SOA). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of SOA.

Section 1: Introduction and Motivation

Cybersecurity insurance, which is designed to transfer the economic losses associated with network and computer incidents to a third party, has attracted much attention from professionals and researchers recently. The problem has been reiterated in the workshop and roundtable of the US Department of Homeland Security's (DHS) National Protection and Programs Directorate (NPPD) (2012, 2013) [8]. The eighth Emerging Risks Survey by the Society of Actuaries indicates that, according to risk managers, cybersecurity risk is the greatest emerging risk. The cybersecurity insurance market continues to broaden, and more and more small to midsize companies show interest in cybersecurity insurance. Many companies are seeking coverages for the value of data loss, lost revenue due to loss of data or downtime, legal expenses for damage to the third party, notification of potentially affected customers and regulatory fines and penalties [13, 2]. It is estimated that the annual gross written premium is $3.25 billion for 2016, compared to $2.75 billion in 2015 [2]. However, the contributions to modeling the cybersecurity risk in the literature are largely descriptive, which is mainly because cyber risk is very different from the traditional risks covered by indemnity insurance. The significant property that distinguishes cyber risk from conventional risk is that information and communication technology (ICT) resources are interconnected in a network, and therefore the analysis of risk and its related potential losses needs to take into account the network topology. Further, if ICT resources are hijacked, then benign sources (e.g., computers) may become threats to other sources [4].

Traditionally, pricing insurance products relies on actuarial tables constructed from historical records. Unlike traditional insurance policies, however, cybersecurity insurance has no standard scoring systems or actuarial tables for rate making. Cybersecurity risks are relatively new, and the data about security breaches and losses do not exist or exist only in small quantities. This difficulty may be further exacerbated by the reluctance of organizations to reveal details of security breaches due to loss of market share, loss of reputation and so forth. Pricing cybersecurity risks is still a challenging question, although many insurance companies do provide cybersecurity insurance products. The insurers tend to increase the premiums for the larger companies, and the coverage may be limited and very expensive for the companies without good cybersecurity protection [2].

The literature reveals several efforts to study the cybersecurity risk via mathematical models. For example, Gordon et al. [11] discuss a general framework on pricing and the adverse selection issues of cyber insurance, and they propose a four-step cyber risk decision plan. Bohme and Kataria [3] consider the correlation between cyber risks and use the beta-binomial and one-factor latent risk model for modeling purposes. In particular, Bohme and Kataria discuss the internal correlation of cybersecurity risk within a firm and the global correlation of cybersecurity risk at the global level. Bohme and Schwartz [4] discuss a framework for dealing with the specific properties of cybersecurity risk, including interdependent security, correlated risk and information asymmetry. They also present a survey on existing models of cybersecurity insurance. A discrepancy between informal arguments in favor of cybersecurity insurance as a tool to improve network security is discussed there.

A Bayesian brief network approach is proposed in Mukhopadhyay et al. [17] for modeling the cybersecurity risk. They use the multivariate Gaussian copula to model the joint distribution and conditional distribution of each node on the network. The premiums are calculated as a function of expected value of claim severity. Herath and Herath [12] propose a copula-based actuarial model for pricing cybersecurity risk, where they model three risk variables: occurrence of the event, the time of payment and the amount of payment. The premiums for first-party losses due to epidemic attacks are calculated by using three types of insurance policy models: policy with a zero deductible, policy with deductibles and policy with coinsurance and limits. Schwartz and Sastry [22] present a framework for managing cybersecurity risk in a large-scale interdependent network. They consider the cyber insurers as strategic players, and they derive the solution for user optimal security in environments with and without cyber insurers.

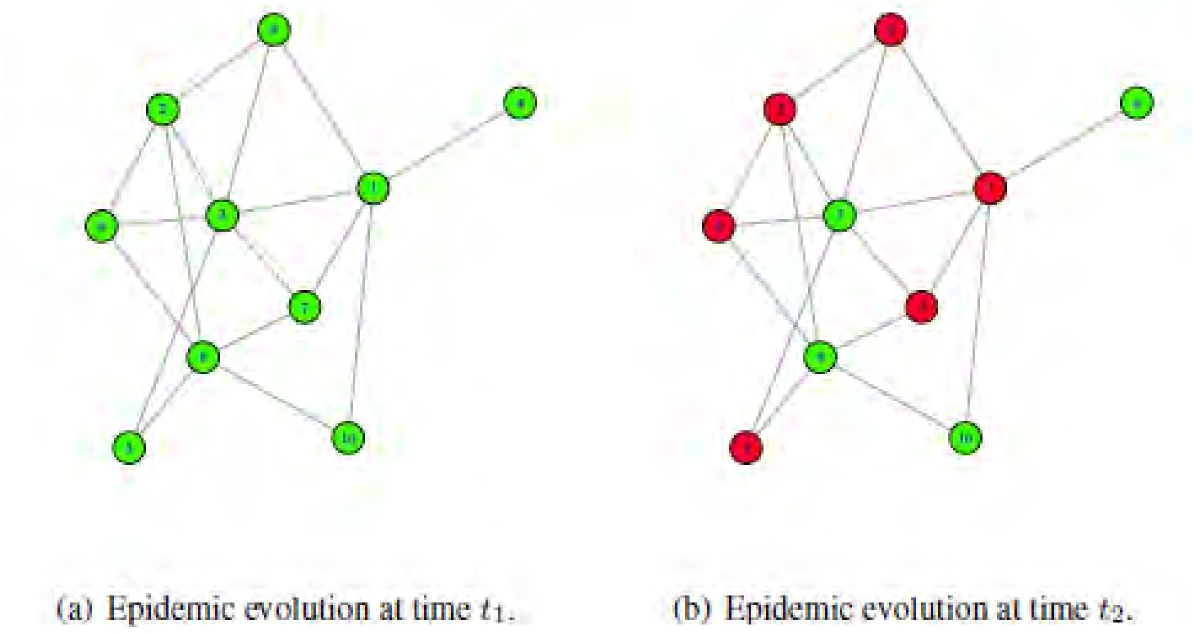
Yang and Lui [29] use the Bayesian network game to model the security investment, where the network externality effect is considered. It is shown there that nodes with more degrees are more likely to be infected and have higher chances to be affected by others' decisions. One may refer to Kosub [16] and Eling and Schnell [10] for comprehensive reviews on cybersecurity risk modeling and management of cybersecurity risk.

Our work in this paper is different from those in the literature in the following aspects: (1) We use stochastic processes (Markov and non-Markov) to describe the dynamics of epidemic spreading over time, while most of the aforementioned works are static. (2) We propose to use the copula to capture the dependence among the time-to-infection distributions, whereas in the literature it is often assumed to be independent. (3) We suggest using Monte Carlo simulations to evaluate the security level of networks, and the security level includes the number of incidents, the infection probabilities of nodes and the total losses.

To further motivate our study, assume that a company whose ICT resources have the network structure in Figure 1 wants to buy cyber insurance, where the nodes represent computers (and/or servers). It is seen at time t_1 that none of the computers is infected. However, at time t_2, six computers are infected. For an insurance company that wants to offer cybersecurity insurance policies, the key step is to understand the evolution of epidemic spreading over the network, as the infection will cause losses in practice. It is also important for the insurance company to know the total loss during a specific time period, as premiums are determined based on the losses.

Figure 1

Cyber epidemic spreading over network for a company with 10 computers/servers at time t_1 and t_2.
The red dots represent the infected computers.



(a) Epidemic evolution at time t_1.

(b) Epidemic evolution at time t_2.

The purpose of this paper is to establish a robust and systematic approach for modeling and pricing cybersecurity risks. We make the following contributions:

- We model the evolution of cybersecurity risk via both Markov and non-Markov models. In particular, we propose to use copula to model the dependence among risks, since it is very flexible in accounting for nonlinear dependence.

- We study the dynamic upper bounds for the infection probabilities of nodes over time. We show that the independence assumption among risks would lead to upper bounds for the infection probabilities.
- We propose to use Monte Carlo simulations to study the pricing strategies in practice. Specifically, we simulate the evolution of epidemic spreading over a network, and hence, derive three key quantities: the number of incidents, the infection probabilities and the total losses.

The rest of the paper is organized as follows. Section 2 discusses the framework for modeling cybersecurity risks by using a renewal process. In Section 3, the evolution of epidemic spreading is modeled by both Markov and non-Markov models, and some upper bounds are discussed. Section 4 presents the simulation and pricing strategies. In the last section, we conclude our results and present some points for discussion.

Section 2: Models for Cybersecurity Risks

Assume that a company has a network that could be described as an undirected graph $\Gamma = (V; \mathbb{E})$, where V is the node set and $\mathbb{E}$ is the edge set. Note that Γ abstracts the network structure according to which the cyber attacks take place (e.g., malware spreading), where $(u, v) \in \mathbb{E}$ abstracts that nodes u and v can attack each other (undirected graph). In principle, Γ can range from a complete graph (i.e., any $u \in V$ can attack any $v \in V$) to any specific graph structure. Denote by $A = (a_{vu})$ the adjacency matrix of Γ, where $a_{vu} = 1$ if and only if $(u, v) \in \mathbb{E}$, and $a_{vu} = 0$ otherwise. Note that the problem setting naturally implies $a_{vv} = 0$. Denote by $deg(v)$ the degree of node v, and $N = |V|$ the total number of nodes. Node $v \in V$ is either *secure* (but vulnerable to attacks) or *infected* (and can attack other nodes) at any time $t = 0,1,\cdots$. The status of this network at time t can be represented as

$$(I_1(t), \cdots, I_N(t))$$

where $I_v(t) = 1$ represents that node v is in infection status at time t, while $I_v(t) = 0$ represents that node v is secure at time t. The infection probability vector is denoted by

$$\boldsymbol{p}^T(t) = (p_1(t), \cdots, p_N(t)),$$

where $p_j(t) = P\big(I_j = 1\big)$, for $j = 0,1, \cdots, N$.

We consider two threats faced by each node: (1) threats outside the network (i.e., node v is infected because it is attacked or its user visits a malicious website); and (2) threats inside the network (i.e., node v is infected, then node v attacks its neighbors). We also assume that if node v is infected, it will be repaired or cleaned to return to secure status. Extensive work has been done modeling the epidemic spreading over the network in the communities of physics and cybersecurity. One may refer to [1, 26, 28, 27, 19] for comprehensive discussions and reviews on this topic.

For illustration purposes, consider the scenario in Figure 2. Node v is secure at time $T = 0$, and the first infection occurs at time $T = t_1$. The infection would incur two types of losses: (1) loss caused by the infection, such as information stolen, data damaged, records exposed and first-party legal costs; and (2) loss caused by restoring the node to secure status. The first type of loss is modeled by a random cost $\eta_v(L_{v,1})$, where $L_{v,1}$ means the loss of information (e.g., data damaged), and it can also be used to model the first-party legal cost. The second type of loss is related to the duration of out-of-service (or repair), and it is modeled by a cost function $C_v(R_{v,1})$, where $R_{v,1}$ is the duration of out-of-service. At time $T = t_2$, node v is secure but vulnerable to attacks, and it will be infected at times t_3 and t_5 again. Therefore, for node v, the loss cumulative to time t can be represented as

$$S_v(t) = \sum_{i=1}^{M_v(t)} [\eta_v(L_{v,i}) + C_v(R_{v,i})] \,,$$

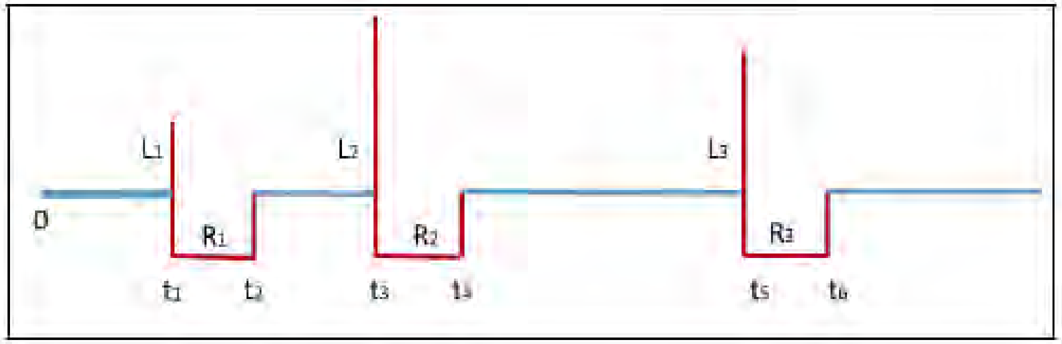
where $\eta_v(\cdot)$ represents the cost due to infection, and $C_v(\cdot)$ represents the cost function associated with the time length $R_{v,i}$ of out-of-service. For each node v, in fact, it is a renewal reward process. The total loss faced by the company during $(0, t]$) is

$$S(t) = \sum_{v=1}^{N} S_v(t) = \sum_{v}^{N} \sum_{i=1}^{M_v(t)} [\eta_v(L_{v,i}) + C_v(R_{v,i})] \qquad (2.1)$$

where $M_v(t)$ is the total number of infections of node v up to time t. Eq. (2.1) shows that the key quantity is the infection vector $(I_1(t), \cdots, I_N(t))$, which requires the epidemic theory [25]. In the next section, we discuss the epidemic models that can be used for modeling cybersecurity risks.

Figure 2

Cybersecurity risk for node v



Section 3: Epidemic Spreading Models

In this section, we discuss two epidemic models for modeling the cybersecurity risks. In particular, we study the dynamic upper bounds for infection probabilities over time, which may be used as conservative estimates for pricing purposes.

3.1 Markov Model

For this model, we assume the recovering process of any infected node v is a Poisson process with δ_v. The infection process per link is also a Poisson process with β due to the infected neighbors inside the network. We also assume that for any infected node, it may be infected with a Poisson ϵ_v due to the threat outside the network. The infection processes and recovering processes are assumed to be independent. This model, in fact, is known as ϵ-SIS model [26] or push-pull model [27] in the literature. For any node v, the infection and recovery processes form the following Markov process:

$$I_v(t) = 0 \to 1 \text{ at rate } \beta \sum_{j=1}^{n} a_{vj} I_j(t) + \epsilon_v;$$

$$I_v(t) = 1 \to 0 \text{ at rate } \delta_v.$$

The following result provides a dynamic upper bound for infection probabilities, which may be used as a conservative estimate for infections over the network.

Theorem 3.1 Let $Q = \operatorname{diag}\big((\beta\delta_v)/(\delta_v + \epsilon_v)\big)A - \operatorname{diag}(\delta_v + \epsilon_v)$. Then the dynamic upper bound for the infection probability is

$$p^*(t) = e^{Qt} p^*(0) + Q^{-1}(e^{Qt} - I)\boldsymbol{\epsilon},$$

where $\epsilon^T = (\epsilon_1, \cdots, \epsilon_n)$, and

$$e^{Qt} = \sum_{k=1}^{\infty} \frac{Q^t k^t}{k!}.$$

Proof: The epidemic spreading process can be written as the following master equation:

$$\frac{d\mathbb{E}[I_v(t)]}{dt} = \mathbb{E}\big[\big(1 - I_v(t)\big)\big(\beta \sum_{j=1}^{N} a_{vj} I_j(t) + \epsilon_v\big)\big] - \delta_v \mathbb{E}[I_v(t)], \quad v = 1, \cdots, N. \tag{3.1}$$

That is,

$$p_v'(t) = \mathbb{E}\big[\big(1 - I_v(t)\big)\big(\beta \sum_{j=1}^{N} a_{vj} I_j(t) + \epsilon_v\big)\big] - \delta_v \mathbb{E}[I_v(t)],$$

which could be rewritten as

$$p_v'(t) = \beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v - \beta \sum_{j=1}^{N} a_{vj} \mathbb{E}\big[I_j(t) I_v(t)\big] - \epsilon_v p_v(t) - \delta_v p_v(t).$$

Note that the dependence among $I_j(t)$ and $I_v(t)$ are generally positive [6]. Then we have

$$p_v'(t) \le \beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v - (\delta_v + \epsilon_v) p_v(t) - \beta \sum_{j=1}^{N} a_{vj} p_j(t) p_v(t). \tag{3.2}$$

It can be represented in the matrix form as

$$\boldsymbol{p}(t) \le [\beta A - diag(\delta_v + \epsilon_v)]\boldsymbol{p}(t) + \boldsymbol{\epsilon} - \beta diag(p_v(t)) A \boldsymbol{p}(t).$$

Note that, for any $t \ge 0$,

$$p_v(t) \ge \frac{\epsilon_v}{\delta_v + \epsilon_v}, \quad v = 1, \cdots, N.$$

This is because we could consider the infection $\beta = 0$, which would lead to a two-state continuous Markov chain. Now, let

$$Q = \mathrm{diag}\left(\frac{\beta\delta_v}{\delta_v+\epsilon_v}\right)A - \mathrm{diag}(\delta_v + \epsilon_v).$$

Therefore, it holds that

$$\boldsymbol{p}'(t) \leq Q\boldsymbol{p}(t) + \boldsymbol{\epsilon}.$$

Consider the following equation:

$$\boldsymbol{p}^{*\prime}(t) - Q\boldsymbol{p}^*(t) = \boldsymbol{\epsilon}. \qquad (3.3)$$

This is a nonhomogeneous linear differential equation of order 1, and it can be solved explicitly as follows:

$$\begin{aligned}\boldsymbol{p}^*(t) &= e^{Qt}\boldsymbol{p}^*(0) + \int_0^t e^{Q(t-s)}\boldsymbol{\epsilon}ds \\ &= e^{Qt}\boldsymbol{p}^*(0) + Q^{-1}[e^{Qt} - I]\boldsymbol{\epsilon}\end{aligned}$$

where

$$e^{Qt} = \sum_{k=0}^{\infty}\frac{Q^t k^t}{k!}.$$

Note that given the same initial probabilities $\boldsymbol{p}^*(0) = \boldsymbol{p}(0)$, it holds that $\boldsymbol{p}(t) \leq \boldsymbol{p}^*(t)$ for any $t > 0$. The proof is completed. ■

Remarks: Note that Q is symmetric, and it can be diagonalized as

$$Q = MDM^{-1},$$

where M is a real orthogonal matrix and D is a diagonal matrix. If all eigenvalues of the matrix Q have a negative real part, then Eq. (3.3) is stable [7], and the solution of Eq. (3.3) could be rewritten as

$$\boldsymbol{p}^*(t) = \boldsymbol{p}^* + e^{Qt}[\boldsymbol{p}^*(0) - \boldsymbol{p}^*],$$

where $\boldsymbol{p}^* = -Q^{-1}\boldsymbol{\epsilon}$ if Q is invertible.

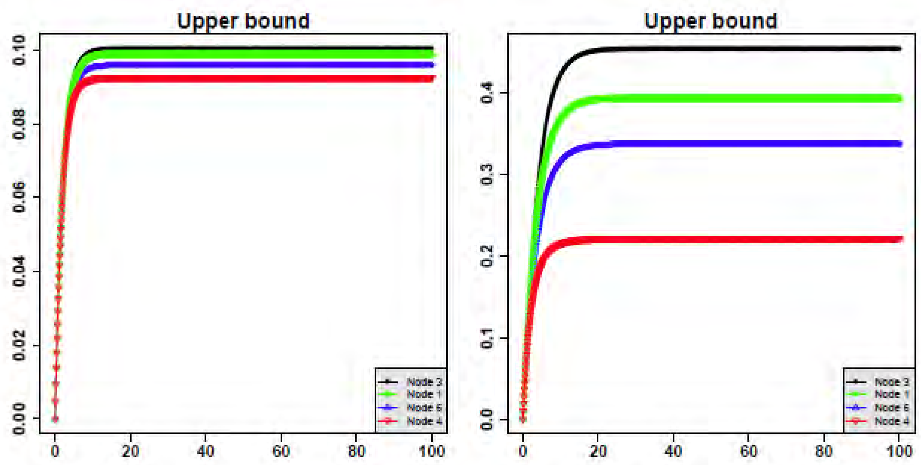
To illustrate, we present the following examples. For different scenarios presented in the paper, we use letter $M$for those based on Markov models and N for those based on non-Markov models.

Example 3.2 Consider the network in Figure 1. For simplicity, we assume that all the nodes have the same infection and recovery rates.

- Scenario M1: Assume the initial infection probability is 0, and $\beta = .01$, $\epsilon = .05$ and $\delta = .5$. In Figure 3(a), we plot the upper bounds $p^*(t)'s$ for nodes 3, 1, 6 and 4. We observe that the upper bounds for the probabilities of all nodes increase during the initial period and then become stable. It is also seen that node 3 has the largest infection probability, while node 4 has the smallest infection probability. This can be explained by the degree of nodes, since node 3 has the largest number of degrees, 6, but node 4 has only 1 degree.
- Scenario M2: Assume the initial infection probability is $.0005$, $\beta = \epsilon = .01$ and $\delta = .5$. In Figure 3(b), we plot the upper bounds $p^*(t)'s$ for nodes 3, 1, 6 and 4. We again observe that the upper bound for probabilities of all nodes

increase during the initial period and then become stable. Node 3 has the largest infection probability, while node 4 has the smallest infection probability.

Figure 3
Upper bounds for the infection probabilities of epidemic spreading over the network in Figure 1, where x-axis is time, and y-axis is the corresponding infection probabilities.



(a) Scenario M1

(b) Scenario M2

Comparing Scenarios M1 and M2, it is seen that the infection probabilities are larger in Scenario M2. This is because the infections β and ϵ are larger in Scenario M2. We observe that both scenarios have the constant upper bounds for the infection probabilities after the initial periods. This is, in fact, not surprising, as the evolution of epidemic spreading would enter the stable state, that is, the infection probabilities are constant, and this is known as the stationary state in the epidemic literature. Refer to Van Mieghem, Van Mieghem and Cator, and Xu, Da, and Xu [25, 26, 27] for more discussions on this topic.

Since the evolution of epidemic spreading can enter the stationary state, the following result presents the stationary probabilities.

Proposition 3.3 If the evolution of epidemic spreading enters the stationary state, then the stationary probability of infection for node v is

$$p_v = \frac{\beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v}{\beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v + \delta_v}, v = 1, \cdots, N.$$

Proof: If the evolution of epidemic spreading enters the stationary state, then for any node v, $p_v'(t) = 0$. According to Eq. (3.2), we have

$$0 = \beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v - p_v[\beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v + \delta_v] .$$

Note that here we use the equality instead of the inequality in Eq. (3.2) for the approximation, which can be considered as conservative estimates of stationary probabilities; see also Van Mieghem and Cator [26]. Therefore, we have

$$p_v = \frac{\beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v}{\beta \sum_{j=1}^{N} a_{vj} p_j(t) + \epsilon_v + \delta_v} , v = 1, \cdots, N.$$

Since the stationary probability is relatively easy to use, it can be used as an estimate of infection probabilities in practice. We present the following result for illustration.

Example 3.4 (Example continued.) Consider the scenarios in Example 3.2. We compute the stationary probabilities for both scenarios according to Eq. (3.4).

- For Scenario M1, the stationary probabilities are $p = (.0988, .0973, .1004, .0925, .0942, .0958, .0958, .0987, .0958, .0942)$. Compared to the upper bounds in Figure 3(a), the upper bounds here are very tight for this case.
- For Scenario M2, the stationary probabilities can be calculated as $p = (.3225, .3098, .3538, .2092, .2511, .2843, .2856, .3223, .2843, .2475)$. Compared to the upper bounds in Figure 3(b), the upper bounds here may be used as conservative estimates for infection probabilities.

The dynamic bounds in Theorem 3.1 may be used as conservative estimates of dynamic infection probabilities. The stationary probabilities may also be used as the estimates for infection probabilities as long as the evolution enters the stable state. The advantage of the Markov model is that it is simple and straightforward. However, in practice, the infection time may not be exponential [9]. Further, there may exist dependence among the infection processes. In the next section, we discuss a general model that would allow not only a general distribution for the infection time but also dependence among the infection processes.

3.2 Non-Markov Model

For the non-Markov model, we assume that for any node v, there exists D_v infected by neighbors launching attacks via links, where the times to infections from neighbors are modeled as random variables $(Y_{v_1}, \cdots, Y_{v_{Dv}})$ with the same marginal distribution F. The time to infection by the threats outside the network is modeled by random variable Z_v with distribution G_v. Therefore, the time to infection for node v is

$$T_v = \min(Y_{v_1}, \cdots, Y_{v_{Dv}}, Z_v).$$

We further assume that if node v is infected, then the attacks will stop, and after node v is recovered, the attacks will resume. The recovery time needed for an infected node v is R_v. Note that

$$D_v = \sum_{j=1}^{N} a_{vj} I_j ,$$

where I_j is the status of node j and

$$p_j = P(I_j = 1).$$

Therefore, we have

$$\mathbb{E}[D_v] = \sum_{j=1}^{N} a_{vj} p_j.$$

The non-Markov model may be considered as the stationary state of epidemic spreading. It is known from the theory of renewal process that the infection and recovery processes of node v can be regarded as an alternating renewal process with renewal interval $R_v + T_v$ [21, 15]. By the standard theory of alternating renewal processes, it holds that

$$p_v = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v]+\mathbb{E}[T_v]}. \qquad (3.4)$$

Therefore, the key quantity is $\mathbb{E}(T_v)$, the average infection time for node v, and the quantity can be represented as follows:

$$\begin{aligned} \mathbb{E}[T_v] &= \mathbb{E}[\min(Y_{v_1}, \cdots, Y_{v_{D_v}}, Z_v)] \\ &= \mathbb{E}[\mathbb{E}[\min(Y_{v_1}, \cdots, Y_{v_{D_v}}, Z_v)|D_v]] \\ &= \sum_{d_v=0}^{\deg(v)} P(D_v = d_v) \int_0^\infty \bar{H}_{d_v}(x, \cdots, x) \bar{G}_v(x)\, dx \end{aligned}$$

where

$$\bar{H}_{d_v}(x, \cdots, x) = P(Y_{v_1} > x, \cdots, Y_{v_{d_v}} > x) \qquad (3.5)$$

for $d_v \geq 1$, $\bar{H}_0 \equiv 1$, and

$$\bar{G}_v(x) = P(Z_v > x).$$

In the literature, there are only a few works on the non-Markov model of epidemic spreading [28, 5, 24, 19]. Our model is different from those in the literature in two respects: (1) It is often assumed that the infected neighbors may still attack node v even if node v is infected, while our model assumes that the infected neighbors stop attacking when node v is infected; and (2) the attack processes are often assumed to be independent, while our model can accommodate the dependence among attacks. The work in Xu and Xu [28] is mostly related to our proposed model, but the network topology is not utilized there.

Eq. (3.5) indicates that the dependence among attacks from neighbors is modeled by the joint survival distribution. The literature demonstrates that copula can be an efficient and flexible way for capturing high-dimensional dependence among various univariate marginals. In what follows, we briefly review the notion of copulas.

Copula is widely used for modeling dependence between random variables [14, 18]. The idea is to separate the modeling of univariate marginals and their dependence structures. The function C: $[0;\ 1] \to [0;\ 1]$ is referred to as a copula of dimension *n* if it has the following properties:

- $C(u_1, \cdots, u_n)$ is increasing in u_z for $z \in \{1, \cdots, n\}$.

- $C(u_1, \cdots, u_{z-1}, 0, u_{z+1}, \cdots, u_n) = 0$ for all $u_j \in [0,1]$ where $j \in 1, \cdots, n$ and $j \neq z$.
- $C(1, \cdots, 1, u_z, 1, \cdots, 1) = u_z$ for all $u_z \in [0.1]$ where $z = 1, \cdots, n$.
- C is n-increasing, namely, for all $(u_{1,1}, \cdots, u_{1,n})$ and $(u_{2,1}, \cdots, u_{2,n})$ in $[0;\ 1]^n$ with $u_{1,j} \leq u_{2,j}$ for all $j = 1, \cdots, n$, it holds that

$$\sum_{z_1=1}^{2} \cdots \sum_{z_n=1}^{2} (-1)^{\sum_{j=1}^{n} z_j} C\left(u_{z_1,1}, \cdots, u_{z_n,n}\right) \geq 0.$$

Let $X_1, \cdots, X_n$ be random variables with distribution functions respectively denoted by $F_1, \cdots, F_n$. Consider the joint distribution function $F(x_1, \cdots, x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n)$. The famous Sklar's theorem [23] says that there exists a copula C such that

$$F(x_1, \cdots, x_n) = C(F_1(x_1), \cdots, F_n(x_n)).$$

There are many copula structures [14, 18]. As examples, we will consider the following two families of dependence structures. The first example is the Gaussian copula

$$C(u_1, \cdots, u_n) = \Phi_\Sigma(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_n)),$$

where Φ^{-1} is the inverse cumulative distribution of the standard normal distribution, and Φ_Σ is the joint cumulative distribution of a multivariate normal distribution with mean vector zero and covariance matrix equal to the correlation matrix Σ. For simplicity, we will assume that the correlation matrix has the form

$$\Sigma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \tag{3.6}$$

where ρ is the correlation between the two relevant random variables. In this case, the Gaussian copula can be rewritten as

$$C(u_1, \cdots, u_n) = \Phi_\rho(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_n)). \tag{3.7}$$

The other example is the Archimedean copula, namely,

$$C(u_1, \cdots, u_n) = \psi_\rho(\psi^{-1}(u_1), \cdots, \psi^{-1}(u_n)),$$

where ψ is the Archimedean generator of C. A particular case is the Clayton copula when the generator takes the form $\psi_\theta(s) = (1+s)^{-1/\theta}$, and

$$C(u_1, \cdots, u_n) = [\sum_{j=1}^{n} u_j^{-\theta} - n + 1]^{-1/\theta}, \ \ \theta > 0. \tag{3.8}$$

The Clayton copula models a positive dependence, especially a lower-tail dependence [14, 18].

Note that the joint survival function $\overline{H}_{d_v}$ can be rewritten as

$$\overline{H}_{d_v}(x, \cdots, x) = C(\overline{F}_1(x), \cdots, \overline{F}_{v_{d_v}}(x)), \tag{3.9}$$

where C is the survival copula of $(Y_1, \cdots, Y_{v_d})$. We remark that if there exists *positive lower orthant dependence* among $Y_1, \cdots, Y_{v_{d_v}}$, then it follows that, for $d_v \geq 1$,

$$\bar{H}_{d_v}(x, \cdots, x) \geq \prod_{i=1}^{d_v} \bar{F}_i(x) = \bar{F}^{d_v}(x), \qquad (3.10)$$

where the right side of the equation is, in fact, the independent case. We use the term *positive dependence* for the positive lower orthant dependence in the follow discussion.

It is often reasonable to consider that if two nodes are connected with each other directly, then the dependence is stronger than that for those disconnected. If two nodes are not connected directly, the dependence between them can be weaker and even independence can be assumed. Now, we use a copula C to model the dependence between the time-to-infection random variables $(Y_1, \cdots, Y_{v_{d_v}})$ as in Eq. (3.9). However, for those v-neighbors $(v_1, \cdots, v_{d_v})$, we assume that there is an adjacency matrix A_v that describes whether two of those v-neighbors are connected or not, with 1: connected, and 0: otherwise.

Then a multivariate Gaussian copula for those v-neighbors has the following correlation matrix:

$$A_v \cdot \Sigma = A_v \cdot \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \qquad (3.11)$$

where $\cdot$ is the element-wise multiplication and ρ is the correlation between the two relevant random variables. Such a neighboring effect due to A_v, together with the case without neighboring effects as in Eq. (3.6), will be considered in a simulation study in Section 4.2. The result follows immediately from Eq. (3.10), which presents an upper bound for the infection probability.

Proposition 3.5 If there exists a positive dependence among the successful infection times among neighbors, then an upper bound for infection probability of node v is given by

$$p_v \leq \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + \sum_{d_v=0}^{\deg(v)} P(D_v = d_v) \int_0^\infty \bar{F}^{d_v}(x) \bar{G}_v(x) dx}. \qquad (3.12)$$

Note that Eq. (3.12) simply implies that an upper bound for p_v is achieved when the times to infection from neighboring random variables Y_{v_i}'s are independent. However, the infection information of degree distribution (i.e., the distribution of D_v) is required for the upper bound. One may refer to Xu and Xu [28] for the discussion of upper bounds when the degree distribution is known. In the following discussion, we examine how to approximate the upper bounds without the infection information of degree distributions.

Note that for the independent case, we have

$$\mathbb{E}[T_v] = \mathbb{E}\left[\int_0^\infty \bar{F}^{D_v}(x) \bar{G}_v(x)\, dx\right].$$

By Jensen's inequality, it follows that

$$\mathbb{E}\left[\int_0^\infty \bar{F}^{D_v}(x)\bar{G}_v(x)\,dx\right] \geq \int_0^\infty \bar{F}^{\mathbb{E}[D_v]}(x)\bar{G}_v(x)\,dx.$$

Therefore,

$$p_v \leq \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v]+\int_0^\infty \bar{F}^{\mathbb{E}[D_v]}(x)\bar{G}_v(x)dx}\,. \tag{3.13}$$

Now, let us consider the following epidemic equation,

$$p_v^* = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v]+\int_0^\infty \bar{F}^{\sum_{j=1}^N a_{vj}p_j^*}(x)\bar{G}_v(x)dx} \quad , \tag{3.14}$$

which may be used as an approximation for the upper bound in practice.

Next, we derive the upper bounds for several general distributions for the time-to-infection random variables. Note that here we do not need the dependence structures to derive such an upper bound. The dependence structures discussed earlier will be used in Section 4.2 for simulation studies on how dependence structures and network topologies affect the infection probabilities.

- *Exponential infection and recovery*. In this case, we assume that the infection processes follow the exponential distributions. Specifically, we assume that

$$\bar{F}(x) = e^{-\beta x}$$

and

$$\bar{G}_v(x) = e^{-\epsilon_v x}.$$

Then, we have

$$p_v^* = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v]+1/(\epsilon_v+\beta\sum_{j=1}^N a_{vj}p_j^*)}\ .$$

If we further assume that the recovery process also follows an exponential distribution, we get

$$\bar{S}_v(x) = P(R_v > x) = e^{-\beta_v x}.$$

It then reduces to the Markov model in the previous section (see Proposition 3.3). That is, we have

$$p_v^* = \frac{1/\delta_v}{1/\delta_v+1/(\epsilon_v+\sum_{j=1}^N a_{vj}p_j^*)} = \frac{\epsilon_v+\beta\sum_{j=1}^N a_{vj}p_j^*}{\epsilon_v+\beta\sum_{j=1}^N a_{vj}p_j^*+\delta_v}\,.$$

We remark that when $\epsilon_v = 0$, this case coincides with the one in Cator, Van de Bovenkamp, and Van Mieghem [5], where the authors study the general model from a different perspective.

- *Weibull infection and recovery*. In this case, we assume that the infection processes follow Weibull distributions. That is,

$$\bar{F}(x) = e^{-(\beta x)^{\alpha_1}}$$

and

$$\bar{G}_v(x) = e^{-(\epsilon_v x)^{\alpha_2}}$$

where β and ϵ_v are scale parameters, and α_1 and α_2 are the shape parameters. Then,

$$\begin{aligned}\mathbb{E}[T_v^*] &= \int_0^\infty e^{-[(\epsilon_v x)^{\alpha_2}+(\beta x)^{\alpha_1}\sum_{j=1}^N a_{vj}p_j^*]}\,dx \\ &= \int_0^\infty e^{-[\epsilon_v^{\alpha_2}\,x^{\alpha_2}+(\beta x)^{\alpha_1}\sum_{j=1}^N a_{vj}p_j^*]}\,dx \\ &= \phi\big(\epsilon_v,\beta,\alpha_1,\alpha_{2,}\,\boldsymbol{p}^*\big).\end{aligned}$$

Note that when $\alpha_1 = \alpha_2 = \alpha$, it holds that

$$\phi\big(\epsilon_v,\beta,\alpha_1,\alpha_{2,}\,\boldsymbol{p}^*\big) = \frac{1}{\left[\epsilon_v^\alpha+\beta^\alpha\sum_{j=1}^N a_{vj}p_j^*\right]^{1/\alpha}}\,\Gamma(1+\frac{1}{\alpha}).$$

If we further assume that the recovery also follows a Weibull distribution with survival function

$$\bar{S}_v(x) = e^{-(\delta_v x)^{\alpha_3}},$$

then

$$\mathbb{E}[R_v] = \frac{1}{\delta_v}\,\Gamma(1+\frac{1}{\alpha}).$$

Hence, the infection probability can be rewritten as

$$p_v^* = \frac{\Gamma(1+\frac{1}{\alpha_3})}{\Gamma(1+\frac{1}{\alpha_3})+\delta_v\phi\big(\epsilon_v,\beta,\alpha_1,\alpha_{2,}\,\boldsymbol{p}^*\big)}\,. \qquad (3.16)$$

- *Log-normal infection and recovery*. For this case, we assume that the infection processes follow log-normal distributions. Given that, the density function of Y_{v_j} can be written as

$$f_{v_j}(x) = \frac{1}{x\sigma_1\sqrt{2\pi}}\exp[-\frac{\ln(x)-\mu_1}{2\sigma_1^2}],$$

and the density for Z_v is

$$g(x) = \frac{1}{x\sigma_2\sqrt{2\pi}}\exp[-\frac{\ln(x)-\mu_2}{2\sigma_2^2}].$$

Therefore, we have

$$\begin{aligned}\mathbb{E}[T_v^*] &= \int_0^\infty[1-\Phi(\frac{\ln(x)-\mu_1}{\sigma_1})][1-\Phi(\frac{\ln(x)-\mu_2}{\sigma_2})]^{\sum_{j=1}^N a_{vj}p_j^*}\,dx \\ &=:\ \Psi(\mu_1,\mu_2,\sigma_1,\sigma_2,\boldsymbol{p}^*).\end{aligned}$$

If we assume that the recovery process also follows a log-normal distribution with distribution function

$$S_v(x) = \Phi(\frac{\ln(x)-\mu_v}{\sigma_v})\,,$$

then it holds that

$$\mathbb{E}[R_v] = \exp(\mu_v+\sigma_v^2/2)\,.$$

Hence, we have

$$p_v^* = \frac{\exp(\mu_v+\sigma_v^2/2)}{\exp(\mu_v+\sigma_v^2/2)\,+\Psi(\mu_1,\mu_2,\sigma_1,\sigma_2,\boldsymbol{p}^*)}\,.$$

Note that the choices of the distributions for recovery processes, infection processes from outside sources and infection processes from neighbors all affect the infection probability simultaneously. Here we choose the same distribution family for the infection and recovery processes to illustrate the idea, and the general model proposed allows different distributions for those processes. To illustrate, we present the following examples for the Weibull distribution.

Example 3.6 Consider the network in Figure 1. Assume that the infection processes follow Weibull distributions. We consider the following two scenarios.

- Scenario N1: The parameters for the Weibull distributions are set to be

$$(\beta,\sigma_v,\alpha,\alpha_3) = (1,.5,2,2).$$

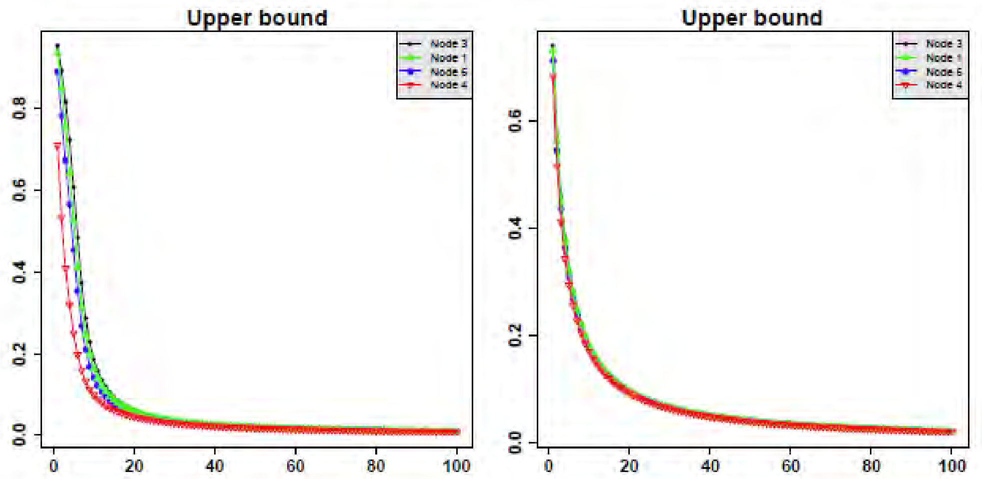
For this case, we calculate the infection probability for different values of δ's. In Figure 4(a), we plot the infection probabilities of Eq. (3.16) for different values of δ's. It is seen that when δ is larger, that is, the recovery power is strong, the infection probability is small. It fits expectations that if the recovery process is quickly completed, it would increase the security of the network. We again observe that node 3 has the largest infection probability, and node 4 has the smallest infection probability.

- Scenario N2: The parameters for the Weibull distributions are set to be

$$(\beta,\sigma_v,\alpha,\alpha_3) = (1,.5,.5,2).$$

For this case, we plot the infection probabilities of Eq. (3.16) for different values of δ's in Figure 4(a). We again observe that when δ is larger, the infection probability is small.

Figure 4
Upper bounds for the infection probabilities of epidemic spreading over the network in Figure 1, where x-axis represents the value of δ's.



(a) $\alpha = .5$

(b) $\alpha = 2$

Comparing Scenarios N1 and N2, we observe that the shape of parameter α also has a significant effect on the infection probabilities. The probabilities in Scenario N2 drop more slowly than those in Scenario N1, and this is mainly because a larger α would result in a smaller infection time.

Note that although the non-Markov model proposed in this section is able to model the multivariate dependence, the multivariate dependence structure can be challenging to implement in practice. The main reasons are (1) the high-dimensional dependence structure is challenging to specify in practice; and (2) there is not enough data to verify the dependence structure. Therefore, we propose to develop the upper bound for the conservative estimates for the infection probabilities. If an insurance company seeks more accurate estimates, then the Monte Carlo simulation approach can be used. In what follows, we discuss the pricing strategies based on the simulation approach, and it is seen that the dependence structure in the general model can be easily implemented in the simulation algorithm.

Section 4: Simulation and Pricing

In this section, we discuss a pricing framework for cybersecurity risk based on simulation. Assume that for a node v, the initial wealth (or information) is e_v. Since the infection event may not result in a total loss of information, we assume that the loss of node v is distributed according to beta distribution with the density function as

$$f_{L_v}(x) = \frac{1}{\omega_v^{a+b-1}} \frac{1}{B(a,b)} x^{a-1}(\omega_v - x)^{b-1}, \qquad 0 \leq x \leq \omega_v$$

where $a, b > 0$ are shape parameters, and B is the beta function. The cost functions are defined as

$$\eta_v(l_v) = cl_v, \quad C_v(r_v) = c_1\omega_v + c_2 r_v \tag{4.1}$$

where c means the cost rate due to infection, c_1 represents the cost rate based on the initial value and c_2 represents the cost rate of the recovery process. It is seen that the cost function defined in Eq. (4.1) depends on not only the duration of downtime but also the wealth of the node.

We study a one-year insurance contract, and two premium principles are considered. The first one is the *standard deviation premium principle:*

$$H(x) = \mathbb{E}[X] + \lambda\sqrt{Var(X)}\,, \tag{4.2}$$

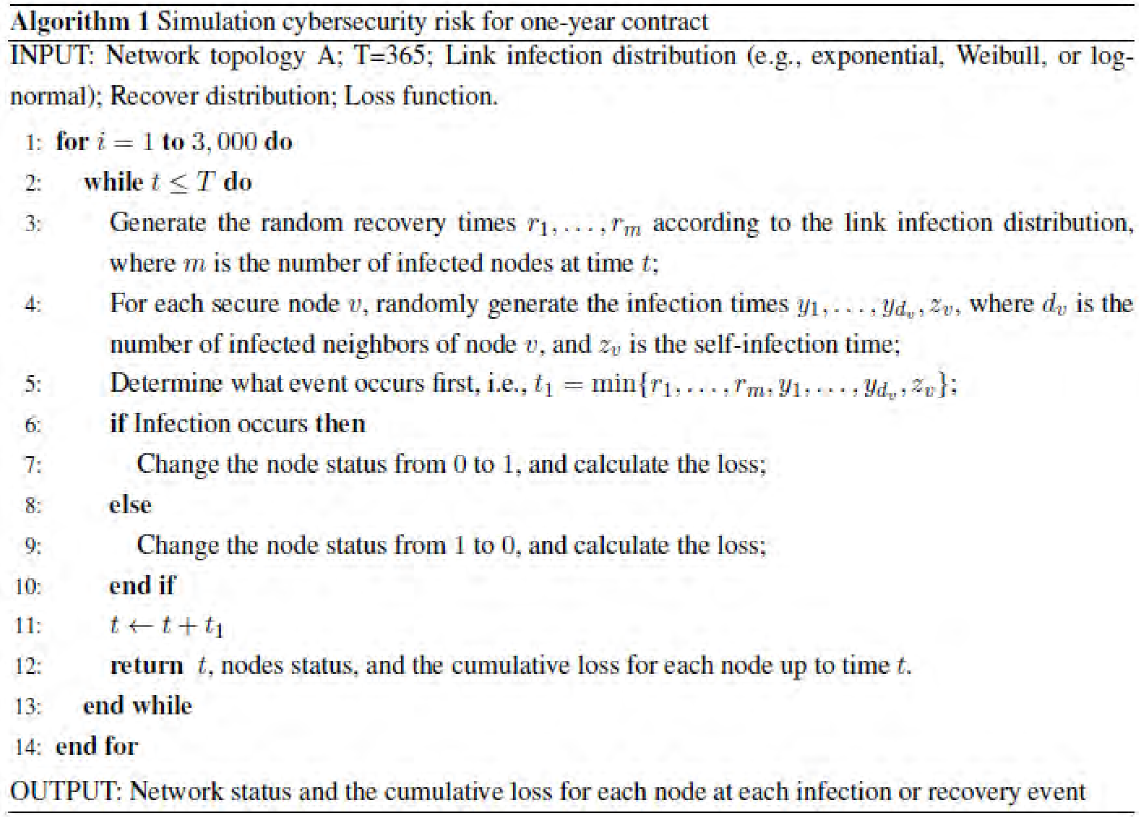
where $\lambda > 0$ is the risk loading. The second one is the principle of equivalent utility, where the premium $H(X)$ solves the equation

$$u(\omega_v)\text{=}\ \mathbb{E}[u(\omega_v - X + H(X))]\,, \tag{4.3}$$

where u is an increasing concave utility of wealth and ω is the initial wealth. In the rest of the discussion, we consider the constant relative risk-averse utility function, which is commonly used in the literature [20, 4]:

$$\begin{cases} \dfrac{\omega^{1-\gamma}}{1-\gamma} & ,\gamma \neq 1 > 0 \\ \log(\omega) & ,\gamma = 1 \end{cases},$$

where γ is the parameter for the degree of risk aversion. In what follows, we study the pricing strategies based on the proposed models. The experiment is based on $3{,}000$ Monte Carlo simulations. The parameters for the loss model are assumed to be $(a, b, c, c_1, c_2) = (2, 4, .001, .1 \times 10^{-6},\ .5 \times 10^{-4})$, and we assume that the initial wealth of each node is $\omega_v = 1000$ dollars. The simulation algorithm is shown in Algorithm 1.



Algorithm 1 Simulation cybersecurity risk for one-year contract

INPUT: Network topology A; T=365; Link infection distribution (e.g., exponential, Weibull, or log-normal); Recover distribution; Loss function.

```
for $i = 1$ to $3,000$ do
    while $t \leq T$ do
        Generate the random recovery times $r_1, \dots, r_m$ according to the link infection distribution, where $m$ is the number of infected nodes at time $t$;
        For each secure node $v$, randomly generate the infection times $y_1, \dots, y_{d_v}, z_v$, where $d_v$ is the number of infected neighbors of node $v$, and $z_v$ is the self-infection time;
        Determine what event occurs first, i.e., $t_1 = \min\{r_1, \dots, r_m, y_1, \dots, y_{d_v}, z_v\}$;
        if Infection occurs then
            Change the node status from 0 to 1, and calculate the loss;
        else
            Change the node status from 1 to 0, and calculate the loss;
        end if
        $t \leftarrow t + t_1$
        return $t$, nodes status, and the cumulative loss for each node up to time $t$.
    end while
end for
```

OUTPUT: Network status and the cumulative loss for each node at each infection or recovery event

Algorithm 1 allows us to record the evolution of network status during the contract year, and we can calculate the cumulative loss for each node at any time t.

4.1 Independent cybersecurity risks

In this section, the simulation is based on the assumption that the infection processes are independent. The quantities we are interested in for each node include (1) the total number of incidents; (2) the infection probability; and (3) the total loss. The network topology used for the simulation is from Figure 1. We assume that there is no infection at the beginning, $T = 0$.

4.1.1 Exponential Distribution

For this section, we consider the Markov model in Section 3.1. The following two scenarios are considered.

a) Scenario M3: We assume that for any node v, $v = 1, \cdots, N$, the parameters are

$$(\beta, \epsilon_v, \delta_v) = (.2, .5, 1).$$

Then, it is easy to see that

$$E(R_v) = 1.$$

Using Eq. (3.15), we can solve the upper bounds for infection probabilities as

$$(.4833, .4667, .5092, .3737, .4112, .4419, .4429, .4831, .4419, .4094),$$

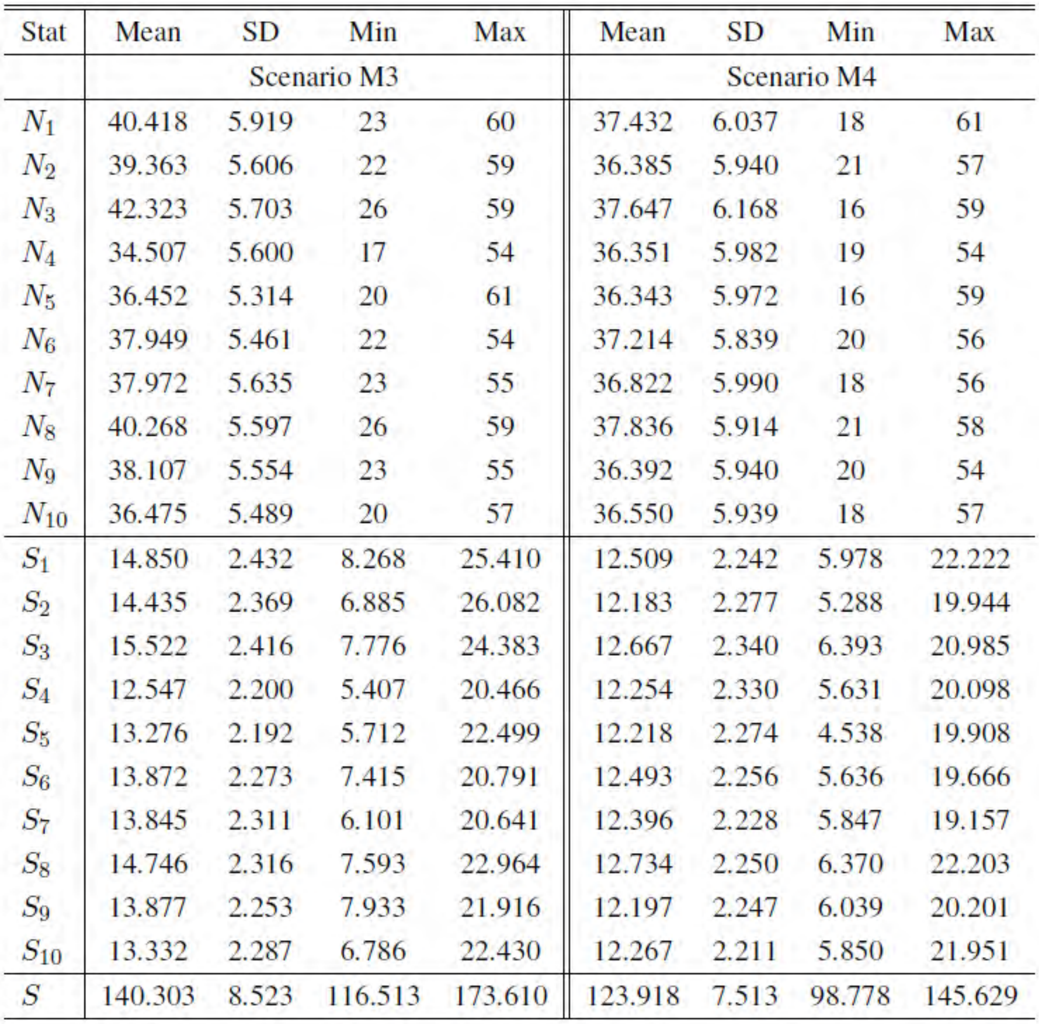
and the expected successful infection times can be computed as

$E(T^{*}) = (1.0691, 1.1427, .9639, 1.6759, 1.4319, 1.2630, 1.2578, 1.0700, 1.2630, 1.4426).$

It is observed that the successful infection time for node 3 (.9639) is the smallest, which indicates that node 3 has the largest chance to be infected. Table 1 presents the infections and related losses for all nodes based on 3,000 simulations. We observe that the average number of infections for node 3 is 42.323, which is the largest among all the nodes. Node 4 has the smallest number of incidents with 34.507. For the related loss, we observe that the loss of node 3 again is the largest, and the loss of node 4 is the smallest. Figure 5(a) shows the evolution of simulated infection probabilities. It's observed that node 3 has the largest infection probability during the whole year except for the initial period. Compared to the upper bounds, all the simulated probabilities are smaller than those of the upper bounds. This indicates that for this scenario, the upper bounds are rather conservative.

Table 1

Simulation Based on the Markov Model, Where N_J Represents the Number of Infections and S_J Means the Total Loss forNode J during One Year. S Represents the Total Loss for the Network.



Stat	Mean	SD	Min	Max	Mean	SD	Min	Max
	Scenario M3				Scenario M4			
N_1	40.418	5.919	23	60	37.432	6.037	18	61
N_2	39.363	5.606	22	59	36.385	5.940	21	57
N_3	42.323	5.703	26	59	37.647	6.168	16	59
N_4	34.507	5.600	17	54	36.351	5.982	19	54
N_5	36.452	5.314	20	61	36.343	5.972	16	59
N_6	37.949	5.461	22	54	37.214	5.839	20	56
N_7	37.972	5.635	23	55	36.822	5.990	18	56
N_8	40.268	5.597	26	59	37.836	5.914	21	58
N_9	38.107	5.554	23	55	36.392	5.940	20	54
N_{10}	36.475	5.489	20	57	36.550	5.939	18	57
S_1	14.850	2.432	8.268	25.410	12.509	2.242	5.978	22.222
S_2	14.435	2.369	6.885	26.082	12.183	2.277	5.288	19.944
S_3	15.522	2.416	7.776	24.383	12.667	2.340	6.393	20.985
S_4	12.547	2.200	5.407	20.466	12.254	2.330	5.631	20.098
S_5	13.276	2.192	5.712	22.499	12.218	2.274	4.538	19.908
S_6	13.872	2.273	7.415	20.791	12.493	2.256	5.636	19.666
S_7	13.845	2.311	6.101	20.641	12.396	2.228	5.847	19.157
S_8	14.746	2.316	7.593	22.964	12.734	2.250	6.370	22.203
S_9	13.877	2.253	7.933	21.916	12.197	2.247	6.039	20.201
S_{10}	13.332	2.287	6.786	22.430	12.267	2.211	5.850	21.951
S	140.303	8.523	116.513	173.610	123.918	7.513	98.778	145.629

b) Scenario M4: The parameter vector is set as follows:

$$(\beta, \epsilon_v, \delta_v) = (.2, .5, 5).$$

Compared to Scenario M3, the infection rates are small and the recovery rate is large. It can be computed that

$$E(R_v) = 2$$

and

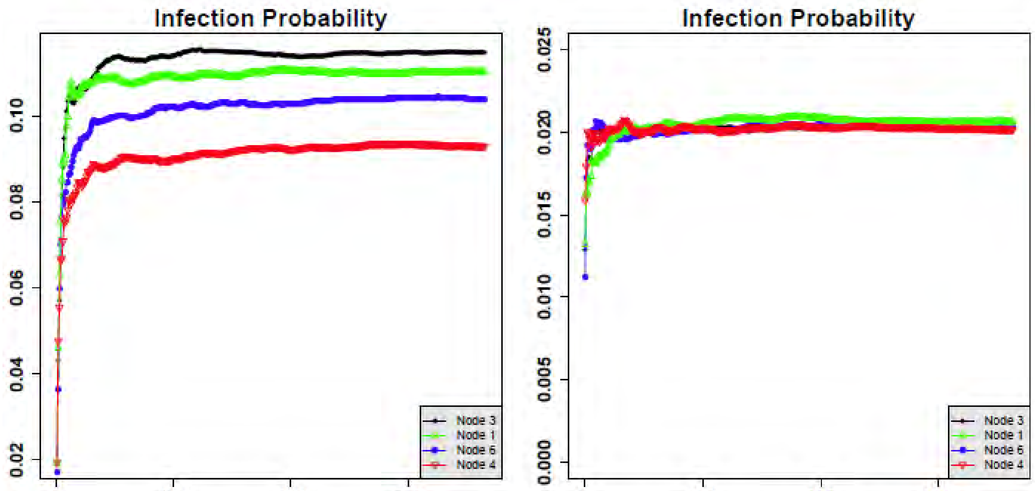
$$E(T^*) = (1.6639, 1.7120, 1.6051, 1.9186, 1.8387, 1.7704, 1.7704, 1.6639, 1.7704, 1.8429).$$

The upper bounds for the infection probabilities are

$$(.1073, .1046, .1108, .0944, .0981, .1015, .1015, .1073, .1015, .0979).$$

Compared to Scenario M3, the expected successful infection times are longer and upper bounds are much smaller. This means that we have a better network environment or recovery ability. Table 1 also shows that the number of infections is overall less than in Scenario M3, except that nodes 4 and 10 are slightly larger. All the losses in this scenario are smaller than the corresponding ones in Scenario M3. The total loss in Scenario M4 is reduced by 11.68% of that in Scenario M3. In particular, the simulated infection probabilities as shown in Figure 5(b) are very small.

Figure 5
Simulated infection probabilities of epidemic spreading over the network in Figure 1 based on the Markov model



(a) Scenario M3

(b) Scenario M4

4.1.2 Weibull Distribution

In this section, we consider the Weibull infection processes as well as the Weibull recovery process, as in Section 3.2. The following two scenarios are considered, where the shape parameters of the infection distributions are assumed to be the same, namely, α.

a) Scenario N3: For any node *v*, the parameter vector is set as follows:

$$(\beta, \epsilon_v, \delta_v, \alpha, \alpha_3) = (.2, .5, 1, 2, 2).$$

Then, it can be computed that

$$\mathbb{E}(R_v) = .8862.$$

By Eq. (3.16), we have

$$\mathbb{E}(T^*) = (1.5660, 1.5990, 1.5319, 1.7234, 1.6773, 1.6365, 1.6358, 1.5667, 1.6365, 1.6781)$$

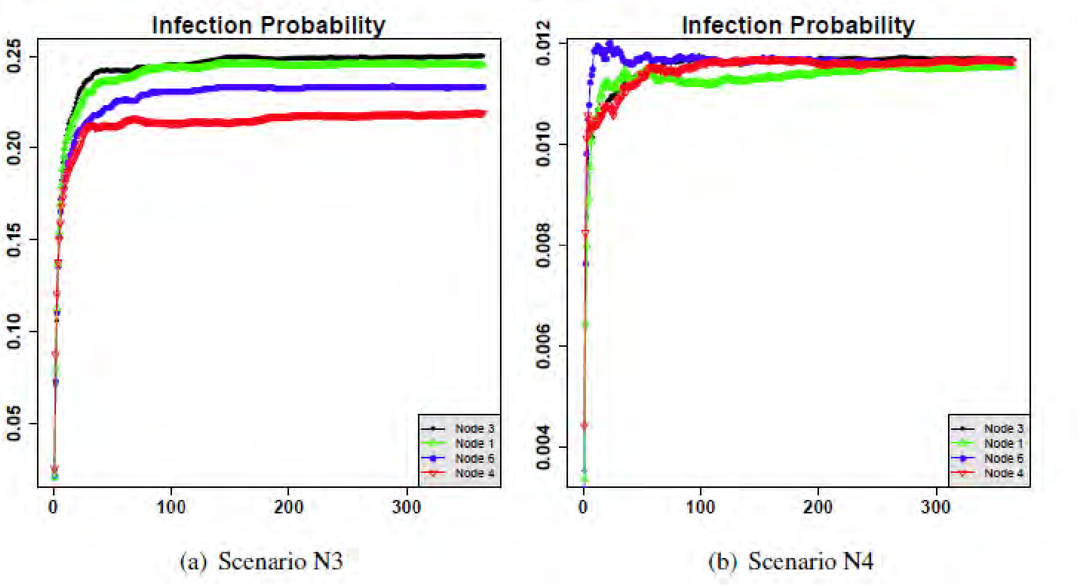
and the upper bounds for infection probabilities are

$$(.3614, .3566, .3665, .3396, .3457, .3513, .3514, .3613, .3513, .3456).$$

Figure 6(a) shows the simulated infection probabilities for one year. Again, we observe that node 3 has the largest overall infection probability. The upper bounds are reasonably close to the simulated infection probabilities. Table 2 shows that node 3 has the largest number of infection incidents and the largest loss, while node 1 has the smallest number of infection incidents and the smallest loss.

Figure 6

Simulated infection probabilities of epidemic spreading over the network in Figure 1 based on Weibull infection and recovery processes.

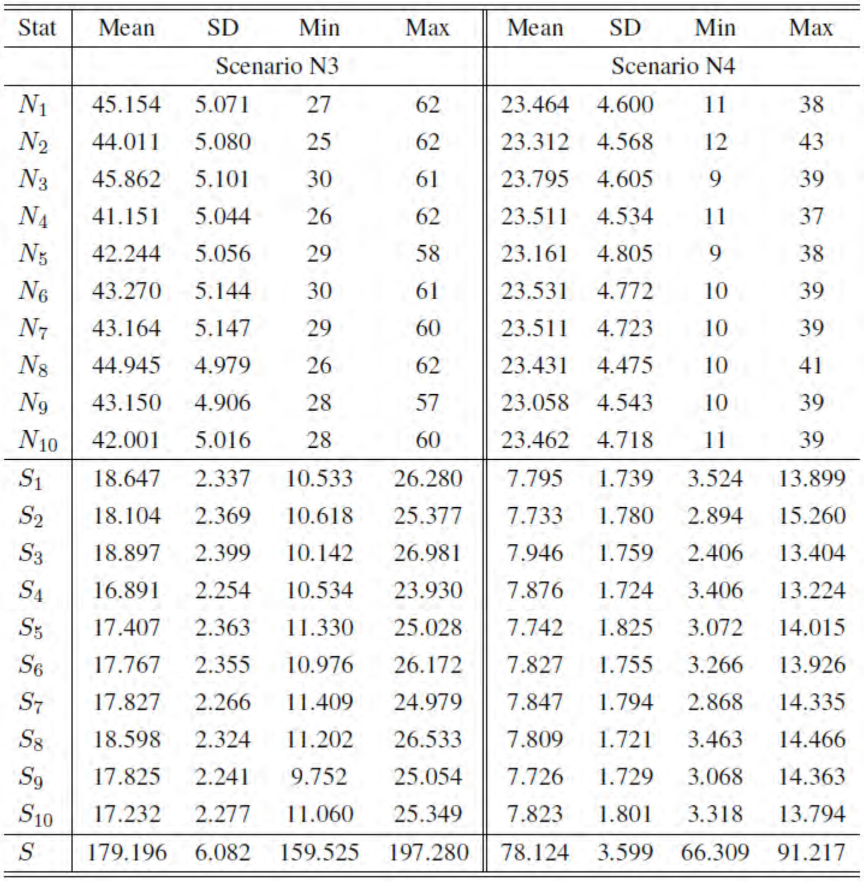


(a) Scenario N3

(b) Scenario N4

Table 2

Simulation Based on the Weibull Infection and Recovery Processes, Where N_J Represents the Number of Infections and S_J Means the Total Loss forNode J during One Year. S Represents the Total Loss for the Network.



Stat	Mean	SD	Min	Max	Mean	SD	Min	Max
	Scenario N3				Scenario N4			
N_1	45.154	5.071	27	62	23.464	4.600	11	38
N_2	44.011	5.080	25	62	23.312	4.568	12	43
N_3	45.862	5.101	30	61	23.795	4.605	9	39
N_4	41.151	5.044	26	62	23.511	4.534	11	37
N_5	42.244	5.056	29	58	23.161	4.805	9	38
N_6	43.270	5.144	30	61	23.531	4.772	10	39
N_7	43.164	5.147	29	60	23.511	4.723	10	39
N_8	44.945	4.979	26	62	23.431	4.475	10	41
N_9	43.150	4.906	28	57	23.058	4.543	10	39
N_{10}	42.001	5.016	28	60	23.462	4.718	11	39
S_1	18.647	2.337	10.533	26.280	7.795	1.739	3.524	13.899
S_2	18.104	2.369	10.618	25.377	7.733	1.780	2.894	15.260
S_3	18.897	2.399	10.142	26.981	7.946	1.759	2.406	13.404
S_4	16.891	2.254	10.534	23.930	7.876	1.724	3.406	13.224
S_5	17.407	2.363	11.330	25.028	7.742	1.825	3.072	14.015
S_6	17.767	2.355	10.976	26.172	7.827	1.755	3.266	13.926
S_7	17.827	2.266	11.409	24.979	7.847	1.794	2.868	14.335
S_8	18.598	2.324	11.202	26.533	7.809	1.721	3.463	14.466
S_9	17.825	2.241	9.752	25.054	7.726	1.729	3.068	14.363
S_{10}	17.232	2.277	11.060	25.349	7.823	1.801	3.318	13.794
S	179.196	6.082	159.525	197.280	78.124	3.599	66.309	91.217

b) Scenario N4: For this scenario, the parameter vector is assumed to be

$$(\beta, \epsilon_v, \delta_v, \alpha, \alpha_3) = \ (.1, .2, 5, 2, 2).$$

The expected recovery time for node v is

$$\mathbb{E}(R_v) = .1772.$$

Again by Eq. (3.16), we have

$$\mathbb{E}(T^*) = (4.3214, 4.3443, 4.3100, 4.4146, 4.3909, 4.3675, 4.3675, 4.3214, 4.3675, 4.3909)$$

and the upper bounds for infection probabilities are

$$(.0394, .0392, .0395, .0386, .0388, .0390, .0390, .0394, .0390, .0388).$$

We see that the recovery times and infection probabilities are much smaller than the corresponding ones in Scenario N3. The expected successful infection times are longer in this case. This indicates that Scenario N4 has a better network environment and stronger recovery power. The simulated infection probabilities are very small, say, less than .012, and again, the upper bounds are relatively closer to the simulated infection probabilities. Table 2 shows that the numbers of incidents and losses for all nodes are close. The total loss in this case is 78.124 compared to 179.196 in Scenario N3, a reduction of 56.4%.

4.1.3 Log-Normal Distribution

In this section, we consider the log-normal infection processes as well as the log-normal recovery process, as in Section 3.2. The following two scenarios are considered.

a) Scenario N5: We assume that the parameter vector is as follows:

$$(\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_v, \sigma_v) = (1.1094, 1, .1931, 1, -.5, 1). \qquad (4.4)$$

Then, it is easy to compute that

$$\mathbb{E}(R_v) = 1.$$

By Eq. (3.17) we can solve that

$$E(T^*) = (1.1053, 1.1650, 1.0288, 1.6274, 1.4027, 1.2614, 1.2578, 1.1048, 1.2614, 1.4091),$$

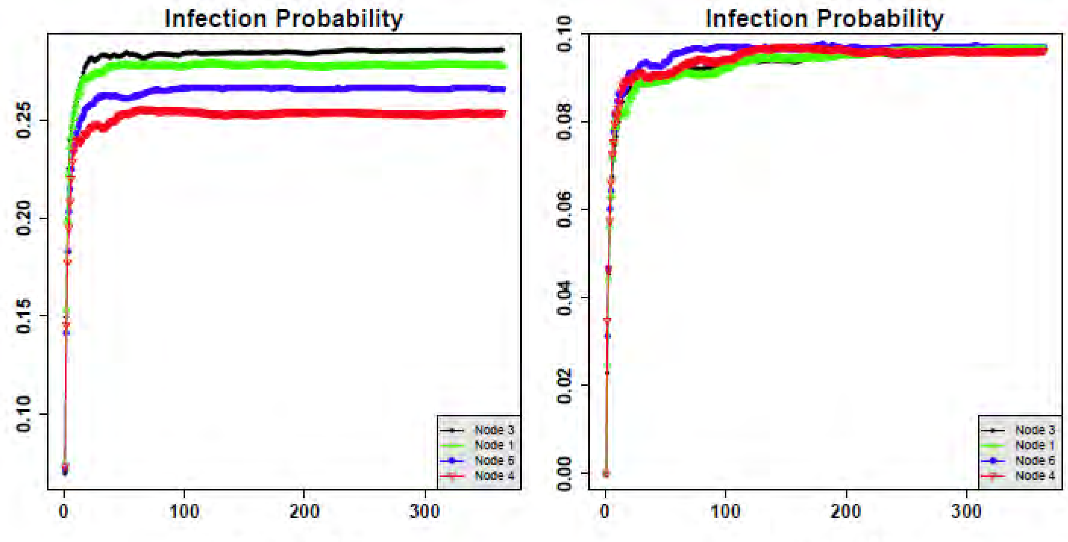
and the upper bounds for the infection probabilities are

$$(.4750, .4619, .4929, .3806, .4162, .4422, .4429, .4751, .4422, .4151).$$

The simulated infection probabilities are plotted in Figure 7(a). It is seen that the upper bounds are reasonably close to the simulated ones. We again observe that node 3 has the largest infection probabilities, and node 1 has the smallest infection probabilities. Table 3 shows that similar conclusions about the numbers of incidents and losses can be drawn for this case.

Figure 7

Simulated infection probabilities of epidemic spreading over the network in Figure 1 based on log-normal infection and recovery processes.

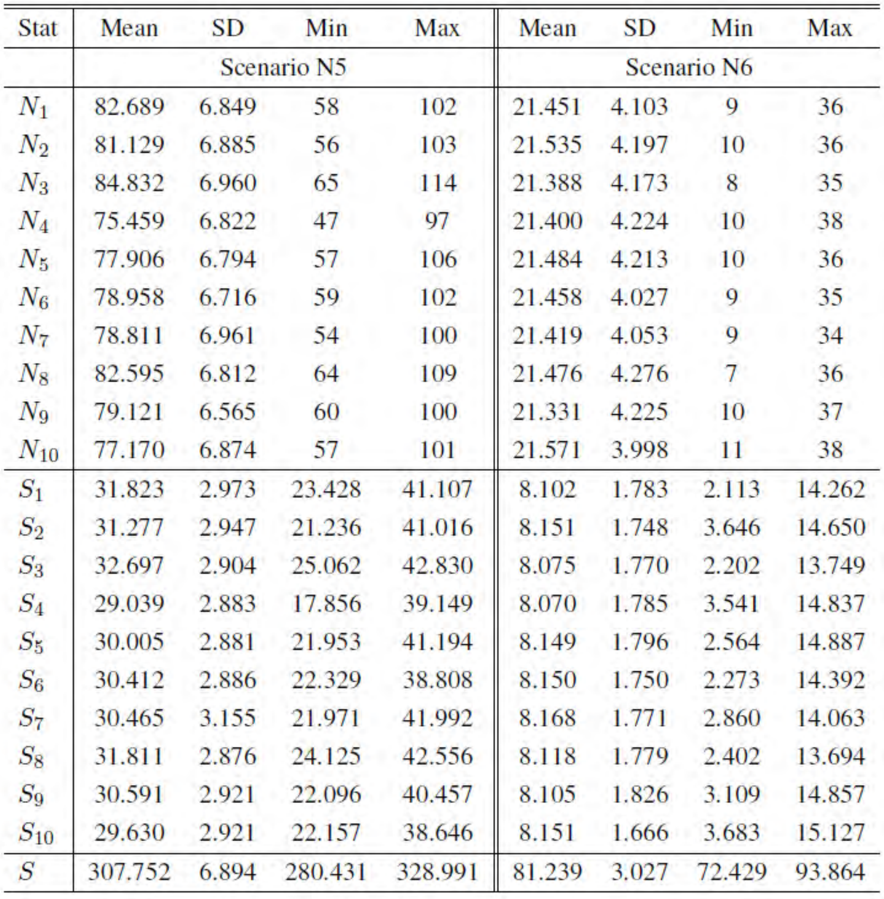


(a) Scenario N5

(b) Scenario N6

Table 3

Simulation Based on the Log-Normal Infection and Recovery Processes, Where N_J Represents the Number of Infections and S_J Means the Total Loss for Node J during One Year. S Represents the Total Loss of the Network.



Stat	Mean	SD	Min	Max	Mean	SD	Min	Max
	Scenario N5				Scenario N6			
N_1	82.689	6.849	58	102	21.451	4.103	9	36
N_2	81.129	6.885	56	103	21.535	4.197	10	36
N_3	84.832	6.960	65	114	21.388	4.173	8	35
N_4	75.459	6.822	47	97	21.400	4.224	10	38
N_5	77.906	6.794	57	106	21.484	4.213	10	36
N_6	78.958	6.716	59	102	21.458	4.027	9	35
N_7	78.811	6.961	54	100	21.419	4.053	9	34
N_8	82.595	6.812	64	109	21.476	4.276	7	36
N_9	79.121	6.565	60	100	21.331	4.225	10	37
N_{10}	77.170	6.874	57	101	21.571	3.998	11	38
S_1	31.823	2.973	23.428	41.107	8.102	1.783	2.113	14.262
S_2	31.277	2.947	21.236	41.016	8.151	1.748	3.646	14.650
S_3	32.697	2.904	25.062	42.830	8.075	1.770	2.202	13.749
S_4	29.039	2.883	17.856	39.149	8.070	1.785	3.541	14.837
S_5	30.005	2.881	21.953	41.194	8.149	1.796	2.564	14.887
S_6	30.412	2.886	22.329	38.808	8.150	1.750	2.273	14.392
S_7	30.465	3.155	21.971	41.992	8.168	1.771	2.860	14.063
S_8	31.811	2.876	24.125	42.556	8.118	1.779	2.402	13.694
S_9	30.591	2.921	22.096	40.457	8.105	1.826	3.109	14.857
S_{10}	29.630	2.921	22.157	38.646	8.151	1.666	3.683	15.127
S	307.752	6.894	280.431	328.991	81.239	3.027	72.429	93.864

b) Scenario N6: For this case, we set the parameter vector as follows:

$$(\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_v, \sigma_v) = (1.5294, .4, .6131, .4, -.08, .4).$$

It is easy to compute that

$$\mathbb{E}(R_v) = 1,$$

that is, we have the same expected recovery time. In fact, the expected values of Y_{v_j}'s and Z_v's are all equal to those in Scenario N5, while the variances are smaller. By Eq. (3.17), we have

$$E(T^*) = (1.9403, 1.9499, 1.9317, 1.9851, 1.9727, 1.9612, 1.9612, 1.9403, 1.9612, 1.9727)$$

and the upper bounds for the infection probabilities are

$$(.3401, .3390, .3411, .3350, .3364, .3377, .3377, .3401, .3377, .3364).$$

Compared to those in Scenario N5, the successful infection times are longer and the upper bounds are smaller. Figure 7(b) shows the simulated infection probabilities, which are smaller than those in Scenario N5. The numbers of incidents and losses for nodes are much less than the corresponding ones in Scenario N5. This indicates that the smaller variances would lead to less risk. The total loss in Scenario N6 is 81.239 compared to 307.752 in Scenario N5, a reduction of 73.6%. Therefore, we conclude that larger variances of infection processes result in larger risks.

4.2 Dependent Cybersecurity Risks

In this section, we study the dependence effect on the evolution of epidemic spreading and related losses. The Gaussian copula in Eq. (3.7) and the Clayton copula in Eq. (3.8) are considered in the simulation. For the Gaussian copula, we consider the cases of using the correlation matrices (3.6) and (3.11), without neighboring effects and with neighboring effects, respectively. For comparison, we consider the log-normal infection and recovery processes in what follows.

4.2.1 Gaussian Copula

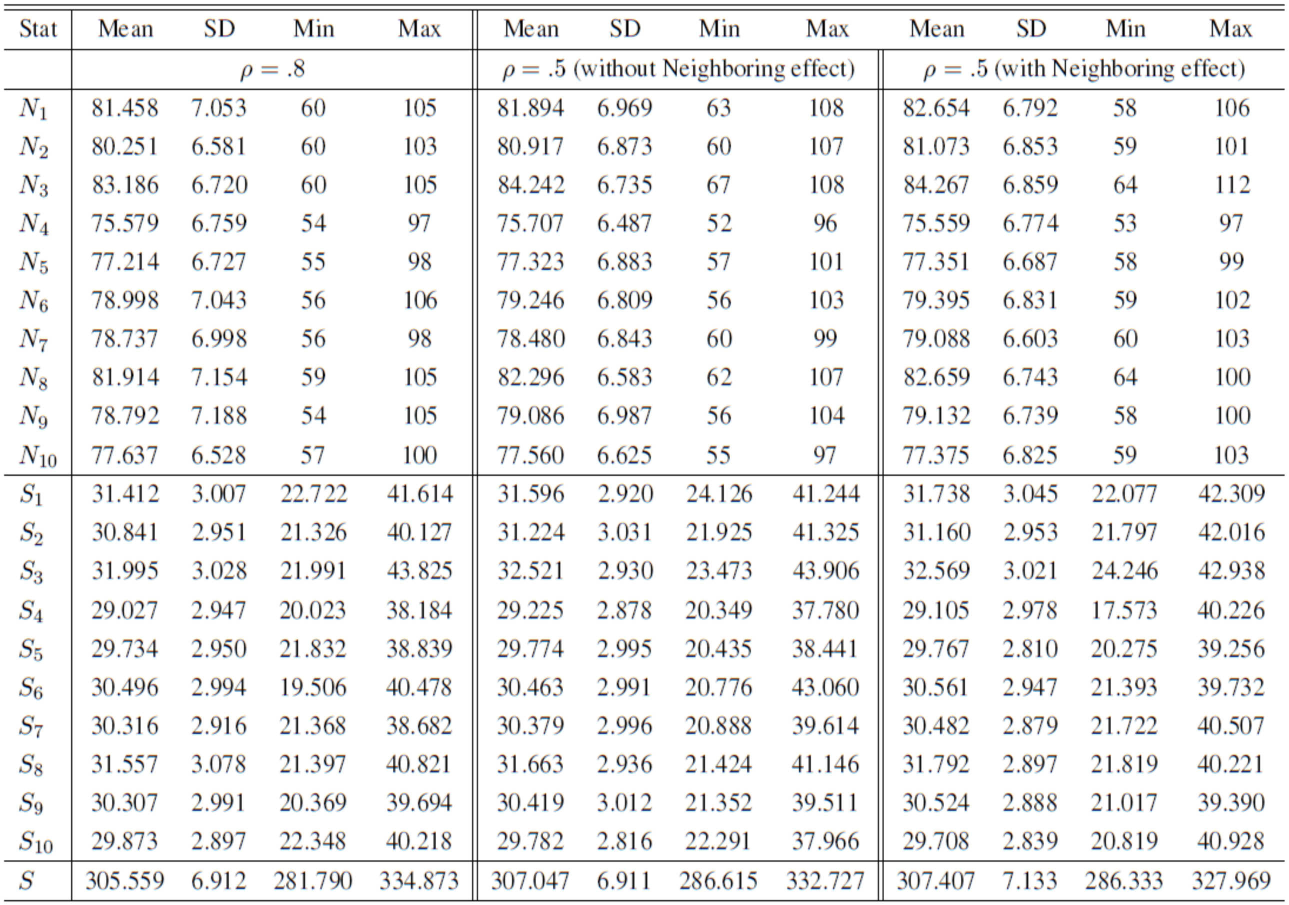
We assume that the parameter vector is as follows:

$$(\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_v, \sigma_v) = (1.1094, 1, .1931, 1, -.5, 1),$$

which is the same as that in Scenario N5 in Section 4.1.3. We consider three cases: $\rho = .8, \rho = .5$ (without neighboring effects), and $\rho = .5$ (with neighboring effects). It is known that a larger ρ implies more positive dependence. Further, the simulated infection probabilities in Figure 8(b) and 8(c) are slightly larger than those in Figure 8(a), and this indicates that the more positive the dependence, the smaller the infection probabilities. All of them show that node 3 has the largest overall infection probabilities, and node 4 has the smallest infection probabilities.

Table 4 shows the numbers of incidents and related losses. We can see that stronger dependence would lead to fewer overall losses. For example, the total loss for $\rho = .8$ is 305.559 while the total losses for $\rho = .5$ are 307.047 and 307.407, respectively. The difference between the two cases of $\rho = .5$, with and without the neighboring effects, is very subtle due to the assumed mechanism of the attack spreading process; that is, as we have assumed throughout the paper, as long as a node is infected, the other nodes will stop attacking it until it recovers. Therefore, dependence between neighbors plays only a moderate role, as Figure 8(a) and Table 4 have illustrated. A similar case can also be observed from the next case with Clayton copulas. Nevertheless, it is interesting to compare the losses in Table 4 to Table 3, and we observe that the independent case (i.e., $\rho = 0$) has the largest total loss.

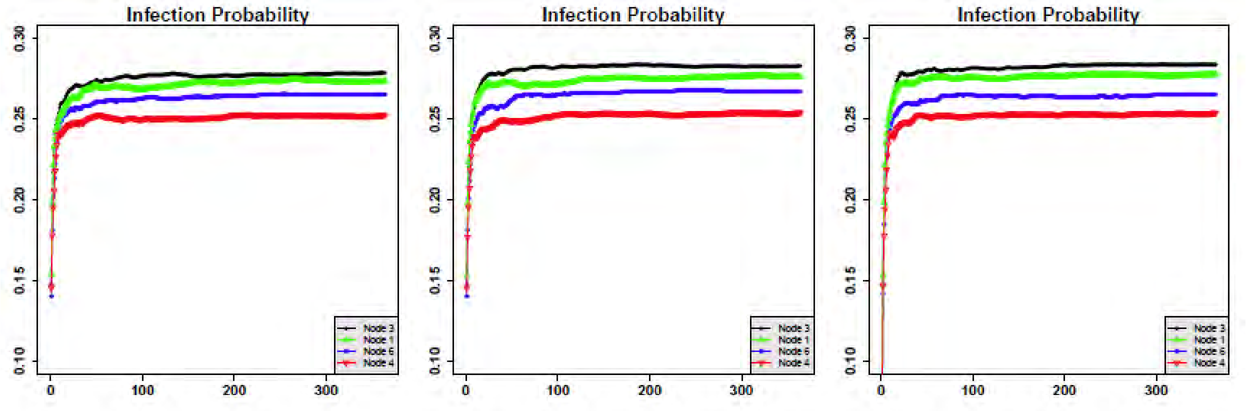
Table 4
Simulation Based on the Log-Normal Infection and Recovery Processes with Gaussian Copula, Where N_J Represents the Number of Infections for Node J during One Year and S_J the Total Loss for Node J during One Year. S Represents the Total Loss for the Network.



Stat	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
	$\rho = .8$				$\rho = .5$ (without Neighboring effect)				$\rho = .5$ (with Neighboring effect)			
N_1	81.458	7.053	60	105	81.894	6.969	63	108	82.654	6.792	58	106
N_2	80.251	6.581	60	103	80.917	6.873	60	107	81.073	6.853	59	101
N_3	83.186	6.720	60	105	84.242	6.735	67	108	84.267	6.859	64	112
N_4	75.579	6.759	54	97	75.707	6.487	52	96	75.559	6.774	53	97
N_5	77.214	6.727	55	98	77.323	6.883	57	101	77.351	6.687	58	99
N_6	78.998	7.043	56	106	79.246	6.809	56	103	79.395	6.831	59	102
N_7	78.737	6.998	56	98	78.480	6.843	60	99	79.088	6.603	60	103
N_8	81.914	7.154	59	105	82.296	6.583	62	107	82.659	6.743	64	100
N_9	78.792	7.188	54	105	79.086	6.987	56	104	79.132	6.739	58	100
N_{10}	77.637	6.528	57	100	77.560	6.625	55	97	77.375	6.825	59	103
S_1	31.412	3.007	22.722	41.614	31.596	2.920	24.126	41.244	31.738	3.045	22.077	42.309
S_2	30.841	2.951	21.326	40.127	31.224	3.031	21.925	41.325	31.160	2.953	21.797	42.016
S_3	31.995	3.028	21.991	43.825	32.521	2.930	23.473	43.906	32.569	3.021	24.246	42.938
S_4	29.027	2.947	20.023	38.184	29.225	2.878	20.349	37.780	29.105	2.978	17.573	40.226
S_5	29.734	2.950	21.832	38.839	29.774	2.995	20.435	38.441	29.767	2.810	20.275	39.256
S_6	30.496	2.994	19.506	40.478	30.463	2.991	20.776	43.060	30.561	2.947	21.393	39.732
S_7	30.316	2.916	21.368	38.682	30.379	2.996	20.888	39.614	30.482	2.879	21.722	40.507
S_8	31.557	3.078	21.397	40.821	31.663	2.936	21.424	41.146	31.792	2.897	21.819	40.221
S_9	30.307	2.991	20.369	39.694	30.419	3.012	21.352	39.511	30.524	2.888	21.017	39.390
S_{10}	29.873	2.897	22.348	40.218	29.782	2.816	22.291	37.966	29.708	2.839	20.819	40.928
S	305.559	6.912	281.790	334.873	307.047	6.911	286.615	332.727	307.407	7.133	286.333	327.969

Figure 8

Simulated infection probabilities of epidemic spreading over the network in Figure 1 based on log-normal infection and recovery processes with Gaussian copula.



(a) $\rho = .8$ (b) $\rho = .5$ (without neighboring effects) (c) $\rho = .5$ (with neighboring effects)

4.2.2 Clayton Copula

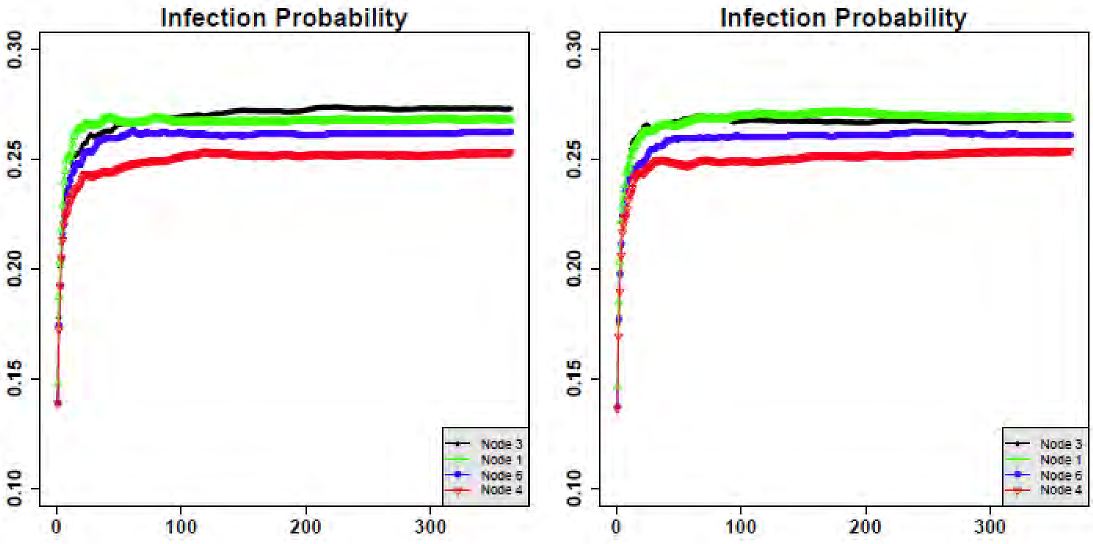
In this section, we discuss the case that the dependence structure could be modeled by Clayton copula. The parameter vector is set to be the same as the one in the previous section:

$$(\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_v, \sigma_v) = (1.1094, 1, .1931, 1, -.5, 1).$$

We consider two cases: $\theta = 2$ and $\theta = 20$. It is known in the literature that the larger value of θ's indicates the more positive dependence [14].

Figures 9(a) and 9(b) display the simulated infection probabilities for Clayton copulas. Here we see the evolutions of infections are similar in both cases. We also observe that the evolutions of node 3 and node 1 are very similar. This may be because nodes 3 and 1 are neighbors and both have a large number of degrees. The infections based on $\theta = 20$ are slightly less than that based on $\theta = 2$, which implies again that the more positive dependence among infection processes would lead to smaller infection probabilities.

Figure 9
Simulated infection probabilities of epidemic spreading over the network in Figure 1 based on log-normal infection and recovery processes with Clayton copula.


Infection Probability

Infection Probability

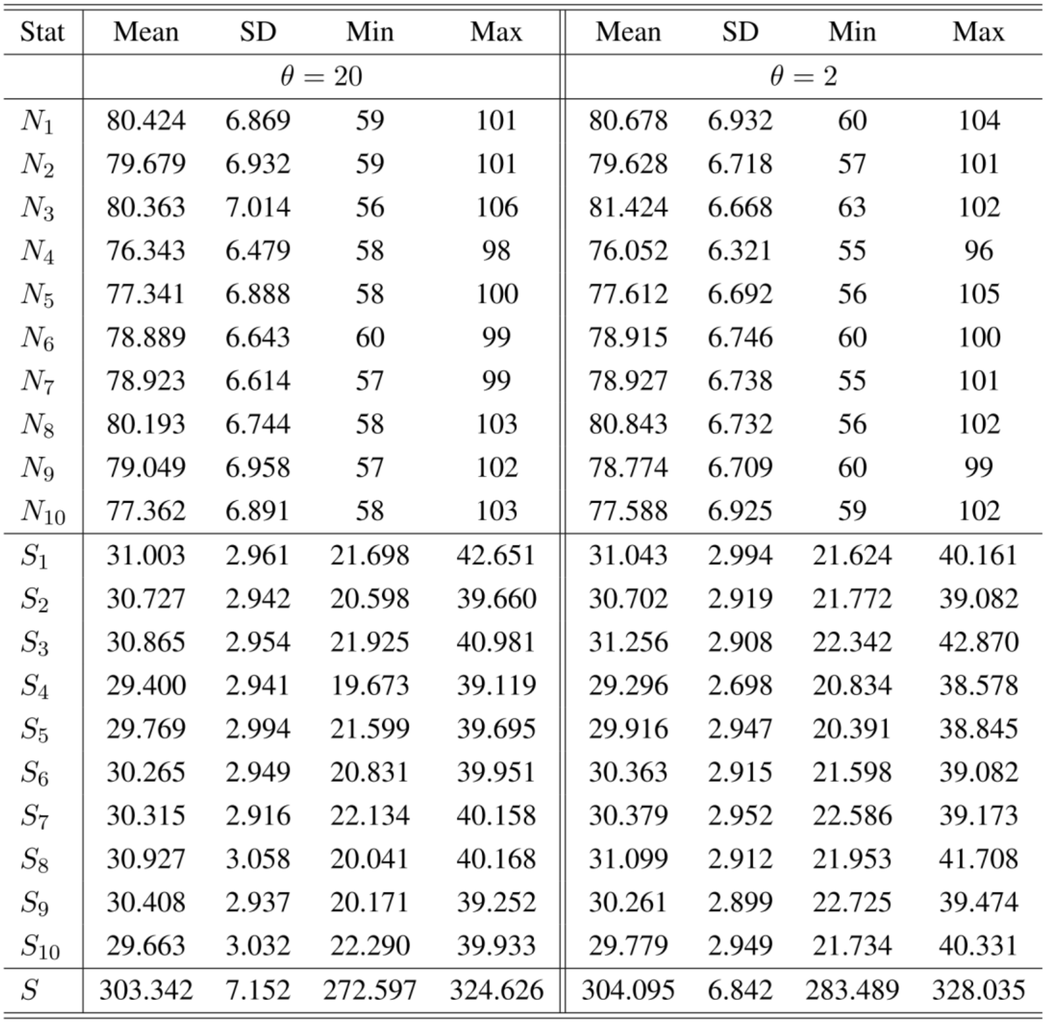
(a) $\theta = 2$

(b) $\theta = 20$

Table 5 shows the number of incidents and related losses. We again observe that more dependence results in less number of incidents and smaller losses. The total losses in both cases are smaller than that in the independence case (see Table 3).

Table 5

Simulation Based on the Log-Normal Infection and Recovery Processes with Clayton Copula, Where N_J Represents the Number of Infections and S_J Means the Total Loss for Node J during One Year. S Represents the Total Loss for the Network.



Stat	Mean	SD	Min	Max	Mean	SD	Min	Max
	$\theta = 20$				$\theta = 2$			
N_1	80.424	6.869	59	101	80.678	6.932	60	104
N_2	79.679	6.932	59	101	79.628	6.718	57	101
N_3	80.363	7.014	56	106	81.424	6.668	63	102
N_4	76.343	6.479	58	98	76.052	6.321	55	96
N_5	77.341	6.888	58	100	77.612	6.692	56	105
N_6	78.889	6.643	60	99	78.915	6.746	60	100
N_7	78.923	6.614	57	99	78.927	6.738	55	101
N_8	80.193	6.744	58	103	80.843	6.732	56	102
N_9	79.049	6.958	57	102	78.774	6.709	60	99
N_{10}	77.362	6.891	58	103	77.588	6.925	59	102
S_1	31.003	2.961	21.698	42.651	31.043	2.994	21.624	40.161
S_2	30.727	2.942	20.598	39.660	30.702	2.919	21.772	39.082
S_3	30.865	2.954	21.925	40.981	31.256	2.908	22.342	42.870
S_4	29.400	2.941	19.673	39.119	29.296	2.698	20.834	38.578
S_5	29.769	2.994	21.599	39.695	29.916	2.947	20.391	38.845
S_6	30.265	2.949	20.831	39.951	30.363	2.915	21.598	39.082
S_7	30.315	2.916	22.134	40.158	30.379	2.952	22.586	39.173
S_8	30.927	3.058	20.041	40.168	31.099	2.912	21.953	41.708
S_9	30.408	2.937	20.171	39.252	30.261	2.899	22.725	39.474
S_{10}	29.663	3.032	22.290	39.933	29.779	2.949	21.734	40.331
S	303.342	7.152	272.597	324.626	304.095	6.842	283.489	328.035

To conclude this section, we observe that dependence among infection processes affects the evolution of epidemic spreading and related losses. Stronger positive dependence among time-to-infection random variables would result in fewer incidents and losses. One interpretation is that a stronger positive dependence structure tends to give rise to longer waiting times to infection, and thus fewer infection events and losses. Since the high-dimension dependence for a complex network topology can be very challenging in practice and the challenge is further increased by the lack of enough cybersecurity data, the independent model may be used in practice as conservative estimates in the aforementioned cases.

4.3 Pricing

In this section, we discuss the premiums for the node level and the company level, respectively. The premiums are calculated based on two premium principles: (i) standard deviation premium principle in Eq. (4.2); and (ii) principle of equivalent utility in Eq. (4.3). For each principle, we consider three scenarios based on log-normal infection and recovery processes discussed in the previous sections: (1) independent model in Eq. (4.4); (2) Gaussian dependent model with $\rho = .8$; and (3) Clayton dependent model with $\theta = 20$. For principle (i), $\lambda = .2$, and for principle (ii), $\gamma = .8$. We assume that the parameter vector is as follows:

$$(\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_v, \sigma_v) = (1.1094, 1, .1931, 1, -.5, 1),$$

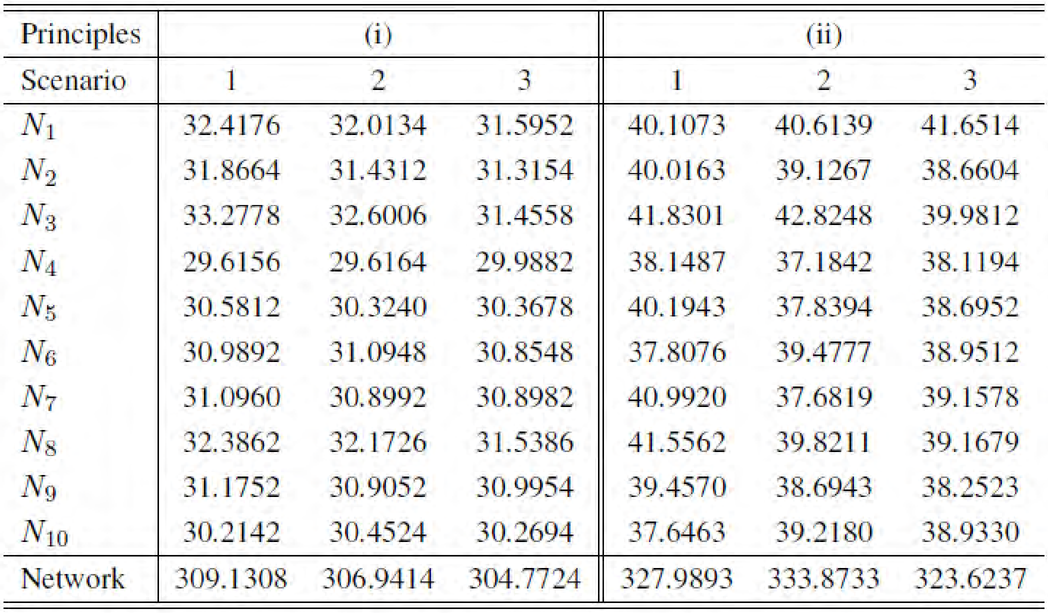
which is the same as that in Scenario N5 in Section 4.1.3.

Table 6 shows the premiums for each principle. For principle (i), we observe that node 3 is charged with the largest premiums for Scenarios 1 and 2. For Scenario 3, we see that the premiums for nodes 8, 1 and 3 are close while the premium for node 1 is the largest. Node 4 has the smallest premium for all the scenarios. From the network level, Scenario 1 has the largest premium, 309.1308, which is the independent scenario. Therefore, for principle (i), the independence model may be used for conservative pricing. The premiums charged based on principle (ii) are overall larger than those based on principle (i), and this is caused by risk aversion. We again observe that node 3 is charged with the largest premiums for Scenarios 1 and 2, and node 1 has a slightly larger premium than that of node 3 in Scenario 3. For the network level, it is interesting to observe that Scenario 2, the Gaussian copula with $\rho = .8$, has the largest premium, 333.8733, which may be due to the large variability and risk aversion utility in this case (e.g., the minimum loss is 281.790, and the largest loss is 334.873; see Table 4).

Due to the dynamic nature of epidemic spreading, it is infeasible to compute the theoretical premiums for nodes. However, the theoretical premiums based on the upper bound in Eq. (3.14) may be calculated. In the following discussion, we examine the theoretical premiums based on principle (i). After some tedious calculations, we have the premiums for nodes based on the upper bound as

$$(63.9958, 62.2444, 66.4075, 51.2775, 56.0738, 59.57678, 59.671, 64.0228, 59.5767, 55.9256).$$

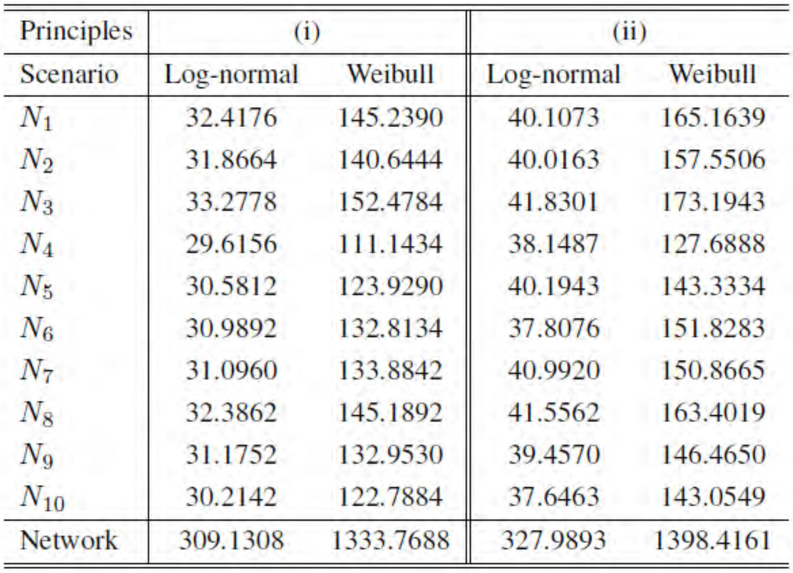
Table 6
Premiums for Each Computer and the Network Based on Two Different Premium Principles



Principles	(i)			(ii)		
Scenario	1	2	3	1	2	3
N_1	32.4176	32.0134	31.5952	40.1073	40.6139	41.6514
N_2	31.8664	31.4312	31.3154	40.0163	39.1267	38.6604
N_3	33.2778	32.6006	31.4558	41.8301	42.8248	39.9812
N_4	29.6156	29.6164	29.9882	38.1487	37.1842	38.1194
N_5	30.5812	30.3240	30.3678	40.1943	37.8394	38.6952
N_6	30.9892	31.0948	30.8548	37.8076	39.4777	38.9512
N_7	31.0960	30.8992	30.8982	40.9920	37.6819	39.1578
N_8	32.3862	32.1726	31.5386	41.5562	39.8211	39.1679
N_9	31.1752	30.9052	30.9954	39.4570	38.6943	38.2523
N_{10}	30.2142	30.4524	30.2694	37.6463	39.2180	38.9330
Network	309.1308	306.9414	304.7724	327.9893	333.8733	323.6237

Compared to the first column of Table 6, the independent case, it is seen that the premiums are rather conservative, i.e., much larger than the simulation results. Therefore, we recommend using the premiums based on simulations in practice, while the upper bound can be employed for worst-scenario testing. The other interesting question is to compute the premiums based on the misspecified distributions. In the following discussion, we assume that the real scenario is the aforementioned independent log-normal distributions although it is misspecified as Weibull distributions. Specifically, we assume that the misspecified Weibull distributions have the same means and variances as those of log-normal distributions. Then, we simulate the premiums based on principles (i) and (ii), respectively. The results are presented in Table 7. For comparison, we also copy the premiums based on log-normal distribution in Tables 6 to 7.

Table 7
Log-Normal and Weibull Premiums for Each Computer and the Network Based on Two Different Premium Principles



Principles	(i)		(ii)	
Scenario	Log-normal	Weibull	Log-normal	Weibull
N_1	32.4176	145.2390	40.1073	165.1639
N_2	31.8664	140.6444	40.0163	157.5506
N_3	33.2778	152.4784	41.8301	173.1943
N_4	29.6156	111.1434	38.1487	127.6888
N_5	30.5812	123.9290	40.1943	143.3334
N_6	30.9892	132.8134	37.8076	151.8283
N_7	31.0960	133.8842	40.9920	150.8665
N_8	32.3862	145.1892	41.5562	163.4019
N_9	31.1752	132.9530	39.4570	146.4650
N_{10}	30.2142	122.7884	37.6463	143.0549
Network	309.1308	1333.7688	327.9893	1398.4161

It is seen that the premiums based on the Weibull distributions are very high for both principles. This indicates that the premiums are very sensitive to the specifications of attack and recovery processes that need to be carefully selected in practice.

In conclusion, different premium principles result in significantly different premiums. If the dependence is unknown, the independent model may be used as a conservative approximation. The specifications of attack and recovery distributions are critical in determining the premiums.

Section 5: Conclusion

Cyber attacks can lead to different types of losses, such as loss of information, loss of revenue, loss of service, and recovery costs. The current work makes a significant contribution to modeling cybersecurity insurance. We propose a novel cybersecurity insurance model, one that models not only the general infection and recovery processes but also the related losses. Moreover, the proposed model employs copulas to account for dependence among infection processes. We derive the dynamic upper bounds for the infection probabilities that may be used as conservative estimates. For pricing purposes, we propose a simulation approach to study the evolution of cyber risks. Three quantities are calculated based on simulations, and those include the number of incidents, infection probabilities, and total loss for the network. This information would help insurance companies to price the cybersecurity insurance products.

We also discuss two different premium principles for calculating premiums based on simulations. Granular information about the network topology and granular data for historical loss events will be helpful in improving the accuracy of rates. Nevertheless, the proposed framework for modeling and pricing of cyber risks for such a network-based system can also be used as a scoring system for the purpose of internal and external cyber risk management.

The proposed approach can be considered as microlevel modeling of cybersecurity risks. That is, the dynamics of attack and recovery processes are modeled, and the related losses are simulated. This proposed approach relies on the underlying stochastic processes and epidemic theory, and it may require a large number of simulations based on the scale and complexity of the network. Some other interesting future research includes exploration into the macrolevel modeling of cybersecurity risks. That is, it becomes feasible to use information of network configurations, network flows, historical cyber incidents, security protocols, and so forth to develop statistical models for modeling and predicting cybersecurity risks, and, therefore, risk assessments for a large-scale network.

References

[1] A. Barrat, M. Barthlemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge: Cambridge University Press, 2008.

[2] R. S. Betterley. Cyber/privacy insurance market survey: A tough market for larger insureds, but smaller insureds finding eager insurers. *The Betterley Report*, June 2016.

[3] R. Bohme and G. Kataria. Models and measures for correlation in cyber-insurance. In *Fifth Presented at Workshop on the Economics of Information Security*, University of Cambridge, UK, June 2006.

[4] R. Bohme and G. Schwartz. Modeling cyber-insurance: Towards a unifying framework. In *Ninth Presented at Workshop on the Economics of Information Security*, Harvard, June 2010.

[5] E. Cator, R. Van de Bovenkamp, and P. Van Mieghem. Susceptible-infected-susceptible epidemics on networks with general infection and cure times. *Physical Review E*, 87(6):062816, 2013.

[6] E. Cator and P. Van Mieghem. Nodal infection in Markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated. *Physical Review E*, 89(5):052802, 2014.

[7] E. A. Coddington. *An Introduction to Ordinary Differential Equations*. North Chelmsford, MA: Courier Corporation, 2012.

[8] US Department of Homeland Security. Cybersecurity insurance. https://www.dhs.gov/cybersecurity-insurance.

[9] C. Doerr, N. Blenn, and P. Van Mieghem. Log-normal infection times of online information spread. *PloS One*, 8(5):e64349, 2013.

[10] M. Eling and W. Schnell. What do we know about cyber risk and cyber risk insurance? *The Journal of Risk Finance*, 17(5), 2016.

[11] L. A. Gordon, M. P. Loeb, and T. Sohail. A framework for using insurance for cyber- risk management. *Communications of the ACM*, 46(3):81–85, 2003.

[13] V. S. B. Herath and T. C. Herath. Copula-based actuarial model for pricing cyber-insurance policies. *Insurance Markets and Companies: Analyses and Actuarial Computations*, 2:7–20, 2011.

[14] H. Joe. *Dependence Modeling with Copulas*. Boca Raton, FL: CRC Press, 2014.

[15] S. Karlin. *A First Course in Stochastic Processes*. Cambridge, MA: Academic Press, 2014.

[16] T. Kosub. Components and challenges of integrated cyber risk management. *Zeitschrift fur die gesamte Versicherungswissenschaft*, 104(5):615–634, 2015.

[17] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, and S. K. Sadhukhan. e-Risk management with insurance: A framework using copula aided Bayesian belief networks. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, vol. 6, 126.1–126.6. Hoboken, NJ: IEEE, 2006.

[18] R. B. Nelsen. *An Introduction to Copulas*. Vol. 139. New York: Springer Science & Business Media, 2013.

[19] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, 2015.

[20] J. W. Pratt. Risk aversion in the small and in the large. In *Foundations of Insurance Economics*, 83–98. New York: Springer, 1992.

[21] S. Ross. *Stochastic Processes*. Hoboken, NJ: Wiley and Sons, 1996.

[22] G. A. Schwartz and S. S. Sastry. Cyber-insurance framework for large-scale interdependent networks. In *Proceedings of the Third International Conference on High Confidence Networked Systems*, 145–154. New York: ACM, 2014.

[23] A. Sklar. Distribution Functions of n dimensions and their margins. *Publications de l'Institut de statistique de l'Universite' de Paris*, 8:229–231, 1959.

[24] P. Van Mieghem and R. Van de Bovenkamp. Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. *Physical Review Letters*, 110(10):108701, 2013.

[25] P. Van Mieghem. *Performance Analysis of Complex Networks and Systems*. Cambridge: Cambridge University Press, 2014.

[26] P. Van Mieghem and E. Cator. Epidemics in networks with nodal self-infection and the epidemic threshold. *Physical Review E*, 86(1):016116, 2012.

[27] M. Xu, G. Da, and S. Xu. Cyber epidemic models with dependences. *Internet Mathematics*, 11(1):62–92, 2015.

[28] M. Xu and S. Xu. An extended stochastic model for quantitative security analysis of networked systems. *Internet Mathematics*, 8(3):288–320, 2012.

[29] Z. Yang and J. C. S. Lui. Security adoption and influence of cyber-insurance markets in heterogeneous networks. *Performance Evaluation*, 74:1–17, 2014.

About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world, dedicated to serving more than 27,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policy makers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

Objectivity: The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

Quality: The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

Relevance: The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

Quantification: The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org