How Many Down?
Toward Understanding Systematic Risk in Networks

Benjamin Johnson
University of California,
Berkeley

Aron Laszka
Budapest University of
Technology and Economics

Jens Grossklags
The Pennsylvania State
University

ABSTRACT

The systematic risk of a networked system depends to a large extent on its topology. In this paper, we explore this dependency using a model of risk propagation from the literature on interdependent security games. Our main area of focus is on the number of nodes that go down after an attack takes place. We develop a simulation algorithm to study the effects of such attacks on arbitrary topologies, and apply this simulation to scale-free networks. We investigate by graphical illustration how the outcome distribution of such networks exhibits correlation effects that increase the likelihood of losing more nodes at once – an effect having direct applications to cyber-insurance.

Categories and Subject Descriptors

K.6.m [**Management of Computer and Information Systems**]: Miscellaneous—*Security, Insurance*; C.2.0 [**Computer Communication Networks**]: General—*Security and protection*

General Terms

Security, Economics, Measurement, Management, Theory

Keywords

Networks, Security, Topology, Cyber-insurance, Risk Mitigation, Economics of Security, Scale-Free Networks

1. INTRODUCTION

Researchers across the spectrum from industry, government, and academia have struggled for years to understand and remedy market failures in the cyber-insurance industry. These failures serve to motivate a tremendous need for a better understanding of systematic risk for real-world networks.

One reason for these market failures is a lack of good data,[1] but this is certainly not the only problem. Risk correlation and interdependency combine together in networked systems to make understanding the problem hard. For example, risks to two different networked individuals can be highly correlated, due to vulnerabilities in a shared common platform, or due to a shared communication channel, or for a variety of other reasons. This problem is further exacerbated by misaligned incentives whereby network-benefiting security practices may not be cost-effective for some individuals.

This paper aims to address this problem by analyzing the systematic risk of a networked system that is subject to both direct risks based on individual investments, and indirect risks based on the network's topology. By *systematic risk*, we mean vulnerability to events which cause catastrophic losses, for example, a very large number of individuals in the network being compromised. Systematic risks are of special interest to cyber-insurers because, in contrast to non-systematic risks, they cannot be diversified by having a sufficiently large portfolio.

Our risk propagation model is borrowed from the literature on interdependent security games, where it has been used primarily to study the incentives of individuals within a networked system. Our focus here is rather on the state of the entire networked system. In particular, we are interested in understanding the distribution on the number of lost nodes within a network after an attack occurs.

To evaluate this focus, we implement a simulation algorithm that approximates the loss distribution for an arbitrary network topology. We then apply this simulation to randomly-generated scale-free networks. We find that the loss distribution derived from this topology differs significantly from the binomial distribution (i.e., the distribution that we would see if the risks to individuals were independent). Quantifying the risk in this scenario has applications for cyber-insurance.

The rest of the paper is organized as follows. In Section 2, we discuss related work. Section 3 describes our network risk model, which comes from the literature on interdependent security games. In Section 4, we apply a risk simulation based on our model to randomly-generated scale-free networks, and discuss the corresponding applications to cyber-insurance. Finally, Section 5 concludes the paper.

ASIACCS '14 Kyoto, Japan
Copyright 2014 ACM 978-1-4503-2800-5/14/06 ...$15.00.
http://dx.doi.org/10.1145/2590296.2590308.

[1] Industry specialists gain valuable information from cybersecurity incident reporting. But these data may suffer from selection bias, and for a variety of reasons many incidents go unreported [13].

2. RELATED WORK

Our work is relevant to the fields of cyber-insurance, and security in networked systems. We also utilize insights from studies about the network structure of real-world networks. We review important works in these areas in the following subsections.

2.1 Cyber-insurance

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 that cause widespread damage across many systems [1, 6]. However, the market for cyber-insurance is developing at a frustratingly slow pace due to several complicating factors, which are discussed in the detailed review of the security economics and cyber-insurance literature by Böhme and Schwartz [8]. The outlined key challenges and the progress made by academics in addressing them serve as important motivators for our work.

A group of defenders might appear as a particularly appealing target to an attacker because of a high correlation in the risk profiles of the defended resources. For example, even though systems may be independently owned and administrated, they may exhibit similar software configurations leading to so-called monoculture risks [5]. Böhme and Kataria study the impact of correlation that is readily observable for an insurer, and found that the resulting insurance premiums to make the risks insurable would likely endanger a market for cyber-insurance [7]. Chen et al. study correlated risks by endogenizing node failure distribution and node correlation distribution [10]. In their work, they allow for different risk mitigation measures, but do not consider the impact on the insurability of risks, or different cases of interdependence.

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 Security in Networked Systems

The problem of interdependence in networked systems has been considered in a variety of different ways in the academic literature [22]. Varian, for example, studies security compromises that result from the failure of independently-owned systems to contribute to an overall prevention objective (i.e., security is a public good) [29]. In this model, security compromises are often the result of misaligned incentives which manifest as coordination failures, such as free-riding on others' prevention investments. Grossklags et al. extend this work to allow for investments in system recovery (i.e., self-insurance) and find that it can serve as a viable investment strategy to sidestep such coordination failures [14]. However, the availability of system recovery will further undermine incentives for collective security investments. 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 [16].

A second group of economic models derives equilibrium strategies for containing the propagation of a virus or an attack in a network. 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 [2, 26]. 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 how a virus spreads or extinguishes in a network. In the literature on epidemic models, the results of Kephart and White [20] are the closest to our analysis. Kephart and White study one of the simplest of the standard epidemic models, the susceptible-infected-susceptible (SIS) model, using various classes of networks.

Finally, a popular approach to model interdependent risk is taken by Kunreuther and Heal, and forms the basis for our formal analysis [15, 21]. 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 [17] and strategic attackers [9]. Similarly, Ogut et al. proposed a related model that allows for continuous (rather than binary) security investments [27]. Our model draws from these extensions by implicitly considering a continuum of risk parameters to study the distribution of outcomes.

2.3 Real-World Networks

The Internet and many other real-world networks have been shown to be scale-free [3]. For example, several studies have measured and made attempts to characterize properties of online social networks. Mislove et al. [25] studied a dataset with over 10 million nodes and over 300 million links to characterize properties of Flickr, YouTube, LiveJournal, and Orkut. They confirmed power law and scale free properties of these networks. More recent work has concentrated on Facebook, today's most popular online social networking platform (e.g., [12]). More generally, a scale-free network's degree distribution is a scale-free power law distribution, which is generally attributed to robust self-organizing phenomena. Recent interest in scale-free networks started with [4], 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. Li et al. introduce a new, mathematically more precise, and structural definition of "scale-free" graphs [24], which promises to offer a more rigorous and quantitative alternative. The networks discussed in our paper satisfy this definition as well.

3. MODEL OVERVIEW

Our modeling framework is grounded in a set of models introduced by Kunreuther and Heal [15, 21] in the context of interdependent security games. These models supply a mechanism for nodes in a network to be attacked, and to attack their immediate neighbors. The primary motivating purpose of these models and their myriad of extensions [9, 11, 17, 19] has been to understand what motivates networked individuals to invest in security.

Our focus in this paper differs from prior work in that we concentrate primarily on the inter-node risk transfer mechanism as opposed to individual operator decisions. Our research questions are to understand the factors contributing to the loss of a catastrophically large number of nodes, and to assess the probabilities with which these bad events may happen.

The Kunreuther–Heal risk propagation structure provides a convenient mechanism to address these questions. In this regime, each network node is exposed to risk from two different types of attacks – external and internal. A successful external attack results in a complete compromise of the node, allowing the node to attack its neighbors. A successful internal attack can results in the complete failure of a node but it does not allow the failed node to attack its neighbors.

3.1 Network Risk Model

More formally, consider a network of size N. Each node is connected to its neighbors and also to an outside system such as the internet. We thus divide threats against nodes into two types – those that originate from outside the network, and those that originate from within the network.

If a node is successfully attacked from outside the network, then the node becomes an attacker; while if a node is successfully attacked from within the network, it is simply damaged. We do not directly model the attacker in this framework. Rather, attacks occur probabilistically. An external attack against node i succeeds with probability p_i. If node i is compromised by an external attack, then q_{ij} is the conditional probability that an attack by node i against node j succeeds.

3.2 Loss Distribution

Once a set of attacks happen with their respective probabilities, we are interested in the aftermath from the perspective of the whole network. That is, we want to know how many total nodes went down from this attack.

To formalize this, let NL be the random variable that counts the total number of compromised nodes in an outcome of the model. Then we define the *loss distribution* as a set of $N+1$ numbers giving $\Pr[NL = k]$ for $k = 0, \dots, N$.

3.3 Simulation

To compute a probability distribution on these outcomes, we use simulation. The simulation repeatedly chooses outcomes from a simulated attack following the external and internal attack success probabilities. The simulation proceeds as follows:

- For each node i, decide whether node i is directly compromised (or not) at random according to p_i.
- For each externally compromised node i, and for each of its non-compromised neighbors j, decide whether node i successfully attacks node j at random according to q_{ij}.
- Count the total number of nodes that have been compromised and call this the outcome. Add 1 to the number of occurrences of this outcome.
- After the empirical distribution becomes Cauchy within some epsilon, or 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 simulation is polynomial in the size of the network, given a constant number of iterations. Furthermore, we know from the strong law of large numbers that this simulation converges to the actual function almost surely. To verify that the simulation is working, we randomly generated scale free networks using the Barabási-Albert (BA) model [4], and ran the simulation with varying number 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. As can be seen from the figure, once the number of iterations is sufficiently high, the empirical distribution reaches a fixed state.

4. SCALE-FREE NETWORKS

The BA model generates scale-free graphs in stages, based on the principle of preferential attachment, meaning that the degrees of nodes with high degrees tend to get even higher as the graph grows. The parameters of the model are the initial clique size m_0, the number of edges added at each stage m, and the total size of the graph N. An instance of a graph with these parameters is constructed randomly as follows. We first initialize with a clique of $m_0 > m$ nodes. The remaining $N - m_0$ nodes are added one at a time, with each new node being randomly connected to m existing nodes with probabilities proportional to the degrees of the existing nodes.

4.1 Simulation Results

We study the risk of scale-free networks using various metrics. For each network, we compute the mean $E[NL]$, the variance $Var(NL)$, and the value of the quantile function Q_{NL} for probability 99.9%. In probability theory, the quantile function (also called the inverse cumulative distribution function) of a random variable gives, for a probability p, the lowest value such that the outcome of the random variable is less than or equal to the value with probability p. This value will play an important role in our cyber-insurance application example.

Besides comparing the loss distributions of various networks to each other, we also compare them to loss distributions without any interdependence effects. To derive loss distributions without interdependence effects, we assume that node compromises are completely independent events. In other words, we use binomial distributions with population size N for the comparisons. Finally, to make the comparisons fair, we set the mean of each binomial distribution equal to the mean of the actual distribution. Formally, we use the binomial distribution $B(\frac{E[NL]}{N}, N)$.

First, we study the effects of varying the compromise probabilities p and q. Figure 2 and Table 1 show the actual loss distributions and the corresponding binomial distributions for various internal and external compromise probability values. In each case, the network consists of $N = 500$ nodes, and it was generated using the B-A model with parameters $m_0 = 15$ and $m = 4$. We see that, for lower internal probabilities, the relationship between external compromise probability and the mean of total loss is roughly linear. For the other metrics, the relationships to the probabilities seem much less predictable.

By comparing the actual loss distributions to binomial distributions, we see a huge difference in variability. This means that, as expected, interdependence in networks causes

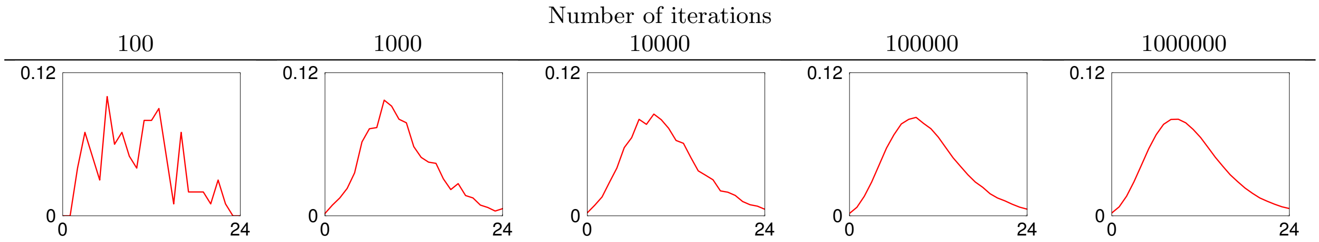

Figure 1: Loss distributions obtained from simulations with various numbers of iterations.

Table 1: Comparison of the actual loss distribution to the binomial distribution for various direct compromise and propagation probabilities, and constant network size $N = 500$.

p	q	$E[NL]$	Variance $Var(NL)$		Quantile $Q_{NL}(0.999)$	
			actual	binomial	actual	binomial
0.005	0.025	3.01	4.25	2.99	12	10
	0.05	3.51	6.51	3.49	15	10
	0.1	4.52	12.45	4.48	22	12
	0.2	6.51	29.80	6.43	35	16
	0.4	10.42	84.54	10.21	59	22
0.01	0.025	6.01	8.43	5.94	17	15
	0.05	7.01	12.81	6.92	22	16
	0.1	9.00	24.28	8.84	30	19
	0.2	12.91	56.92	12.58	47	25
	0.4	20.46	156.24	19.62	77	35
0.02	0.025	11.99	16.49	11.70	27	24
	0.05	13.97	24.85	13.58	33	27
	0.1	17.85	46.10	17.21	44	32
	0.2	25.39	104.48	24.10	66	42
	0.4	39.58	270.16	36.45	105	59

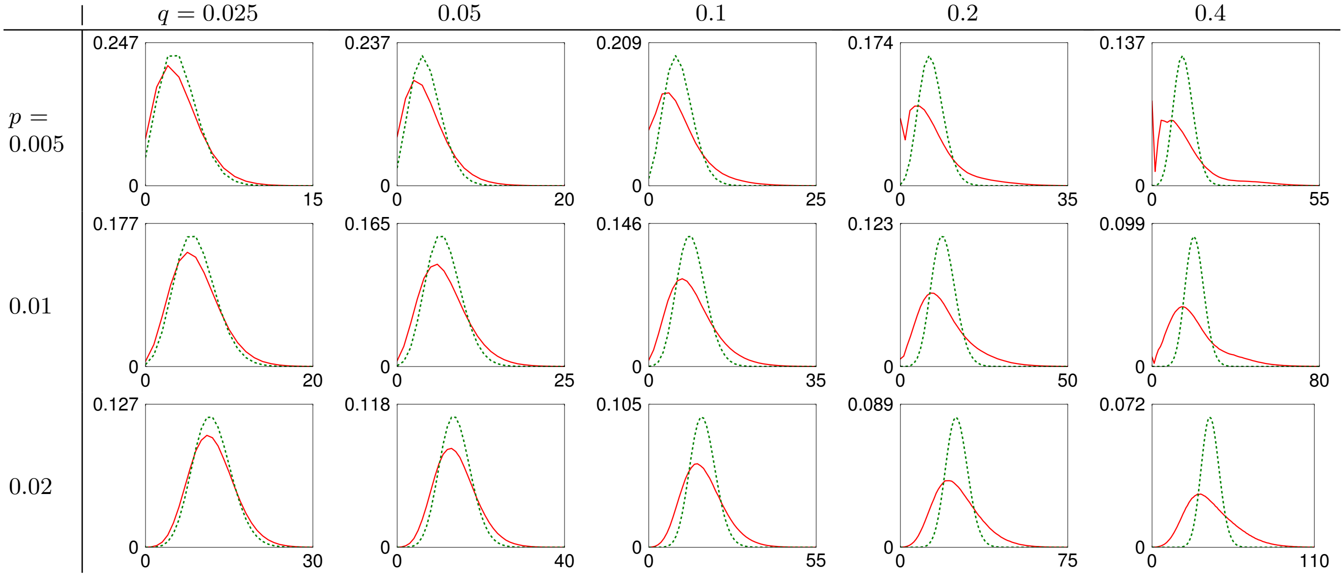

Figure 2: Comparison of the actual loss distribution (solid red) to the binomial distribution (dotted green) for various direct compromise and propagation probabilities, and constant network size $N = 500$. (Note that the slightly irregular subfigures for $p = 0.005$ and $q = 0.2$ or $q = 0.4$ are correctly drawn.)

high systematic risk. Furthermore, we see that increasing internal compromise probability has a much higher impact on systematic risk than increasing external compromise probability. Again, this is unsurprising, since internal compromises model the interdependence between the nodes, while external compromises model independent events.

Second, we study the effects of varying the size N of the network. Figure 3 and Table 2 show the actual loss dis-

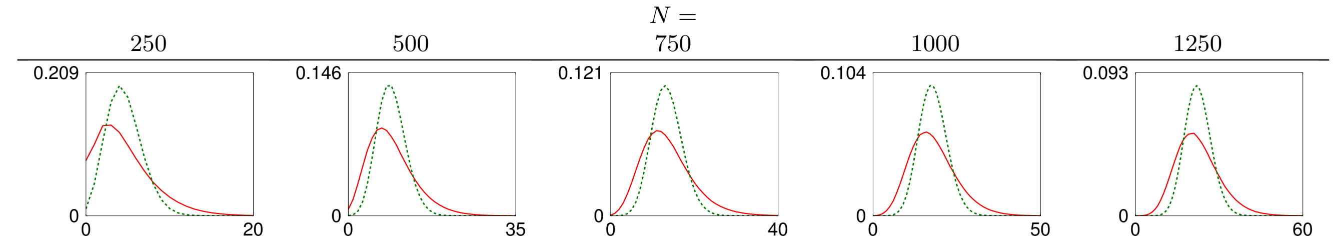

Figure 3: Comparison of the actual loss distribution (solid red) to the binomial distribution (dotted green) for various network sizes and constant $p = 0.01$, $q = 0.1$.

Table 2: Comparison of the actual loss distribution to the binomial distribution for various network sizes and constant $p = 0.01$, $q = 0.1$.

N	$E[NL]$	Var. $Var(NL)$		Quant. $Q_{NL}(0.999)$	
		actual	binomial	actual	binomial
250	4.55	12.13	4.46	21	12
500	9.01	23.98	8.84	30	19
750	13.45	37.16	13.21	39	26
1000	17.92	49.56	17.60	47	32
1250	22.37	61.75	21.97	55	38

tributions and the corresponding binomial distributions for various network sizes. In each case, the internal and external compromise probabilities are $p = 0.01$ and $q = 0.1$, and the network was generated using the B-A model with the same parameters as for the previous figure and table. The main observation here is that systematic risk does not disappear as the size of the network increases. This can be observed in both the figure and the table, which show that the difference between the actual loss distributions and the binomial distributions does not diminish. For example, the ratio between the variance of the actual distribution and the binomial distribution remains around 2.77 (more precisely, the ratios are 2.72, 2.71, 2.81, 2.82, and 2.81). In the following subsection, we will show that this has important implications for the viability of cyber-insurance.

4.2 Cyber-Insurance Example

The viability of cyber-insurance – just like any other insurance – depends on the diversifiability of risks. Diversifiability means that, in a large set of entities, the individual risks of the entities cancel out, and the aggregate risk of the whole set is relatively low. More specifically, for a sufficiently large set of entities, it is very unlikely that the total loss is much higher than its expected value, even when the individual entities themselves are very risky. In a competitive market, the diversifiability of risks leads to insurance premiums that are only marginally higher than the expected loss. Here, we show that underestimating the systematic risk of networked systems can have catastrophic consequences for an insurance provider, which highlights the importance of studying systematic risk.

Suppose that an insurer plans to provide coverage to a set of $N = 500$ networked entities, whose external and internal compromise probabilities are $p = 0.01$ and $q = 0.1$. We assume that the insurance provider is able to measure the average probability of a node being compromised, which is equal to $E[NL]/N = 1.8\%$ in this example. In practice, the insurer can obtain this value by randomly choosing nodes and measuring their probability of being compromised. On the other hand, the insurance provider is unable to learn the topology of the network, since this would require detailed data collection from a large number of nodes. Note that, for learning the network topology, random sampling is not a viable solution either [28].

To avoid bankruptcy, the insurance premiums are chosen so that the total losses exceed the sum of the premiums with only a very low probability, which is called the probability of ruin. Let us assume that the highest probability of ruin that an insurance provider can tolerate is 0.1%. Then, in the above example, the sum of the premiums has to be at least $Q_{NL}(0.999) = 30$ (see the actual quantile in Table 1). Now, consider an insurance provider who underestimates systematic risk, and assumes that individual risks are mostly independent. Based on this assumption, the insurer will calculate an incorrect sum premium $Q_{NL}(0.999) = 19$ (see the binomial quantile in Table 1). The consequences of this mistake can be catastrophic, as the probability of ruin for the incorrect sum premium is $\Pr[NL > 19] = 3.3\%$, more than 30 times the intended value.

5. CONCLUSIONS AND FUTURE WORK

Systematic risks in networked systems depend on both the network's topology and the security levels of individual nodes. In this work, we explored the loss outcomes of a model in which nodes face both external and internal risks. Applying the model to randomly generated scale-free networks, we found that the risk of catastrophe was substantially higher than if the node compromises were independent events, yielding an application to cyber-insurance.

Alongside this introductory work are many key questions, some of which we have already begun to study. The computational complexity of computing loss distributions for various classes of networks is addressed in [18]. Methods for better extrapolating the correct risk portfolio of a network from sample data on subnets are addressed in [23]. An important remaining research goal is to develop mechanisms for a cyber-insurer to estimate the direct risks of compromise and the probabilities of internode risk transfer for relevant real-world networks.

Acknowledgements: We thank the reviewers for their comments on an earlier draft of the paper. We gratefully acknowledge the support of the Penn State Institute for CyberScience and the National Science Foundation under ITR award CCF-0424422 (TRUST). In addition, this research was partially sponsored by the Army Research Laboratory and was partially accomplished under Cooperative

Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

6. REFERENCES

[1] R. Anderson. Liability and computer security: Nine principles. In *Proceedings of the Third European Symposium on Research in Computer Security (ESORICS)*, pages 231–245, Nov. 1994.

[2] J. Aspnes, K. Chang, and A. Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *Journal of Computer and System Sciences*, 72(6):1077–1093, Sept. 2006.

[3] A.-L. Barabási. Scale-free networks: A decade and beyond. *Science*, 325(5939):412–413, July 2009.

[4] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, Oct. 1999.

[5] K. Birman and F. Schneider. The monoculture risk put into context. *IEEE Security and Privacy*, 7(1):14–17, Jan. 2009.

[6] R. Böhme. Towards insurable network architectures. *it - Information Technology*, 52(5):290–293, Sept. 2010.

[7] R. Böhme and G. Kataria. Models and measures for correlation in cyber-insurance. In *Workshop on the Economics of Information Security*, June 2006.

[8] R. Böhme and G. Schwartz. Modeling cyber-insurance: Towards a unifying framework. In *Workshop on the Economics of Information Security*, June 2010.

[9] H. Chan, M. Ceyko, and L. Ortiz. Interdependent defense games: Modeling interdependent security under deliberate attacks. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 152–162, Aug. 2012.

[10] P.-Y. Chen, G. Kataria, and R. Krishnan. Correlated failures, diversification, and information security risk management. *MIS Quarterly*, 35(2):397–422, June 2011.

[11] S. Dhall, S. Lakshmivarahan, and P. Verma. On the number and the distribution of the Nash equilibria in supermodular games and their impact on the tipping set. In *Proceedings of the International Conference on Game Theory for Networks (GameNets)*, pages 691–696, May 2009.

[12] M. Gjoka, M. Kurant, C. Butts, and A. Markopoulou. Practical recommendations on crawling online social networks. *IEEE Journal on Selected Areas in Communications*, 29(9):1872–1892, Oct. 2011.

[13] L. Gordon, M. Loeb, and W. Lucyshyn. Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy*, 22(6):461–485, November–December 2003.

[14] J. Grossklags, N. Christin, and J. Chuang. Secure or insure?: A game-theoretic analysis of information security games. In *Proceedings of the 17th International World Wide Web Conference (WWW)*, pages 209–218, Apr. 2008.

[15] G. Heal and H. Kunreuther. Interdependent security: A general model. Working Paper No. 10706, NBER, August 2004.

[16] B. Johnson, R. Böhme, and J. Grossklags. Security games with market insurance. *Decision and Game Theory for Security*, pages 117–130, 2011.

[17] B. Johnson, J. Grossklags, N. Christin, and J. Chuang. Uncertainty in interdependent security games. *Decision and Game Theory for Security*, pages 234–244, 2010.

[18] B. Johnson, A. Laszka, and J. Grossklags. The complexity of estimating systematic risk in networks. Working paper, Feb. 2014.

[19] M. Kearns and L. Ortiz. Algorithms for interdependent security games. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*, pages 561–568. MIT Press, 2004.

[20] J. Kephart and S. White. Directed-graph epidemiological models of computer viruses. In *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359, May 1991.

[21] H. Kunreuther and G. Heal. Interdependent security. *Journal of Risk and Uncertainty*, 26(2):231–249, 2003.

[22] A. Laszka, M. Felegyhazi, and L. Buttyán. A survey of interdependent security games. Technical Report CRYSYS-TR-2012-11-15, CrySyS Lab, Budapest University of Technology and Economics, Nov 2012.

[23] A. Laszka, B. Johnson, J. Grossklags, and M. Felegyhazi. Estimating systematic risk in real-world networks. In *Proceedings of the 18th International Conference on Financial Cryptography and Data Security (FC)*, 2014.

[24] L. Li, D. Alderson, J. Doyle, and W. Willinger. Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, 2(4):431–523, 2005.

[25] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pages 29–42, 2007.

[26] T. Moscibroda, S. Schmid, and R. Wattenhofer. When selfish meets evil: Byzantine players in a virus inoculation game. In *Proceedings of the ACM Symposium on Principles of Distributed Computing*, pages 35–44, 2006.

[27] H. Ogut, N. Menon, and S. Raghunathan. Cyber insurance and IT security investment: Impact of interdependent risk. In *Workshop on the Economics of Information Security*, 2005.

[28] M. Stumpf, C. Wiuf, and R. May. 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):4221–4224, 2005.

[29] H. Varian. System reliability and free riding. In J. Camp and S. Lewis, editors, *Economics of Information Security*, pages 1–15. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.