A

On the Assessment of Systematic Risk in Networked Systems

ARON LASZKA, University of Houston
BENJAMIN JOHNSON, Technical University of Munich
JENS GROSSKLAGS, Technical University of Munich

In a networked system, the risk of security compromises depends not only on each node's security, but also on the topological structure formed by the connected individuals, businesses, and computer systems. Research in network security has been exploring this phenomenon for a long time, with a variety of modeling frameworks predicting how many nodes we should expect to lose, on average, for a given network topology, after certain types of incidents. Meanwhile the pricing of insurance contracts for risks related to information technology (better known as cyber-insurance) requires determining additional information, for example, the maximum number of nodes we should expect to lose within a 99.5% confidence interval. Previous modeling research in network security has not addressed these types of questions, while research on cyber-insurance pricing for networked systems has not taken into account the network's topology. Our goal is to bridge that gap, by providing a mathematical basis for the assessment of systematic risk in networked systems.

We define a *loss-number distribution* to be a probability distribution on the total number of compromised nodes within a network following the occurrence of a given incident; and we provide a number of modeling results that aim to be useful for cyber-insurers in this context. We prove NP-hardness for the general case of computing the loss-number distribution for an arbitrary network topology, but obtain simplified computable formulas for the special cases of star topologies, ER-random topologies, and uniform topologies. We also provide a simulation algorithm that approximates the loss-number distribution for an arbitrary network topology and that appears to converge efficiently for many common classes of topologies.

Scale-free network topologies have a degree distribution that follows a power law, and are commonly found in real-world networks. We provide an example of a scale-free network in which a cyber-insurance pricing mechanism that relies naively on incidence reporting data will fail to accurately predict the true risk level of the entire system. We offer an alternative mechanism that yields an accurate forecast by taking into account the network topology, thus highlighting the lack/importance of topological data in security incident reporting. Our results constitute important steps towards the understanding of systematic risk, and help to contribute to the emergence of a viable cyber-insurance market.

CCS Concepts: •**Security and privacy** → **Economics of security and privacy;**

Additional Key Words and Phrases: Networks, security, topology, cyber-insurance, risk mitigation, economics of security, scale-free networks

ACM Reference Format:
Aron Laszka, Benjamin Johnson, and Jens Grossklags, 2017. On the assessment of systematic risk in networked systems. *ACM Trans. Internet Technol.* V, N, Article A (January YYYY), 29 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Computer systems, businesses, and individuals often form networks. Computers, for example, are connected by physical and logical links; businesses provide services to one another; and individuals make friends and acquaintances encompassing various implicit levels of trust.

Authors' addresses: A. Laszka, Department of Computer Science, University of Houston; B. Johnson and J. Grossklags, Chair of Cyber Trust, Department of Informatics, Technical University of Munich.

DOI: http://dx.doi.org/10.1145/0000000.0000000

While these networks can be very beneficial, they can also increase risks, as attackers are often able to exploit the access and trust relationships that network connections entail. For example, in 2011, RSA, a major security company, was compromised; and information on about 40 million SecureID tokens were stolen. This successful compromise was later used to attack Lockheed Martin, one of the world's largest defense contractors [Drew 2011]. As another example, hackers calling themselves the Syrian Electronic Army sent e-mails to Financial Times employees containing phishing links, which were used to gain access to FT.com corporate e-mail accounts. These accounts were then used to propagate the social engineering attack to a larger number of FT.com users, eventually compromising the organization's website and Twitter account [Betts 2013]. More recently, the perpetrators of the Energetic Bear cyber-espionage campaign exploited interdependence between energy companies and industrial control system (ICS) manufacturers [Symantec 2014]. In order to penetrate highly-secure targets, such as energy grid and petroleum pipeline operators in the U.S., Germany, and other countries, the attackers compromised ICS manufacturers. Then, they inserted malware into software updates distributed by these manufacturers, which were downloaded and applied by the targets, leading to their compromise.

These examples serve to illustrate that implicit trust from network connections can be used to compromise trusting neighbors through attacks on their peers. From the attacker's perspective, the network structure gives rise to what we might term systematic opportunity, because the opportunity for an attacker to strike a large payoff is a consequence of the system itself. Correspondingly, the users of such systems become subject to *systematic risks*, arising from the structure of their connections.

These systematic artifacts can have consequential effects on the motivations of users of such systems, as they recognize that their security is dependent on the investments of their peers. The resulting environment gives rise to well-documented problems such as under-investment or free-riding [Laszka et al. 2014a], and it may also motivate users to consider alternative risk-mitigation strategies such as purchasing insurance.

Insurance is a promising remedy to many risk-related problems because it facilitates risk diversification; however, structural consequences of networked systems can also affect insurers. Traditionally, insurance is based on the diversifiability of risks: if an insurance provider has enough clients, the variabilities from individual risks cancel, and the aggregate risk is predictable. But if individual risks are correlated, then even for a large number of clients, there may be a non-negligible probability of a catastrophic event in which many clients are compromised at the same time.

This risk of catastrophe is a consequence of the network structure formed by the connected individuals, businesses and computer systems; and this causal relationship warrants our attention. However, to the best of our knowledge, the effects of a network's connective structure on risk-mitigation concerns that would be relevant to a cyber-insurance provider have not been researched. For example, Lelarge and Bolot model interdependent security with insurance, but assume that there is an insurance provider with an *exogenously* priced premium [Lelarge and Bolot 2009],[1] thus sidestepping the question of whether an insurance provider would be willing to offer such a contract. Many elements for understanding the relationship between a network's structure and the resulting risk to its components can be found in related work addressing cyber-insurance, models of interdependent security, or properties of scale-free networks. But a persistent research gap remains.

Contributions: In this paper, we provide a mathematical basis for studying the distribution of the number of compromised nodes from within a set of interconnected nodes, after individual risk propagates through a network structure following a secu-

[1]This means that the price is a parameter of the model that is not chosen by any of the players.

rity event. We illustrate and explain why network-wide risk-mitigation solutions, such as cyber-insurance, must consider the variability of the number of compromised nodes; and that in contrast to its expected value, the variability of the number of compromised nodes cannot be naively estimated from sampling a small part of the network. This failure is especially interesting from a practical point of view, as many real-world business and social networks are resilient against comprehensive data collection, so that the only viable prediction mechanism for determining the risk portfolio of these networks relies on extrapolation from smaller samples.

This article extends our previous papers [Laszka et al. 2014b; Johnson et al. 2014b; Johnson et al. 2014a] with the following new contributions:

(1) We introduce and study a multiple-hop variant of our propagation model.
(2) We provide a new proof for the NP-hardness of computing the loss-number distributions, which—in contrast to the previous proof—covers the multiple-hop model.
(3) Finally, we provide a new example involving the multi-hop model applied to scale-free network topologies to illustrate the biased nature of random samples.

Organization: The rest of the paper is organized as follows. In Section 2, we discuss related work from the areas of risk mitigation, interdependent security, and network structures. Section 3 introduces our network risk propagation models, which are derived from existing models from the literature on interdependent security games. In Section 4, we address the computational complexity of computing the distribution of the number of compromised nodes for these models. Section 5 addresses the question of systematic risk for scale-free networks. Finally, Section 6 concludes the paper. Additional illustrations may be found in Appendix A. Additional proofs may be found in Appendix B.

2. RELATED WORK

First, we present current challenges in risk mitigation with a focus on risk-transfer mechanisms and specifically cyber-insurance. Then, we summarize previous work on interdependent security models, which model how risk is propagated between connected nodes, the main concern of our study. Finally, we discuss scale-free networks, which realistically model many real-world networks and which form the basis of our simulation-based analysis.

2.1. Cyber-Insurance

Markets for risk-transfer mechanisms, such as cyber-insurance, suffer from the difficulty to correctly assess systematic risk in networks. A functioning market for cyber-insurance and a good understanding of the insurability of networked resources are both important because they signal that stakeholders are able to manage modern threats as succinctly stated in Anderson's principle that "[a] trusted component or system is one which you can insure [Anderson 1994; Böhme 2010]." However, the cyber-insurance market is developing at a slow pace due to a number of factors and is still not fully understood from an economic modeling perspective. (See, in particular, the survey by Böhme and Schwartz [Böhme and Schwartz 2010]).

A primary difficulty for insurance providers is risk correlation. A group of defenders might appear as a particularly appealing target to an attacker because of a high correlation in their risk profiles. For example, even though individual computer systems may be independently owned and administrated, they may exhibit highly homogeneous software configurations which in turn can be vulnerable to the same attack vector [Birman and Schneider 2009; Geer et al. 2003]. Böhme and Kataria as well as Chen et al. study the impact of correlations that are due to such so-called monoculture risks [Böhme and Kataria 2006; Chen et al. 2011].

Our research is complementary to the studies cited above: they investigate (the effect of) correlations arising from nodes having the same software configurations, while we study how correlations arise from nodes being connected to each other.

2.2. Scale-Free Networks

Many real-world networks are believed to be scale-free, including social, financial, and biological networks [Barabási 2009]. A scale-free network is one whose degree distribution approximates a power law distribution. That is, the fraction of nodes in the network having degree k approximates $k^{-\gamma}$ for large k. Here γ is a constant parameter.

Recent interest in scale-free networks started with [Barabási and Albert 1999], in which the Barabási-Albert (BA) model is introduced for generating random scale-free networks. The BA model is based on two concepts: network growth and preferential node attachment. We discuss this model in detail in Section 5. Li et al. introduce a new, mathematically more precise, and structural definition of scale-free graphs [Li et al. 2005], which promises to offer a more rigorous and quantitative alternative. The networks discussed in our paper satisfy this definition as well.

One of the most important questions addressed by our paper is whether small samples can be used to predict systematic risks in scale-free networks. Stump et al. show that the degree distributions of randomly sampled subnets of scale-free networks are not scale-free [Stumpf et al. 2005]; thus, subnet data cannot be naively extrapolated to every property of the entire network. However, random samples are unbiased estimators of some properties (e.g., average degree). In Section 5, we investigate whether they are unbiased estimators of systematic risk.

2.3. Interdependent Security

The notion of correlated risks can be extended to capture the underlying interdependent nature of networks. That is, the mere vulnerability of a large number of systems to a particular attack is less significant if an attacker cannot easily execute a sufficiently broad attack. One key attack method to cause wide-ranging compromises is attack propagation, but an attacker can also directly attack multiple defenders, e.g., by using some form of attack automation.

Interdependence has been considered in different ways in the academic literature [Laszka et al. 2014a]. Varian, for example, studies a security-compromise setting where an overall prevention objective (i.e., security is a public good) depends on the (investment-prevention) contributions of independently-owned systems [Varian 2004]. In his model, security compromises are often the result of misaligned incentives of the independently-owned systems which manifest as coordination failures, such as free-riding on others' prevention investments. Grossklags et al. extend this work by allowing independently-owned systems to do a private investment in system recovery, and they find that it can serve as a viable investment strategy to sidestep such coordination failures [Grossklags et al. 2008]. However, the availability of system recovery will further undermine incentives for the overall prevention objective. Johnson et al. add the availability of cyber-insurance to this modeling framework, and identify solution spaces in which these different investment approaches may be used as bundled security strategies [Johnson et al. 2011]. However, due to the fact that those models capture two security outcomes (i.e., everybody is compromised, or nobody is compromised), they can only serve as approximate guidance for realistic insurance models.

A second group of economic models derives equilibrium outcomes for decisions by independently-owned systems to inoculate themselves (or not) against a compromise by a virus or other attack in a network, and thereby also to contain the propagation of the attack. For example, the models by Aspnes et al. as well as Moscibroda et al. would be applicable to the study of loss distributions, however, several simplifying

assumptions in those models limit the generality of the results [Aspnes et al. 2006; Moscibroda et al. 2006]. Those limitations include the assumption that every infected node deterministically infects all unprotected neighbors.

A third class of propagation models that has been widely studied is the class of epidemic models, which describe the process how a virus spreads or extinguishes in a network. In the literature on epidemic models, the results of Kephart and White [Kephart and White 1991] are the closest to our analysis. They study one of the simplest of the standard epidemic models, the susceptible-infected-susceptible (SIS) model, using various classes of networks. The SIS model captures the initial infection process, but also allows for recovery of the nodes (and eventual reinfection of nodes after recovery). For example, for Erdős-Rényi random graphs [Erdős and Rényi 1959; Erdős and Rényi 1960], they approximate both the expected value and the variance of the number of infected nodes using formulas. For the more realistic hierarchical network model, they show that the expected number of infected nodes does not increase with the number of nodes in the graph, but that there are scenarios in which the number of infected nodes fluctuates significantly in an irregular fashion. Kephart and White extend their analysis also to other variations of epidemiological models in follow-up research [Kephart and White 1993]. Pastor-Satorras and Vespignani analyze real data from computer virus infections in order to define a dynamical SIS model for epidemic spreading in scale-free networks [Pastor-Satorras and Vespignani 2001]. Likewise, Eguíluz and Klemm study the spreading of viruses in scale-free networks with high clustering and degree correlations (between the degree of a node and the degrees of its neighbors) [Eguiluz and Klemm 2002]. Pastor-Satorras and Vespignani study epidemic dynamics in finite-size scale-free networks, and show that, even for relatively small networks, the epidemic threshold is much smaller than that of homogeneous systems [Pastor-Satorras and Vespignani 2002]. Wang et al. propose a general epidemic threshold condition, which applies to arbitrary graphs, based on the largest eigenvalue of the adjacency matrix [Wang et al. 2003; Chakrabarti et al. 2008]. In a follow-up work, Ganesh et al. obtain the same epidemic threshold result (along with other results) using another approach [Ganesh et al. 2005].

Finally, a popular approach to model interdependent risk is taken by Kunreuther and Heal, and forms the basis for our formal analysis [Kunreuther and Heal 2003; Heal and Kunreuther 2004]. The basic premise of this work is to separately consider the impact of direct attacks and propagated attacks. We explain the details of the model in Section 3. The model has been generalized to consider distributions of attack probabilities [Johnson et al. 2010] and a strategic attacker [Chan et al. 2012]. Similarly, Ogut et al. proposed a related model that allows for continuous (rather than binary) security investments [Ogut et al. 2005]. Lelarge and Bolot study another extension of the model, which is more similar to our multiple-hop model [Lelarge and Bolot 2008a; Lelarge and Bolot 2008b]. In a follow-up work, they investigate the supply-side of insurance [Lelarge 2009]; however, they do not consider the correlation of risks. Our analysis setup draws from these extensions by implicitly considering a continuum of risk parameters to study the distribution of outcomes.

3. MODEL OVERVIEW

Our modeling framework builds on the interdependent security game introduced by Kunreuther and Heal [Kunreuther and Heal 2003; Heal and Kunreuther 2004]. This model has been studied and extended by many authors (e.g., [Chan et al. 2012; Dhall et al. 2009; Johnson et al. 2010; Kearns and Ortiz 2004]), with a common focus on understanding how individuals in a networked system make security investment decisions in response to potential threats, along with how these decisions affect other individuals in the network.

Although we use the risk propagation structure of this model, our focus is different from prior work. We concentrate exclusively on properties of the network's loss distribution, and study how this distribution is shaped by the direct risk probabilities of individual nodes as well as by the risk propagation probabilities between neighbors.

More specifically, we focus most of our attention on the number of compromised nodes in an outcome of the model, with special attention given to the probability distribution over this number. We refer to this distribution as the *loss-number distribution*.

For a list of symbols used in this paper, see Table I in Appendix A.

3.1. One-Hop Network Risk Propagation Model

We begin with introducing the original risk propagation model proposed by Kunreuther and Heal, to which we will refer as the one-hop model. At a high level, this model describes risk effects of any networked system in which security breaches may have two levels of effect. First, a security breach at a network node damages the breached node itself. Second, the perpetrators may also use the breached node as a "digital beachhead," and exploit the breached node's network connections to compromise and damage its neighbors as well. This propagation structure yields a simple mechanism for studying network risk, and it captures the core dynamics of risk transfer from a node-centric perspective. Any risk to a certain node may be categorized as either originating outside the network or originating within the network. If the risk originates outside the network, we may categorize it in terms of the probability of a direct security breach; while if the risk originates from within the network, from one of its neighbors, we may quantify the magnitude of the risk derived from that connection.

Consider a network of N nodes. Each node has two types of connections: one type which connects to other nodes in the same network, and another type which connects to a system outside the network. For an illustration, see Figure 9 in Appendix A. Threats originate outside the network, and subject each node to some risk of compromise. If an outside threat successfully reaches a node, that node is compromised. This outcome is binary so that each node is either compromised or not.

If a node is compromised, the risk may propagate within the network to that node's direct neighbors. In our interpretation of the model, the risk does not propagate further than one hop, so that each node is threatened only by external threats against itself and its immediate neighbors. While this model does not encompass all conceivable multiple-hop propagation structures, it strikes a good balance between realistic risk transfer properties and conceptual simplicity since it captures a wide range of interdependent security phenomena [Laszka et al. 2014a], yet it is analytically tractable.

Risk of direct compromise threatens each node i with probability p_i, and for the analysis we assume that direct compromises for different nodes are independent events. Our framework is agnostic about the origin of direct risk, although it could be motivated in an active attacker model by assuming that each node has a different attacker.

If node i is directly compromised, it transfers risk to each neighboring node j with probability q_{ij}. If a node is not directly compromised, it cannot transfer risk to any other node. Notice that we can use an $N \times N$ matrix Q, whose elements are the risk transfer probabilities q_{ij}, to directly specify network topology alongside risk propagation simply by requiring $q_{ij} = 0$ whenever node i is not connected to node j.

3.2. Multiple-Hop Network Risk Propagation Model

The propagation of compromises in the model of Kunreuther and Heal, which we introduced in the previous subsection, is essentially limited to one hop: nodes that are directly compromised can transfer risk to their immediate neighbors; however, non-directly compromised nodes cannot transfer the risk any further. While a one-hop model is suitable for many scenarios (e.g., when attacks are propagated manually by

the attacker and, thus, are limited), there are many threats (e.g., computer worms) which can be propagated over multiple hops.

In our multiple-hop propagation model, every node that becomes compromised (either directly or non-directly) may be the origin of a non-direct compromise of its neighbors. More specifically, a compromised node i has a one-time chance of compromising each neighboring node j independently with probability q_{ij}. Note that even if a node could have become compromised due to multiple neighbors (or also due to direct compromise), it has only one chance to propagate the compromise through each outgoing link. Equivalently, we could define the multiple-hop model as follows: each link (i, j) is present in the network with probability q_{ij}, and a node becomes compromised if there is a directed path leading to it from a directly compromised node.

Other natural extensions of our one-hop model include the class of models in which risk is allowed to be transferred exactly n times. Our current work focuses instead on extensions allowing unlimited propagation for essentially two reasons. First, it is much easier to find practical interpretations for models with unlimited propagation, (e.g., viruses), than for exactly-n-hop models with $n > 1$. Second, the most direct n-hop extension to our model – the one which differs only in the number of hops risk is allowed to transfer – is equivalent to a one-hop model in which the matrix $[q_{ij}]$ of risk transfer probabilities between nodes, is replaced by its power graph $[q_{ij}]^n$. In contrast, our multi-hop extension is defined in such a way that there is no obvious direct translation to the one-hop case.

3.3. Game-Theoretic Actors

Many studies have used the one-hop model to understand interdependent security by considering a game in which each individual can reduce the risk of her own node by making a security investment. In this case, actors with various motivations make choices that may decrease the node-centric risk parameters p_i and q_{ij} (e.g., host-based IDS or firewalls). A variety of game-theoretic analyses, using a variety of solution concepts, have been conducted – the aims of which are to inform us about the set of configurations in which the model might be likely to end up after some time. Due to the success of these analyses at determining various equilibrium configurations, we occasionally find it useful to ground our thinking by considering the initial parameter configurations of our model as equilibrium outcomes of some game played by node operators. However, our work is not dependent on this interpretation, or any of its details. In this paper, we focus exclusively on the probability distribution over the number of compromised nodes in an outcome of the model, using the entire parameter space and without making any additional assumptions about how the various values of parameters got to where they are.

3.4. Loss Distribution

A loss outcome is an event in which some nodes are compromised and others are not. To completely specify a loss outcome requires listing the set of compromised nodes. So a complete distribution on loss outcomes is a probability distribution on all subsets of nodes. This distribution is not tractable to analyze since the number of subsets of nodes is exponential in the number of nodes. However, if we consider only the *number* of compromised nodes, then its distribution is tractable to analyze. Moreover, the information obtained from studying this distribution remains highly relevant to network security and insurability. Let NL be the random variable that counts the number of compromised nodes in a loss outcome. Then, the *loss-number distribution* is a set of $N + 1$ numbers giving $\Pr[NL = k]$ for $k = 0, \ldots, N$.

4. COMPUTABILITY OF THE LOSS DISTRIBUTION

Notational Conventions

Whenever necessary for convenience throughout this paper, we adopt the following common mathematical conventions:

$$0^0 = 1, \qquad \sum_{\varnothing} = 0, \qquad \prod_{\varnothing} = 1, \qquad \binom{n}{m} = 0 \quad \text{whenever } m < 0 \text{ or } m > n.$$

4.1. General Formula for the One-Hop Model

We start by giving a general formula for the $N+1$ terms of the loss distribution on NL for the one-hop model.

LEMMA 4.1. *For each* $k = 0, \ldots, N$,

$$\Pr[NL{=}k] = \sum_{\substack{C,D:\\ C\subseteq\{1,\ldots,N\},\\ D\subseteq C,\ |C|=k}} \left[\prod_{i\in D} p_i \prod_{i\in C\setminus D} \Big((1{-}p_i)(1{-}\prod_{j\in D}(1{-}q_{ji}))\Big) \cdot \prod_{i\notin C} \Big((1-p_i)\prod_{j\in D}(1-q_{ji})\Big) \right].$$

The proof of Lemma 4.1 can be found in Appendix B.2.

Notice that the number of terms in each of the lemma's formulas is exponential in the number of nodes N. Consequently, the running time of a straightforward algorithm computing the value of the formula is also exponential. Even for relatively small networks, the number of terms can be considerably large; for example, the number of 150-element-subsets of the set $\{1, \ldots, 300\}$ is approximately 10^{88}, which is greater than the number of atoms in the observable universe. In practice, this prevents us from using the above formula for large networks.

4.2. NP-Hardness of Computing the Loss Distribution

The question naturally arises: is this exponential running time a defect of our formulation or an inherent property of the problem? Here we show that, unfortunately, the general problem of computing the loss-number distribution is indeed NP-hard in both the one-hop and multiple-hop models. Thus, assuming that P $\neq$ NP[2], no polynomial-time algorithm can exist that computes the exact value of $\Pr[NL = k]$ for each k. However, in the subsequent subsections, we also show that the distribution can be computed efficiently for certain classes of networks in the one-hop model.

Our hardness proof is based on reduction from a well-known NP-complete problem, the Minimum Set Cover problem. To perform the reduction, we first define a decision problem that can easily be reduced to computing the distribution of NL.

Definition 4.2. **Loss Probability:** Given an integer N, probabilities p_i, q_{ij} for $i, j = 1, \ldots N$, an integer k, and a real number δ, does the network of N nodes having direct compromise probabilities p_i and indirect risk probabilities q_{ij} satisfy $\Pr[NL \geq k] \geq \delta$?

THEOREM 4.3. *In both the one-hop and the multiple-hop models, the Set Cover problem can be reduced to the Loss Probability problem in polynomial time.*

The proof of Theorem 4.3 can be found in Appendix B.1.

[2] P $\neq$ NP is a widely accepted conjecture; if it were not true, we would be able to solve all NP-hard problems in polynomial time.

4.3. Special Case Topologies in the One-Hop Model

Since the problem of computing the exact distribution is NP-hard, we have two viable options for large networks. First, we can focus on restricted classes of networks. In the following subsubsections, we give efficient formulas for three such classes in the one-hop model. A second option is to use simulations to approximate the general case. We follow this second approach in Section 4.4.

4.3.1. Homogeneous Topologies. For a homogeneous network, the topology of the network is a complete graph; each node has a direct compromise probability of p, and each edge has a propagation probability of q (in both directions). Such topologies arise in practice whenever the network is fully connected. See Figure 10a in Appendix A for an illustration.

LEMMA 4.4. *The probability of k nodes being compromised in a homogeneous network is*

$$\Pr[NL = k] = \binom{N}{k} \sum_{d=0}^{k} \left[\binom{k}{d} p^d (1-p)^{(N-d)} \cdot \left(1 - (1-q)^d\right)^{k-d} \cdot \left((1-q)^d\right)^{(N-k)} \right] .$$

An alternative formulation derived using the binomial distribution is

$$\Pr[NL = k] = \binom{N}{k} \left[1 - (1-p)(1-pq)^{N-1}\right]^k \cdot \left[(1-p)(1-pq)^{N-1}\right]^{N-k} .$$

The proof of Lemma 4.4 can be found in Appendix B.3.

4.3.2. Star Topologies. A star graph is a tree with one internal node and $N-1$ outer nodes. See Figure 10b for an illustration. We let p_0 denote the direct compromise probability of the internal node, and assume that the outer nodes have a uniform direct compromise probability, denoted by p_1. Furthermore, we assume that the probability of propagation is uniform from the internal node to the outer nodes, denoted by q_{out}, and from the outer nodes to the internal node, denoted by q_{in}.

This can model, for example, a network that consists of a single server and $N-1$ clients. We can assume that each client communicates directly only with the server; e.g., there are strict firewalls or no physical connections between the clients. Hence, there is no propagation between the clients.

LEMMA 4.5. *The probability of k nodes being compromised in the star network is*

$$\begin{aligned}\Pr[NL = k] = &\binom{N-1}{k-1} \sum_{d=0}^{k-1} \left[\binom{k-1}{d} \cdot p_0 p_1^d (1-p_1)^{N-1-d} \cdot q_{out}^{(k-1)-d} \cdot (1-q_{out})^{N-k} \right] \\ &+ \binom{N-1}{k-1} p_1^{k-1} (1-p_0)(1-p_1)^{N-k} \cdot \left(1 - (1-q_{in})^{k-1}\right) \\ &+ \binom{N-1}{k} p_1^k (1-p_0)(1-p_1)^{N-1-k} \cdot (1-q_{in})^k .\end{aligned}$$

The proof of Lemma 4.5 can be found in Appendix B.4.

4.3.3. E-R Random Topologies. In the Erdős-Rényi (E-R) random graph model, undirected edges are set between each pair of nodes with equal probability ρ, independently of other edges [Erdős and Rényi 1960].

Assume that the propagation probability of every edge is q. Then, the probability that a directly compromised node i propagates the compromise to any given node j is

$$\Pr[i \text{ and } j \text{ are connected}] \cdot q = \rho q . \quad (1)$$

Consequently, the probability of any given node i being compromised in an E-R random graph with a propagation probability of q and an edge probability of ρ is equal to the probability of i being compromised in a homogeneous network with a propagation probability of ρq. Therefore, the distribution of NL is the same for a random network with parameters p, q, and ρ and for a homogeneous network with parameters p and ρq. See Figure 10c for an illustration.

COROLLARY 4.6. *The probability of k nodes being compromised in an E-R random network is*

$$\Pr[NL = k] = \binom{N}{k} \sum_{d=0}^{k} \left[\quad \binom{k}{d} p^d (1-p)^{(N-d)} \cdot \left(1 - (1-\rho q)^d\right)^{k-d} \cdot \left((1-\rho q)^d\right)^{(N-k)} \right].$$

PROOF. It follows immediately from Lemma 4.4 and the structure of the E-R random network. □

Notice that for each of these network topologies, the number of terms in the formula giving the distribution on the number of losses is at most quadratic in N. Thus, we can compute the distribution efficiently for networks with these topologies.

4.4. Simulation

4.4.1. One-Hop Model. For more general network topologies, we use simulation to obtain an approximate distribution. The simulation computes an empirical distribution by repeatedly choosing outcomes that result from a simulated attack following the direct compromise and propagation probabilities, as follows:

— In each iteration, choose an outcome randomly in the following way:
  — First, for each node i, decide whether node i is directly compromised (or not) at random according to p_i.
  — Second, for each directly compromised node i, iterate over the set of its non-compromised neighbors. For each non-compromised neighbor j, decide whether node i propagates compromise to node j (or not) at random according to q_{ij}.
— Count the nodes that have been compromised and add 1 to the number of occurrences of this outcome.
— After a fixed number of iterations, terminate the simulation and, for each outcome, output the number of occurrences over the number of iterations as the empirical probability of that outcome.

The running time of the above algorithm is polynomial in the size of the network, given a constant number of iterations. Furthermore, we have from the strong law of large numbers that the empirical distribution function converges to the actual function almost surely. To show that convergence is fast enough in practice, we ran simulations for a) randomly generated scale-free networks and b) special case topologies.

First, we randomly generated scale-free networks using the Barabási-Albert (BA) model [Barabási and Albert 1999], and ran the simulation with varying numbers of iterations. In each case, the shape of the distribution settled down to a smooth form within a few tens of thousands of iterations. Figure 1 shows a series of simulated distributions using a single network and with varying numbers of iterations. The network was generated with parameters $N = 600$, $m_0 = 15$, and $m = 4$, and the simulations were performed with parameters $p = 0.01$ and $q = 0.1$. (See Section 5 for a brief explanation of how BA graphs are generated.) As can be seen from the figure, once the number of iterations is sufficiently high, the empirical distribution reaches a fixed state.

Second, we ran the simulations for homogeneous and star graph networks and compared the approximate distributions to the exact ones. Two of these results are pre-

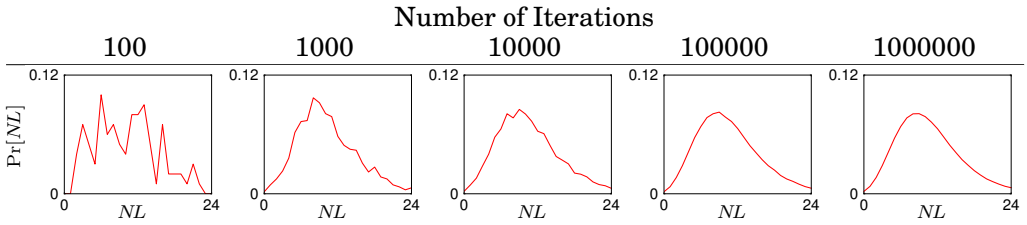
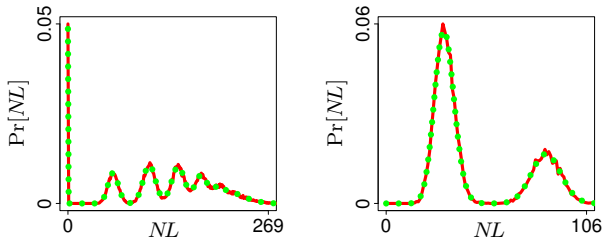


Fig. 1: Distributions obtained from simulations with various numbers of iterations.

sented in Figure 2. The homogeneous network in that figure consists of 300 nodes with $p = 0.01$ and $q = 0.2$. We ran the simulation for 50,000 iterations for this network. The star network in the adjacent sub-figure consists of 300 nodes with $p_0 = 0.3$, $p_1 = 0.1$, and $q_{in} = q_{out} = 0.2$, and we ran this simulation for 20,000 iterations. As can be seen from these plots, the distributions obtained from the simulations, which consisted of only relatively small numbers of iterations, are very good approximations to the exact distributions.



(a) Homogeneous (b) Star

Fig. 2: Comparison of distributions obtained using simulation (solid red) to the exact distributions obtained from the formulas (dotted green).

Notice that these distributions have multiple local maxima, which distinguish them substantially from the common bell shape of a normal or a binomial distribution. To explain this phenomenon in the homogeneous network, the global maximum at the very beginning represents the event in which no nodes are directly compromised, while each consecutive local maximum primarily contains events in which one additional node is directly compromised. In the star network, the first maximum primarily contains events in which the center node is *not* compromised, and the second maximum consists primarily of events in which the center node is compromised.

4.4.2. Multiple-Hop Model. We use a similar simulation to obtain approximate distributions for the multiple-hop model, by choosing outcomes randomly in the following way:

— First, for each link (i, j), decide whether link (i, j) is present (or not) at random according to q_{ij}.
— Second, for each node i, decide whether node i is directly compromised (or not) at random according to p_i. If node i is directly compromised, then mark all nodes that can be reached on a directed path from node i as compromised.

Note that, if we use a breadth- or depth-first search which enqueues only uncompromised nodes in the second step, the running time of choosing a random outcome is linear in the size of the network. Finally, we evaluated the rate of convergence in

practice for the multiple-hop model in the same way as we did for the one-hop model in Figure 1. The results we obtained from this evaluation were very similar to the results from the one-hop model; we omit their detailed discussion here due to space limitations.

5. SYSTEMATIC RISK IN SCALE-FREE NETWORKS

To study how systematic risk is affected by the network topology, we ran a large number of simulations on scale-free networks. The networks were generated according to one of the most prevalent models, the Barabási-Albert (BA) model [Barabási and Albert 1999]. The BA model is based on the concept of preferential attachment, meaning that the more connected a node is, the more likely it is to receive new connections.

The BA model generates scale-free graphs as follows. First, a clique of m_0 initial nodes is constructed. Then, the remaining $N - m_0$ nodes are added to the network one by one. Each new node is randomly connected to $m < m_0$ existing nodes with probabilities proportional to the degrees of the existing nodes.

5.1. Measuring Systematic Risk

We begin this section with two prerequisite definitions involving insurance concepts. For a given probability r and random variable X, the quantile function $Q(r)$ specifies the lowest value k such that $\Pr[X \leq k] = r$. For example, in the case of the loss-number distribution arising from an instance of our model, the 99% quantile gives the maximum number of compromised nodes that an insurer can expect in an outcome of the model 99% of the time. In a similar veign, the *safety loading* at probability r is the excess premium, above and beyond an unbiased premium, that would be required to ensure that the *probability of having to pay out more in total damages than what is collected in total premiums* is at most r. For example, if the variability of the loss-number distribution is high, then the probability of an unusually-large number of nodes being compromised may be non-negligible. In that case, for the insurance market to be viable, the safety loading and – hence – the premiums, have to be relatively high. See Section 5.3 for an example, and [Böhme 2005] for a more detailed definition.

In the remainder of this subsection, we compute the mean, the variance, the 99.9% quantile, and the safety loading requirement at probability 99.9% for the loss distributions of several scale-free networks. We compare these quantities to those of *binomial distributions with the same mean*. Binomial distributions are of special interest to us because our goal is to measure the systematic risk caused by the interdependence between the nodes. If there were no interdependence (i.e., if the nodes were not connected by a network), then the individual node compromise events would be independent of each other. In this case, assuming that all the direct compromise probabilities are the same, the loss distribution would be a simple binomial distribution by definition. Consequently, the binomial distribution is the baseline to which the actual loss distributions should be compared. Note that we use binomial distributions with the same mean for the comparison because we are interested in riskiness (i.e., the variability of the loss) not in the expected loss.

5.1.1. One-Hop Model. Figure 3 and Table II (see Appendix C) compare binomial distributions to the actual loss distributions resulting from various direct compromise and propagation probabilities in the one-hop model. The network consists of $N = 600$ nodes, and it was generated using the parameters $m_0 = 15$ and $m = 4$.

Even though the shapes of the actual loss distributions are in many cases similar to those of the binomial distributions, there is always a substantial difference in variability; for example, the variance of the actual distribution is always almost twice as high as that of the binomial. As expected, increasing the propagation probability

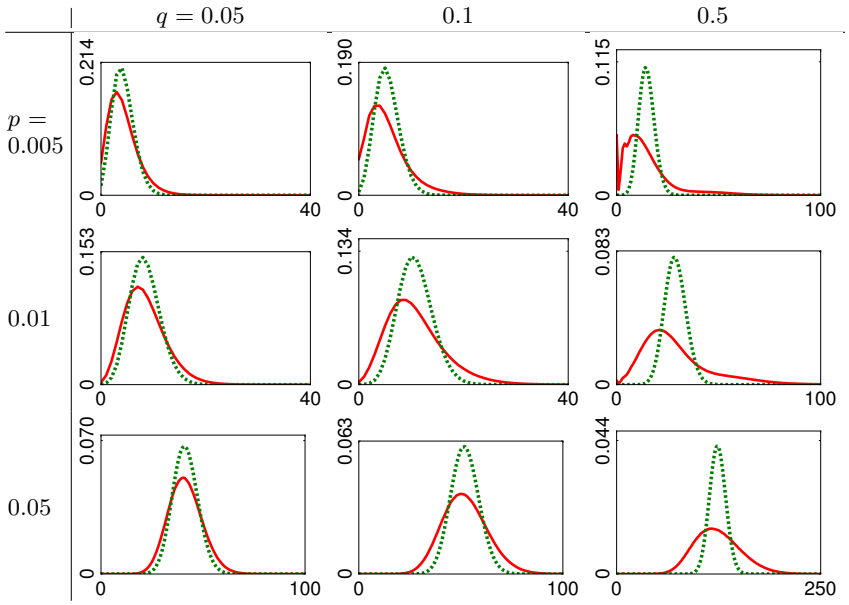


Fig. 3: Comparison of the actual loss distribution in the one-hop model (solid red) to the binomial distribution (dotted green) for various direct compromise and propagation probabilities, and constant network size $N = 600$. (Note that the slightly irregular subfigure for $p = 0.005$ and $q = 0.5$ is correctly drawn.)

increases the difference between the actual loss distribution and the binomial distribution through increasing the interdependence between the nodes of the network. Increasing the direct node compromise probability has a less pronounced effect in the same direction. Again, this is unsurprising, since correlations are caused by interdependence, which is not affected by direct node compromise probabilities in our model.

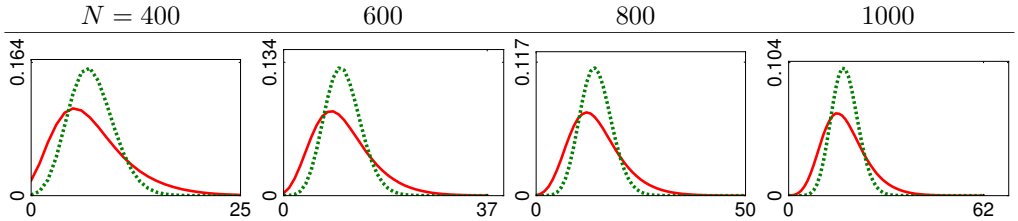


Fig. 4: Comparison of the actual loss distribution in the one-hop model (solid red) to the binomial distribution (dotted green) for various sizes and constant $p = 0.01$, $q = 0.1$.

Figure 4 and Table III (see Appendix C) compare binomial distributions to the actual loss distributions for various network sizes in the one-hop model. The direct compromise and propagation probabilities are $p = 0.01$ and $q = 0.1$, and the networks were generated using the same parameters $m_0 = 15$ and $m = 4$. As can be seen, the difference between the actual loss distribution and the binomial distribution does not diminish as the size of the network increases (in fact, the ratio between the variances slightly increases from 2.76 to 2.82). This observation is very important, since insurance is based on the idea of diversifiability, which means that individual risks cancel out as the number of individuals increases. Since we would see a binomial distribution

if all risks were independent, the fact that the difference does not diminish indicates that these risks are not diversifiable, and highlights the importance of knowing the actual loss distribution.

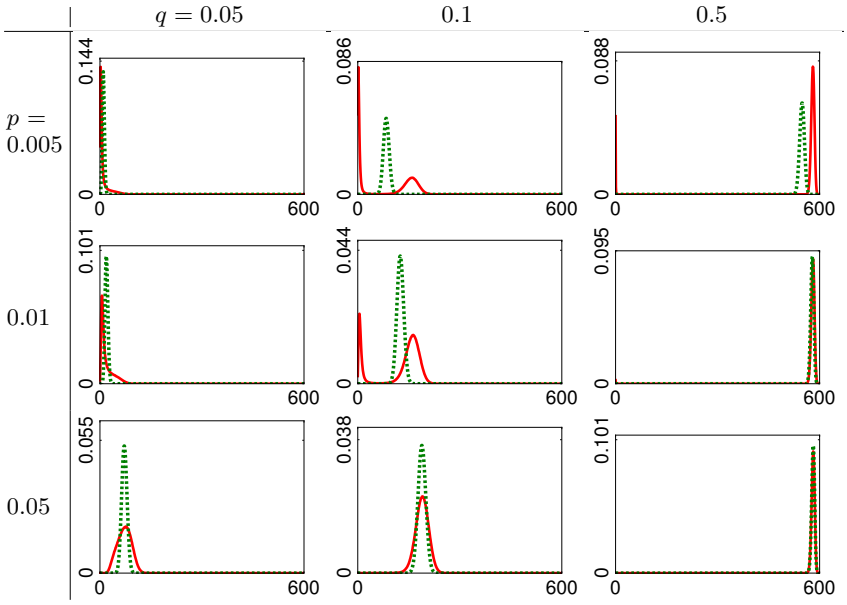


Fig. 5: Comparison of the actual loss distribution in the multiple-hop model (solid red) to the binomial distribution (dotted green) for various direct compromise and propagation probabilities, and constant network size $N = 600$. (Observe that the probability of zero nodes being compromised is non-zero for $q = 0.5$ and $p = 0.005$ or 0.01.)

5.1.2. Multiple-Hop Model. Figure 5 and Table IV (see Appendix C) compare binomial distributions to the actual loss distributions resulting from various direct compromise and propagation probabilities in the multiple-hop model. The network was generated in the same manner as for the one-hop model.

Contrary to the one-hop model, the behavior of the multiple-hop model can vary substantially with different parameter values. Before we discuss specific cases, consider the epidemic threshold for the propagation probability, that is, the minimum probability that is sufficient for the compromise to spread to a large portion of the network. Since the average degree is around $2 \cdot m = 8$, the threshold propagation probability is around $\frac{1}{8}$ for a random graph (otherwise, each new compromise would lead to less than one other compromise on average). However, scale-free networks have high-degree "hub" nodes, which are very likely to get compromised due to propagation from one of their many neighbors. Once compromised, "hubs" can propagate the compromise to large numbers of other nodes since they have many neighbors. Hence, compromise spreads to a large portion of the network even at lower propagation probabilities.

In the first column ($q = 0.05$) of Figure 5, we see that the propagation probability q is so extremely low that the compromise does not spread (note that the fairly high expected value for $p = 0.05$ is mostly due to direct compromises). Nonetheless, the difference between the actual and the binomial distributions is higher than in the one-hop model (see Table IV). Next, in the second column ($q = 0.1$), we see that the propagation probability q is around the epidemic threshold. More specifically, we see

that not spreading (first local maxima, around zero) and spreading to a large portion of the network (second local maxima) are both likely, unless the number of directly compromised nodes is extremely high ($p = 0.05$). Finally, in the third column ($q = 0.5$), we see that the propagation probability q is so high that the compromise spreads to the whole network almost certainly. Note that spreading is only "almost certain", since the chance of no direct compromises occurring can be non-negligible (see the tall thin line at zero for $p = 0.005$ and the short thin line for $p = 0.01$).

Some of these phenomena have already been studied in the context of percolation theory and epidemic thresholds in scale-free graphs (e.g., [Pastor-Satorras and Vespignani 2001]); however, we are rather interested in their implications for risk-mitigation and cyber-insurance. First, for lower propagation probabilities ($q = 0.05$), the results are similar to those in the case of the one-hop model: the actual distributions are long-tailed compared to binomial distributions, which means that interdependence causes systematic risk. Second, for propagation probabilities around the threshold ($q = 0.1$), the variability of the actual loss distributions is extremely high. Hence, an insurance provider has to ask for very high premiums, which the individual nodes will find unfair. Finally, for higher propagation probabilities ($q = 0.5$), the difference between the actual and the binomial distributions is relatively small (see, e.g., $p = 0.05$). However, providing insurance in this case is not viable, since every node is compromised almost certainly (no one would provide insurance for certain events).[3]

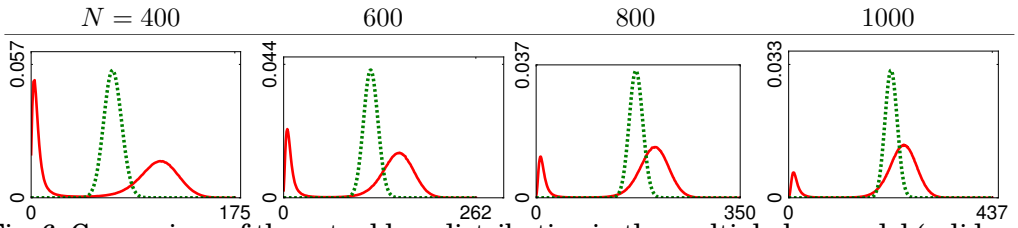


Fig. 6: Comparison of the actual loss distribution in the multiple-hop model (solid red) to the binomial distribution (dotted green) for various network sizes and constant $p = 0.01$, $q = 0.1$.

Figure 6 and Table V (see Appendix C) compare binomial distributions to the actual loss distributions in the multiple-hop model for various network sizes. Again, the networks were generated in the same manner as for the one-hop model, and the same parameters were used. As can be seen, the difference between the actual loss distribution and the binomial distribution is substantial (e.g., the variance of the actual loss distribution is an order of magnitude higher), which again shows high systematic risk.

5.2. Sampling Scale-Free Graphs

In the previous subsection, we illustrated the extent to which systematic risk is present in scale-free networks. In this subsection, to study whether this systematic risk can be estimated from smaller samples (e.g., using incident reports from a subset of the nodes), we investigate the loss distributions of random samples of scale-free networks. The network from which the samples are drawn is a scale-free network with parameters $N = 600$, $m_0 = 15$, and $m = 4$, and with probabilities $p = 0.05$ and $q = 0.1$.

[3]More specifically, if an insurer did provide insurance, then the premiums would have to be almost as high as the potential loss values, so that honest individuals would not purchase it.

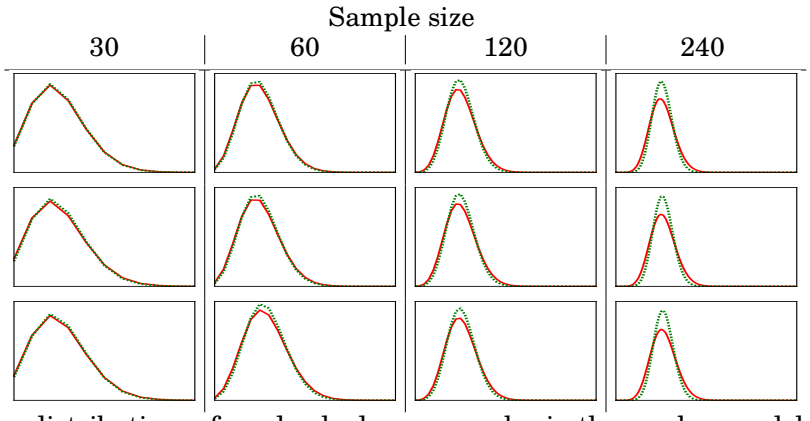


Fig. 7: Loss distributions of randomly drawn samples in the one-hop model (solid red) compared to binomial distributions (dotted green). For each size, three randomly chosen samples are used for the comparison. Parameters are $N = 600$, $p = 0.05$, $q = 0.1$.

5.2.1. One-Hop Model. Figure 7 and Table VI (see Appendix C) compare the actual loss distributions of randomly drawn samples to binomial distributions in the one-hop model. We study four different sample sizes: 30, 60, 120, and 240 nodes. For each sample size, three samples of the given size were drawn uniformly at random from the set of all nodes. For each sample, its loss distribution was computed by running the simulation for the *entire network*, but counting only the compromised nodes belonging to the sample. This models the real-world scenario where incident reports are collected from only a sample, but the security of this sample is affected by the rest of the world through external connections. As before, the binomial distributions to which the samples are compared have exactly the same expected loss $E[NL]$ as the corresponding actual sample distributions.

Figure 7 shows the three random samples for each size, together with the corresponding binomial distributions, while Table VI gives a more detailed comparison in terms of the metrics we are considering. The figure shows that the loss distributions of the samples are almost indistinguishable from binomial distributions for sample sizes of 30 and 60 nodes. Consequently, by observing only a sample of the entire network, one might arrive at the wrong conclusion that individual node compromises are independent events. As the sizes of the samples increase, the loss distributions become more distinguishable from the binomial distribution, eventually approaching the distribution of losses for the full network. This phenomenon can be explained by considering the probability of two nodes sharing a neighbor, which could cause correlation between them, or two nodes being connected. In smaller samples, this probability is negligible, which means that individual risks are almost always independent; hence, the loss follows a binomial distribution.

5.2.2. Multiple-Hop Model. Figure 8 and Table VII (see Appendix C) compare the actual loss distributions of randomly drawn samples to binomial distributions in the multiple-hop model. The samples were drawn and their distributions were computed in the same way as for the one-hop model. Similarly to the one-hop model, we see that for smaller sample sizes, the actual loss distributions can be almost indistinguishable from binomial distributions. Again, this means that by observing only a smaller sample, one can easily underestimate the systematic risk of the complete network.

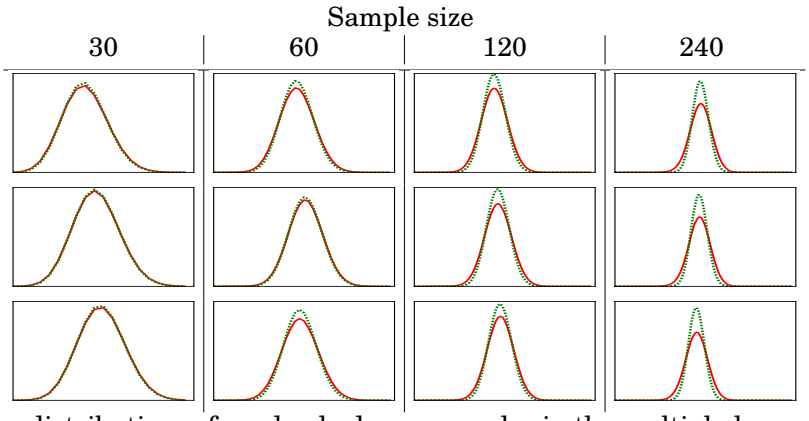


Fig. 8: Loss distributions of randomly drawn samples in the multiple-hop model (solid red) compared to binomial distributions (dotted green). For each size, three randomly chosen samples are used for the comparison. Parameters are $N = 600$, $p = 0.05$, $q = 0.1$.

5.3. Application to Cyber-Insurance

As a motivating example, consider an insurer who provides insurance coverage to individuals forming a network with parameters $N = 600$, $p = 0.01$, and $q = 0.1$, against threats which can be modeled using the one-hop model. Suppose that the insurer uses a smaller sample of incident reports to estimate the risk associated with these insurance policies. Since even small samples are unbiased estimators of the average probability that a given node is compromised[4], it can correctly estimate the average risk as $E[NL]/N = 1.80\%$ based on the individual incident reports. In order to keep its probability of ruin (i.e., the probability that the loss exceeds the premiums) below a given level 0.001, the insurer wants to compute the necessary amount of premiums to be collected as $Q(NL, 0.999)$. In other words, it wants to compute the safety loading $Q(NL, 0.999) - E[NL]$ using the quantile premium principle. Since the insurer observes that risks are very close to independent (i.e., there is no systemic risk), it estimates the necessary safety loading based on a binomial distribution [Böhme 2005], which gives a value of 11.2. However, its safety loading should be in fact 24.2 (see Table II). This mistake has rather harsh consequences for the insurance provider: the probability that the total loss exceeds the erroneously calculated insurance premium is $\Pr[NL > E[NL] + 11.2] = 3.1\%$. In other words, the probability of ruin for the miscalculated premiums is 3.1% instead of the anticipated 0.1%.

6. CONCLUSIONS

Cybersecurity is not exclusively a technological problem. A 2012 article by prominent computer scientists in the Proceedings of the IEEE titled “Privacy and Cybersecurity: The Next 100 Years” summarized that our ability to manage security partly remains haphazard since we are still lacking methods to appropriately assess either the cost or value of security [Landwehr et al. 2012]. With more emphasis, leading cybercrime researchers and security economists added that we remain “*extremely* inefficient at fighting cybercrime” [Anderson et al. 2013].

From an economic perspective, there are at least two high-level challenges to address this troubling status quo. On the one hand, we have to address the shortage of

[4]Table VI shows that the ratio $\frac{E[NL]}{N}$ is independent of the sample size, which means that even small samples are unbiased estimators of average risk. Note that, since Table VI is based on different parameter values, the ratio is approximately 8.7% there.

reliable data about security investments and incidents. On the other hand, we need to overcome the dearth of appropriate models to understand cybersecurity. A particularly important task identified by a committee of cybersecurity experts is to adjust existing risk management approaches so that they address the specific characteristics of cyber-risk [Chong et al. 2009]. The objective is not to find the silver bullet to eliminate all cybersecurity incidents, which would be unrealistic given the existing constraints and complexities. However, progress is needed so that eventually "crime does not pay" [Chong et al. 2009].

Unfortunately, up to this point, many specific risk management approaches remain underdeveloped. In particular, cyber-insurance, which has been identified as one of the most promising approaches for incentivizing the adoption of security best practices and efficient levels of investment in cybersecurity, is suffering from a lack of good methods for reliable risk pricing [Chong et al. 2009]. However, this is needed for the calculation of insurance premiums, internal company decision-making, and appropriate public policy measures.

In our work, we took several concrete steps to improve our understanding of risk pricing in networked systems with a particular focus on systematic risk, which measures vulnerability to catastrophic cyber-incidents that damage a vast number of nodes at the same time. Systematic risk is a consequence of the complexity of the networked system itself. In particular, it depends jointly on the topology of the network and the security contributions of individual nodes. In our work, we study the properties of systematic risk with a general model of one-hop and multiple-hop risk propagation in a networked system with particular emphasis on scale-free topologies.

Our results include theoretical contributions as well as insights driven by robust simulation analyses. We found that the distribution of the number of compromised nodes has a number of interesting properties. For the one-hop model, it is expressible as a simple closed formula; and it is efficiently computable for several interesting special cases. In general, it is NP-hard to compute; and it can be efficiently approximated using simulation for arbitrary topologies.

By applying our methodology to scale-free networks, we found that the full network possesses systematic risks, which may require large amounts of safety capital to properly insure. Yet, we found much lower systematic risk in random samples of the same networks. This observation yields two contrasting applications to cyber-insurance. On the one hand, it may be possible to insure random subsamples of a network with scale-free properties while bearing only a modest loading cost. On the other hand, an insurer cannot readily deduce the systematic risk of a network by taking random samples.

Moving forward, our research agenda contributes to a more principled understanding and management of systematic risks in networked systems. In particular, for cyber-insurance to be viable, insurers need not only to be able to assess the security practices of their customers, but also consider the topology of physical and social connections. Partly, our results are encouraging; still, further progress is needed to derive more precise figures for the occurrence of systematic risks. With our work, we move one step forward towards the emergence of an attractive and robust cyber-insurance market.

Acknowledgment

We thank the reviewers of our manuscript and of our previous conference contributions for their detailed comments and suggestions that helped to improve our work. The research activities of Jens Grossklags and Benjamin Johnson are supported by the German Institute for Trust and Safety on the Internet (DIVSI).

REFERENCES

Ross Anderson. 1994. Liability and Computer Security: Nine Principles. In *Proceedings of the Third European Symposium on Research in Computer Security (ESORICS)*. 231–245.

Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. 2013. Measuring the Cost of Cybercrime. In *The Economics of Information Security and Privacy*, Rainer Böhme (Ed.). Springer Berlin Heidelberg, 265–300.

James Aspnes, Kevin Chang, and Aleksandr Yampolskiy. 2006. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. System Sci.* 72, 6 (Sept. 2006), 1077–1093.

Albert-László Barabási. 2009. Scale-Free Networks: A Decade and Beyond. *Science* 325, 5939 (July 2009), 412–413.

Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286, 5439 (Oct. 1999), 509–512.

Andrew Betts. 2013. A sobering day. Financial Times Labs, http://labs.ft.com/2013/05/a-sobering-day/. (May 29, 2013).

Kenneth Birman and Fred Schneider. 2009. The Monoculture Risk Put into Context. *IEEE Security and Privacy* 7, 1 (Jan. 2009), 14–17.

Rainer Böhme. 2005. Cyber-insurance revisited. In *Workshop on the Economics of Information Security*.

Rainer Böhme. 2010. Towards Insurable Network Architectures. *it - Information Technology* 52, 5 (Sept. 2010), 290–293.

Rainer Böhme and Gaurav Kataria. 2006. Models and Measures for Correlation in Cyber-Insurance. In *Workshop on the Economics of Information Security*.

Rainer Böhme and Galina Schwartz. 2010. Modeling cyber-insurance: Towards a unifying framework. In *Workshop on the Economics of Information Security*.

Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. 2008. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security* 10, 4 (2008), 1.

Hau Chan, Michael Ceyko, and Luis Ortiz. 2012. Interdependent Defense Games: Modeling Interdependent Security under Deliberate Attacks. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI)*. 152–162.

Pei-Yu Chen, Gaurav Kataria, and Ramayya Krishnan. 2011. Correlated failures, diversification, and information security risk management. *MIS Quarterly* 35, 2 (June 2011), 397–422.

Fred Chong, Ruby Lee, Claire Vishik, Alessandro Acquisti, William Horne, Charles Palmer, Anup Ghosh, Dimitrios Pendarakis, William Sanders, Eric Fleischman, Hugo Teufel, Gene Tsudik, Dipankar Dasgupta, Steven Hofmeyr, and Leor Weinberger. 2009. National Cyber Leap Year Summit 2009: Co-Chairs' Report. (Sept. 2009). Available at: https://www.qinetiq-na.com/wp-content/uploads/2011/12/National_Cyber_Leap_Year_Summit_2009_CoChairs_Report.pdf.

Sudarshan Dhall, Sivaramakrishnan Lakshmivarahan, and Pramode Verma. 2009. On the number and the distribution of the Nash equilibria in supermodular games and their impact on the tipping set. In *Proceedings of the Int. Conference on Game Theory for Networks (GameNets)*. 691–696.

Christopher Drew. 2011. Stolen Data Is Tracked to Hacking at Lockheed. New York Times, http://www.nytimes.com/2011/06/04/technology/04security.html. (June 3, 2011).

Victor Eguiluz and Konstantin Klemm. 2002. Epidemic threshold in structured scale-free networks. *Physical Review Letters* 89, 10 (Aug. 2002), 108701.

Paul Erdős and Alfréd Rényi. 1959. On random graphs. *Publicationes Mathematicae (Debrecen)* 6 (1959), 290–297.

Paul Erdős and Alfréd Rényi. 1960. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960), 17–61.

Ayalvadi Ganesh, Laurent Massoulié, and Don Towsley. 2005. The effect of network topology on the spread of epidemics. In *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*. 1455–1466.

Daniel Geer, Charles Pfleeger, Bruce Schneier, John Quarterman, Perry Metzger, Rebecca Bace, and Peter Gutmann. 2003. CyberInsecurity: The Cost of Monopoly. How the Dominance of Microsoft's Products Poses a Risk to Society. (2003).

Jens Grossklags, Nicolas Christin, and John Chuang. 2008. Secure or insure?: A game-theoretic analysis of information security games. In *Proc. of the 17th International World Wide Web Conference*. 209–218.

Geoffrey Heal and Howard Kunreuther. 2004. *Interdependent security: A general model*. NBER Working Paper No. 10706.

Benjamin Johnson, Rainer Böhme, and Jens Grossklags. 2011. Security games with market insurance. *Decision and Game Theory for Security* (2011), 117–130.

Benjamin Johnson, Jens Grossklags, Nicolas Christin, and John Chuang. 2010. Uncertainty in interdependent security games. *Decision and Game Theory for Security* (2010), 234–244.

Benjamin Johnson, Aron Laszka, and Jens Grossklags. 2014a. The Complexity of Estimating Systematic Risk in Networks. In *Proceedings of the 27th IEEE Computer Security Foundations Symposium (CSF)*. 325–336.

Benjamin Johnson, Aron Laszka, and Jens Grossklags. 2014b. How Many Down? Toward Understanding Systematic Risk in Networks. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*. 495–500.

Michael Kearns and Luis Ortiz. 2004. Algorithms for Interdependent Security Games. In *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf (Eds.). MIT Press, 561–568.

Jeffrey Kephart and Steve White. 1991. Directed-graph epidemiological models of computer viruses. In *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*. 343–359.

Jeffrey Kephart and Steve White. 1993. Measuring and modeling computer virus prevalence. In *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*. 2–15.

Howard Kunreuther and Geoffrey Heal. 2003. Interdependent security. *Journal of Risk and Uncertainty* 26, 2 (March 2003), 231–249.

Carl Landwehr, Dan Boneh, John Mitchell, Steven Bellovin, Susan Landau, and Michael Lesk. 2012. Privacy and Cybersecurity: The Next 100 Years. *Proc. IEEE* 100 (May 2012), 1659–1673.

Aron Laszka, Mark Felegyhazi, and Levente Buttyan. 2014a. A Survey of Interdependent Information Security Games. *Comput. Surveys* 47, 2 (August 2014), 23:1–23:38.

Aron Laszka, Benjamin Johnson, Jens Grossklags, and Mark Felegyhazi. 2014b. Estimating Systematic Risk in Real-World Networks. In *Proceedings of the 18th International Conference on Financial Cryptography and Data Security (FC)*. 417–435.

Marc Lelarge. 2009. Economics of malware: Epidemic risks model, network externalities and incentives. In *Proc. of the 47th Annual Allerton Conf. on Communication, Control, and Computing*. IEEE, 1353–1360.

Marc Lelarge and Jean Bolot. 2008a. A local mean field analysis of security investments in networks. In *Proceedings of the 3rd International Workshop on Economics of Networked Systems*. ACM, 25–30.

Marc Lelarge and Jean Bolot. 2008b. Network externalities and the deployment of security features and protocols in the internet. *ACM SIGMETRICS Performance Evaluation Review* 36, 1 (June 2008), 37–48.

Marc Lelarge and Jean Bolot. 2009. Economic Incentives to Increase Security in the Internet: The Case for Insurance. In *Proceedings of the 33rd IEEE International Conference on Computer Communications (INFOCOM)*. 1494–1502.

Lun Li, David Alderson, John Doyle, and Walter Willinger. 2005. Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics* 2, 4 (2005), 431–523.

Thomas Moscibroda, Stefan Schmid, and Roger Wattenhofer. 2006. When selfish meets evil: Byzantine players in a virus inoculation game. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*. 35–44.

Hulisi Ogut, Nirup Menon, and Srinivasan Raghunathan. 2005. Cyber insurance and IT security investment: Impact of interdependent risk. In *Workshop on the Economics of Information Security*.

Romualdo Pastor-Satorras and Alessandro Vespignani. 2001. Epidemic spreading in scale-free networks. *Physical Review Letters* 86, 14 (April 2001), 3200–3203.

Romualdo Pastor-Satorras and Alessandro Vespignani. 2002. Epidemic dynamics in finite size scale-free networks. *Physical Review E* 65, 3 (March 2002), 035108.

Michael Stumpf, Carsten Wiuf, and Robert May. 2005. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America* 102, 12 (March 2005), 4221–4224.

Symantec. 2014. Emerging Threat: Dragonfly / Energetic Bear – APT Group. *Symantec Connect*, http://www.symantec.com/connect/blogs/emerging-threat-dragonfly-energetic-bear-apt-group. (June 2014). Accessed: November 23, 2017.

Hal Varian. 2004. System Reliability and Free Riding. In *Economics of Information Security*, J. Camp and S. Lewis (Eds.). Kluwer Academic Publishers, Dordrecht, The Netherlands, 1–15.

Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. 2003. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Proceedings of the 22nd International Symposium on Reliable Distributed Systems (SRDS)*. 25–34.

A. LIST OF SYMBOLS AND ILLUSTRATIONS OF THE RISK MODEL

Table I: List of Symbols

Symbol	Description
N	number of nodes
p	probability of direct compromise (when it is uniform over the nodes)
p_i	probability of node i being directly compromised
q	probability of compromise propagation (when it is uniform over the links)
q_{ij}	probability of compromise propagation from node i to node j (given that node i is directly compromised)
q_{in}	probability of compromise propagation from an outer node to the internal node in star topologies
q_{out}	probability of compromise propagation from the internal node to an outer node in star topologies
NL	random variable measuring the number of compromised nodes
ρ	edge inclusion probability in ER random graph model
m_0	size of the initial clique in the BA random graph model
m	number of connections per additional node in the BA random graph model

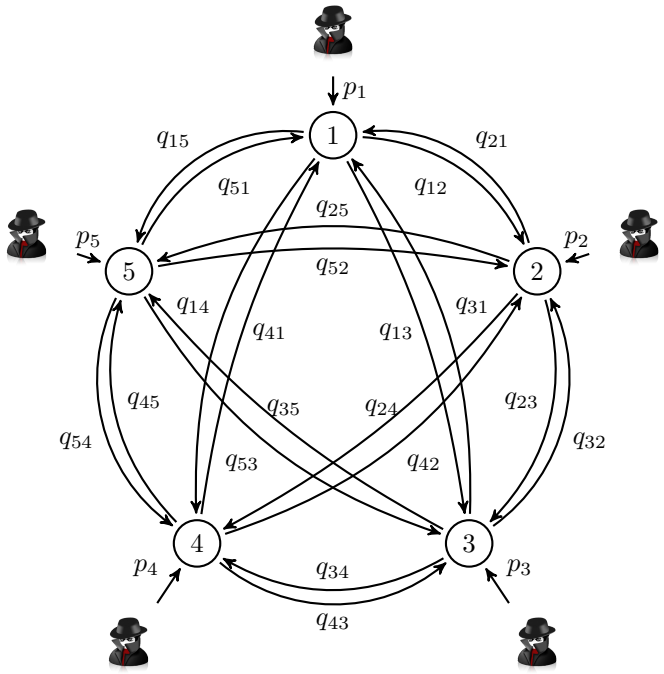
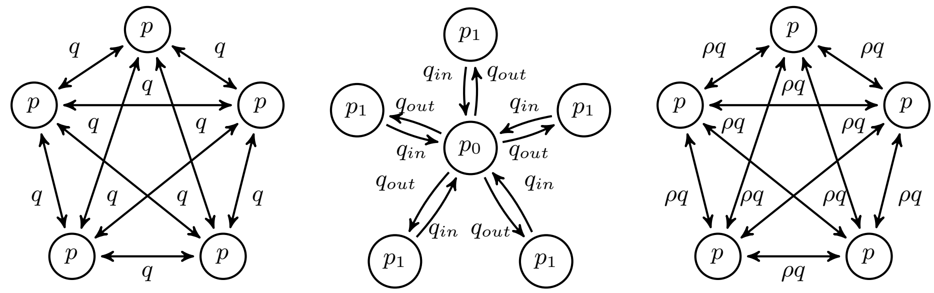


Fig. 9: Network risk arrival and propagation.



(a) Homogeneous network (b) Star network (c) E-R random network

Fig. 10: Special case topologies.

B. PROOFS

B.1. Proof of Theorem 4.3

The Minimum Set Cover problem is defined as follows.

Definition B.1. **Minimum Set Cover:** Given a universe U, a collection $\mathcal{F}$ of subsets of U, and an integer m, is there a collection $\mathcal{C}$ of at most m subsets in $\mathcal{F}$ whose union is U?

Now, we can prove Theorem 4.3.

PROOF. Given an instance $(U, \mathcal{F}, m)$ of the Set Cover problem, we construct an instance $(N, \{p_i\}, \{q_{ij}\}, k, \delta)$ of the Loss Probability problem as follows (see Figure 11 for an illustration).

— Let $N = |\mathcal{F}| + |U| \cdot |\mathcal{F}|$.
— For each subset $S \in \mathcal{F}$, there is a network node, denoted by S, with direct compromise probability $p_S = \frac{1}{|\mathcal{F}|!}$.
— For each element $u \in U$, there are $|\mathcal{F}|$ network nodes, denoted by $u_1, \ldots, u_{|\mathcal{F}|}$, with direct compromise probabilities $p_{u_i} = 0$, $i = 1, \ldots, |\mathcal{F}|$.
— For each pair (u, S) from $U \times \mathcal{F}$ with $u \in S$, there are edges from S to each $u_1, \ldots, u_{|\mathcal{F}|}$ with risk propagation probabilities $q_{S u_i} = 1$, $i = 1, \ldots, |\mathcal{F}|$.
— Let $k = |U| \cdot |\mathcal{F}|$ and $\delta = \frac{1}{|\mathcal{F}|!^m}$.

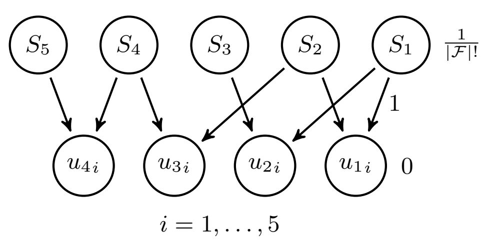


Fig. 11: Illustration for the NP-hardness reduction.

This reduction can be carried out in time and space that is polynomial (more specifically, quartic) in the size of the Minimum Set Cover problem instance. To see this,

observe that the size of the proscribed Loss Probability instance is at most quartic in the size of the Minimum Set Cover instance, since the output is dominated by the $|\mathcal{F}|$ instances of the value $\frac{1}{|\mathcal{F}|!^m}$, which can represented by at most $\log_2 |\mathcal{F}| \cdot |\mathcal{F}| \cdot m \leq |\mathcal{F}|^3$ bits. Furthermore, this value can computed näively in cubic time. It remains to show that, in both the one-hop and the multiple-hop model, $\Pr[NL \geq k] \geq \delta$ if and only if a set cover $\mathcal{C}$ of size at most m exists.

First, observe that in the proscribed network, the one-hop and the multiple-hop models lead to the same loss distribution. To see this, recall that the only difference between the two models is that in the former, indirectly compromised nodes cannot propagate the compromise, while in the latter, they can. In the proscribed network, only the nodes u_i, where $u \in U$ and $i = 1, \ldots, |\mathcal{F}|$, can be indirectly compromised, since only these nodes have incoming edges. However, since these nodes have no outgoing edges, they can never propagate the compromise to other nodes even in the multiple-hop model. Consequently, in the proscribed network, the two models lead to identical loss outcome distributions.

Next, observe that in every outcome for this network, the nodes $u_1, \ldots, u_{|\mathcal{F}|}$ corresponding to the same element $u \in U$ are either all compromised or none of them is. Consequently, if at least $k = |U| \cdot |\mathcal{F}|$ nodes are compromised, then all nodes corresponding to the elements of U are compromised, since any set of $|U| \cdot |\mathcal{F}|$ nodes either contains at least one node for each element of U or it contains all nodes corresponding to the subsets in F.

Now, we show that there exists a collection $\mathcal{C}$ of at most m subsets in $\mathcal{F}$ whose union is U if and only if $\Pr[NL \geq k] \geq \delta$.

For the forward direction, assume that there exists a collection $\mathcal{C}$ of at most m subsets in $\mathcal{F}$ whose union is U. Then, if every node corresponding to a subset in $\mathcal{C}$ is directly compromised, all $|U| \cdot |\mathcal{F}| = k$ nodes corresponding to elements in U get compromised since they are all covered by $\mathcal{C}$. Hence, we have

$$\Pr[NL \geq k] \geq \Pr[\text{every subset in } \mathcal{C} \text{ is directly compromised}] = \left(\frac{1}{|\mathcal{F}|!}\right)^m = \delta \, .$$

Conversely, assume that there does not exist an m-cover of U, so that every collection of sets in $\mathcal{F}$ that covers U has size at least $m + 1$. Then, we have

$$\begin{aligned}
\Pr[NL \geq k] &= \Pr\left[\begin{array}{c}\text{some collection } \mathcal{C} \subseteq \mathcal{F} \text{ that} \\ \text{covers } U \text{ is directly compromised}\end{array}\right] \\
&\leq \Pr\left[\begin{array}{c}\text{some collection } \mathcal{C} \subseteq \mathcal{F} \text{ having at} \\ \text{least } m+1 \text{ subsets is compromised}\end{array}\right] \\
&= \sum_{i=m+1}^{|\mathcal{F}|} \binom{|\mathcal{F}|}{i} \left(\frac{1}{|\mathcal{F}|!}\right)^i < |\mathcal{F}|! \left(\frac{1}{|\mathcal{F}|!}\right)^{m+1} = \left(\frac{1}{|\mathcal{F}|!}\right)^m = \delta \, .
\end{aligned}$$

□

B.2. Proof of Lemma 4.1

PROOF. We compute the probability of the event $NL = k$ by enumerating all events in which k nodes are compromised and summing their probabilities.

Let us first subdivide outcomes meeting the criteria $NL = k$ into disjoint classes according to which nodes were compromised directly, indirectly, or neither. Let C be the set of all compromised nodes, and let D be the set of directly compromised nodes. Then $D \subseteq C$ and, for outcomes in the class specified by this pair (C, D), we know that:

(1) every node in D is directly compromised, and

(2) every node in $C \setminus D$ is not directly compromised but is indirectly compromised by at least one of the nodes in D, and
(3) every node not in C is neither directly compromised nor indirectly compromised by a node in D.

Denoting these events with their numbers, respectively, we have

$$\begin{aligned}
\Pr[1] &= \prod_{i \in D} p_i \\
\Pr[2|1] &= \prod_{i \in C \setminus D} \Big((1-p_i) \Big(1 - \prod_{j \in D} (1-q_{ji}) \Big) \Big) \\
\Pr[3|1] &= \prod_{i \notin C} \Big((1-p_i) \prod_{j \in D} (1-q_{ji}) \Big) \, .
\end{aligned}$$

In any outcome where 1 happens, events 2 and 3 are independent, which implies $\Pr[2 \wedge 3|1] = \Pr[2|1] \cdot \Pr[3|1]$.

The probability of an event in the class (C, D) happening is then

$$\begin{aligned}
\Pr[1 \wedge 2 \wedge 3] &= \Pr[1] \cdot \Pr[2 \wedge 3|1] \\
&= \Pr[1] \cdot \Pr[2|1] \cdot \Pr[3|1] \, . \qquad (2)
\end{aligned}$$

The probability that any event satisfying $NL = k$ happens can now be computed by taking the sum of Equation (2) over all pairs C, D with $D \subseteq C \subseteq \{1, \ldots, N\}$ and $|C| = k$. $\square$

B.3. Proof of Lemma 4.4

PROOF. Suppose that for each node i and j, $p_i = p$ and $q_{ij} = q$. Fix C, D with $D \subseteq C \subseteq \{1, \ldots, N\}$, $|C| = k \leq N$ and $|D| = d \leq k$. Then,

$$\prod_{i \in D} p_i = p^d, \qquad (3)$$

$$\prod_{i \in C \setminus D} \Big((1-p_i) \Big(1 - \prod_{j \in D} (1-q_{ji}) \Big) \Big) = \big((1-p)(1-(1-q)^d) \big)^{k-d} , \qquad (4)$$

and

$$\prod_{i \notin C} \Big((1-p_i) \prod_{j \in D} (1-q_{ji}) \Big) = \big((1-p)(1-q)^d \big)^{N-k} \, . \qquad (5)$$

From Lemma 4.1, $\Pr[NL = k]$ has the form

$$\sum_{d=0}^{k} \sum_{\substack{C,D: \\ C \subseteq \{1,\ldots,N\},\ D \subseteq C,\ |D|=d,\ |C|=k}} \Pr[(C, D)] \, , \qquad (6)$$

where $\Pr[(C, D)]$ is the product of Equations (3), (4) and (5).

The number of pairs (C, D) with $D \subseteq C \subseteq \{1, \ldots, N\}$, $|C| = k$, and $|D| = d$ is exactly $\binom{N}{k} \cdot \binom{k}{d}$; and $\Pr[(C, D)]$ is uniform over all pairs (C, D) satisfying these properties. So

we have

$$\Pr[NL = k] = \sum_{d=0}^{k} \binom{N}{k} \binom{k}{d} \Pr[(C, D)]$$
$$= \binom{N}{k} \sum_{d=0}^{k} \left[\binom{k}{d} p^d (1-p)^{(N-d)} \cdot \left(1 - (1-q)^d\right)^{k-d} \cdot \left((1-q)^d\right)^{(N-k)} \right] .$$

For the alternative formulation, consider that in a homogeneous network topology with parameters p and q, a node is not compromised precisely when it is not directly compromised, and it is neither infected by any of its $N-1$ neighbors. Since these N possibilities are mutually independent, the probability that a node is not compromised may be computed as the product $(1-p) \cdot (1-pq)^{N-1}$. The alternative formula in the lemma now derives directly from the formula for a binomial distribution with parameters N and $(1-p) \cdot (1-pq)^{N-1}$. □

B.4. Proof of Lemma 4.5

PROOF. We divide the set of outcomes into three possibilities. Either

(1) the center node is directly compromised, or
(2) the center node is not directly compromised, but is indirectly compromised, or
(3) the center node is neither directly nor indirectly compromised.

We address each case separately, and then add their probabilities.

(1) In the first case, we further subdivide the space according to the number d of directly-compromised exterior nodes. Fix the total number of compromised nodes k and the number of directly compromised nodes d. In this sub-case we know that $k-d-1$ exterior nodes were not directly but indirectly compromised, and $N-k$ nodes were not compromised at all. The total probability of this case happening is the product of the probabilities that
   (a) the center node is directly compromised
   (b) d exterior nodes are directly compromised
   (c) $k-d-1$ exterior nodes are not directly compromised but are indirectly compromised
   (d) $N-k$ exterior nodes are neither directly nor indirectly compromised

   which gives

$$p_0 p_1^d ((1-p_1) q_{out})^{k-d-1} ((1-p_1)(1-q_{out}))^{N-k}$$
$$= p_0 p_1^d (1-p_1)^{N-1-d} \cdot q_{out}^{(k-1)-d} \cdot (1-q_{out})^{N-k} .$$

   The number of ways to choose d and k in this case is $\binom{N-1}{k-1} \cdot \binom{k-1}{d}$; and the total probability of this case is obtained by summing the probabilities over all possible values for d, i.e.,

$$\binom{N-1}{k-1} \sum_{d=0}^{k-1} \left[\binom{k-1}{d} \cdot p_0 p_1^d (1-p_1)^{N-1-d} \cdot q_{out}^{(k-1)-d} \cdot (1-q_{out})^{N-k} \right] . \quad (7)$$

(2) In the second case, each of the $k-1$ external compromised nodes must be directly compromised, because the center node is not directly compromised, and only the center node can indirectly compromise external nodes. For a fixed choice of these $k-1$ compromised external nodes, the probability of this configuration is the product of the probabilities that

(a) the center node is not directly compromised, but is indirectly compromised
(b) $k-1$ exterior nodes are directly compromised
(c) $N-k$ exterior nodes are not directly compromised
which gives

$$\begin{aligned}&(1-p_0)\cdot(1-(1-q_{in})^{k-1})\cdot p_1^{k-1}\cdot(1-p_1)^{N-k}\\ =&p_1^{k-1}(1-p_0)(1-p_1)^{N-k}\cdot(1-(1-q_{in})^{k-1})\ .\end{aligned}$$

There are $\binom{N-1}{k-1}$ ways to choose the external compromised nodes, so the probability of this case is

$$\binom{N-1}{k-1}p_1^{k-1}(1-p_0)(1-p_1)^{N-k}\cdot(1-(1-q_{in})^{k-1})\ . \tag{8}$$

(3) In the third case, there are k external compromised nodes, each of which must be directly compromised; and for a fixed choice of these k compromised external nodes, the probability of this configuration is the product of the probabilities that
(a) the center node is neither directly nor indirectly compromised
(b) k exterior nodes are directly compromised
(c) $N-1-k$ exterior nodes are not directly compromised
which gives

$$\begin{aligned}&(1-p_0)\cdot(1-q_{in})^{k}\cdot p_1^{k}\cdot(1-p_1)^{N-1-k}\\ =\ &p_1^{k}(1-p_0)(1-p_1)^{N-1-k}\cdot(1-q_{in})^{k}\ .\end{aligned}$$

There are $\binom{N-1}{k}$ ways to choose the external compromised nodes, so the probability of this case is

$$\binom{N-1}{k}p_1^{k}(1-p_0)(1-p_1)^{N-1-k}\cdot(1-q_{in})^{k}\ . \tag{9}$$

Finally, the total probability of k losses is the sum of Equations (7), (8) and (9).

$$\begin{aligned}&\binom{N-1}{k-1}\sum_{d=0}^{k-1}\left[\binom{k-1}{d}\cdot p_0 p_1^d(1-p_1)^{N-1-d}\cdot q_{out}^{(k-1)-d}\cdot(1-q_{out})^{N-k}\right]\\ &+\binom{N-1}{k-1}p_1^{k-1}(1-p_0)(1-p_1)^{N-k}\cdot(1-(1-q_{in})^{k-1})\\ &+\binom{N-1}{k}p_1^{k}(1-p_0)(1-p_1)^{N-1-k}\cdot(1-q_{in})^{k}\ .\end{aligned}$$

□

C. NUMERICAL RESULTS

Table II: Comparison of the actual loss distribution in the one-hop model to the binomial distribution for various direct compromise and propagation probabilities, and constant network size $N = 600$.

p	q	$E[NL]$	Variance $Var(NL)$		Quantile $Q(NL, 0.999)$		Safety loading $Q(NL, 0.999) - E[NL]$	
			actual	binomial	actual	binomial	actual	binomial
0.005	0.05	4.21	7.90	4.18	17	12	12.79	7.79
	0.1	5.41	15.30	5.36	25	14	19.59	8.59
	0.5	14.72	154.90	14.36	83	28	68.28	13.28
0.01	0.05	8.41	15.55	8.29	25	19	16.59	10.59
	0.1	10.78	29.88	10.59	35	22	24.22	11.22
	0.5	28.77	281.91	27.39	107	46	78.23	17.23
0.05	0.05	41.29	68.56	38.45	70	62	28.71	20.71
	0.1	52.00	121.38	47.50	91	74	39.00	22.00
	0.5	123.74	784.98	98.22	223	155	99.26	31.26

Table III: Comparison of the actual loss distribution in the one-hop model to the binomial distribution for various network sizes and constant $p = 0.01$, $q = 0.1$.

N	$E[NL]$	Variance $Var(NL)$		Quantile $Q(NL, 0.999)$	
		actual	binomial	actual	binomial
400	7.21	19.55	7.08	27	17
600	10.80	29.33	10.61	34	22
800	14.34	39.64	14.09	41	27
1000	17.91	49.65	17.59	47	32

Table IV: Comparison of the actual loss distribution in the multiple-hop model to the binomial distribution for various direct compromise and propagation probabilities, and constant network size $N = 600$.

p	q	$E[NL]$	Variance $Var(NL)$		Quantile $Q(NL, 0.999)$		Safety loading $Q(NL, 0.999) - E[NL]$	
			actual	binomial	actual	binomial	actual	binomial
0.005	0.05	10.10	203.77	9.93	82	21	71.90	10.90
	0.1	83.94	6006.82	72.20	213	111	129.06	27.06
	0.5	548.10	17401.00	47.41	592	568	43.90	19.90
0.01	0.05	19.21	326.60	18.59	90	34	70.79	14.79
	0.1	125.18	4366.15	99.06	218	157	92.82	31.82
	0.5	578.21	1010.09	21.00	592	591	13.79	12.79
0.05	0.05	71.86	414.28	63.25	129	97	57.14	25.14
	0.1	188.71	398.83	129.36	245	224	56.29	35.29
	0.5	580.80	20.59	18.59	593	593	12.20	12.20

Table V: Comparison of the actual loss distribution in the multiple-hop model to the binomial distribution for various network sizes and constant $p = 0.01$, $q = 0.1$.

N	$E[NL]$	Variance $Var(NL)$		Quantile $Q(NL, 0.999)$	
		actual	binomial	actual	binomial
400	70.21	2650.25	57.88	155	95
600	118.89	4250.06	95.37	214	150
800	171.37	5271.09	134.66	272	208
1000	219.96	5738.59	171.58	327	261

Table VI: Comparison of the actual loss distribution of randomly drawn samples in the one-hop model to the binomial distribution for various sample sizes and constant $N = 600$, $p = 0.05$, $q = 0.1$.

Sample size	$E[NL]$	Variance $Var(NL)$		Quantile $Q(NL, 0.999)$	
		actual	binomial	actual	binomial
30	2.48	2.39	2.28	8.00	8.00
60	5.22	5.31	4.76	13.33	13.33
120	10.11	11.68	9.26	22.00	20.67
240	20.81	30.14	19.01	40.00	35.33

Table VII: Comparison of the actual loss distribution of randomly drawn samples in the multiple-hop model to the binomial distribution for various sample sizes and constant $N = 600$, $p = 0.05$, $q = 0.1$.

Sample size	$E[NL]$	Variance $Var(NL)$		Quantile $Q(NL, 0.999)$	
		actual	binomial	actual	binomial
30	8.98	6.57	6.27	17.00	17.00
60	19.31	15.44	13.08	31.67	31.00
120	36.83	35.95	25.52	55.33	52.67
240	74.10	98.15	51.21	103.67	96.67