Models and Measures for Correlation in Cyber-Insurance

Rainer Böhme
Institute for System Architecture
Technische Universität Dresden
rainer.boehme@tu-dresden.de

Gaurav Kataria
Heinz School of Policy and Management
Carnegie Mellon University
gauravk@andrew.cmu.edu

WORKING PAPER

Abstract

High correlation in failure of information systems due to worms and viruses has been cited as major impediment to cyber-insurance. However, of the many cyber-risk classes that influence failure of information systems, not all exhibit similar correlation properties. In this paper, we introduce a new classification of correlation properties of cyber-risks based on a twin-tier approach. At the first tier, is the correlation of cyber-risks within a firm i.e. correlated failure of multiple systems on its internal network. At second tier, is the correlation in risk at a global level i.e. correlation across independent firms in an insurer's portfolio. Various classes of cyber-risks exhibit different level of correlation at two tiers, for instance, insider attacks exhibit high internal but low global correlation. While internal risk correlation within a firm influences its decision to seek insurance, the global correlation influences insurers' decision in setting the premium. Citing real data we study the combined dynamics of the two-step risk arrival process to determine conditions conducive to the existence of cyber-insurance market. We address technical, managerial and policy choices influencing the correlation at both steps and the business implications thereof.

Revision 0.3: Workshop on the Economics of Information Security (WEIS)
University of Cambridge, UK, June 2006

Contents

1 Introduction

The usual approach to managing information security risk is similar to other business risks, i.e. first eliminate, then mitigate, absorb and then if possible, transfer. Since eliminating security risks in today's environment is not possible, managers deploy protection technologies like firewall, antivirus, encryption, and instate appropriate security policies like passwords, access control, port blocking etc. to mitigate the probability of a break-in or failure. If the residual risk is manageable it is absorbed, otherwise, transfered by either outsourcing security or buying insurance.

Though this approach seems appropriate, it creates a widening rift between security experts who propose employing standardized best practices and deploying homogeneous software to enhance system manageability thereby reducing vulnerabilities, versus those, who propose using cyber-insurance as a means of transferring risks associated with system vulnerabilities. This is because insurance relies on the principle of independent risks while standardized system environments by themselves create a global monolithic risk manifested in virtually every standardized system. Unlike in physical world where risks are geographically dispersed, in information world, network exploits, worms and viruses span all boundaries. All systems that run standardized software and processes are vulnerable, because bugs in them, once discovered, are common knowledge and can be exploited anywhere. This potentially creates a situation where not only *all* systems within an organization could fail by virtue of their being identical and vulnerable to same exploits, but all similar systems worldwide could fail affecting many organizations simultaneously as seen in case of worms like *SQL Slammer*, *Code Red* etc. Ironically, most techniques for security risk mitigation could themselves induce correlated failures as they too are standardized. For instance, antivirus updates, IDS attack signatures and software patches are all downloaded from web sources, which, if compromised can in turn compromise millions of systems that depend on them for their security [3]. Such possibilities should surely cross the mind of an insurer who plans to offer cyber-insurance to only those businesses which "responsibly" manage their information system by "timely" updating their antivirus, firewall, IDS etc.

The existence of high correlation in breach or failure of information systems adds a new dimension to risk management that has rarely been looked at in the context of information security [17, 10]. Information security risk management has been studied by Soo Hoo [46], Schechter et al. [42], Arora et al. [1], Cavusoglu et al. [9] and Gordon et al. [19, 21]. Majuca et al. [26] propose cyber-insurance as an effective strategy for security risk management. They study the evolution of cyber-insurance market citing *moral hazard* and *adverse selection* as the primary concerns. Ogut et al. [30] and Kunreuther et al. [25] discuss interdependent risks between firms and their suppliers. Yet, most studies in this area have not explicitly modeled correlated risks and the impediments they cause to cyber-insurance except Böhme [4] and Geer et al. [17]. In the insurance and actuarial literature, the research on aggregation of correlated risks and extreme value theory (EVT) is abundant [14]. However, the research in that area has not focused on modeling correlated risks within a single firm seeking insurance. In this paper, we explicitly identify cyber-risk classes that affect internal correlation in failure and model its effect on the cyber-insurance market in general.

While global risk correlation influences insurers' decision in setting the premium, the internal correlation within a single firm influences its individual decision to seek insurance. A risk-averse firm prefers low variance of loss and hence low correlation of failure amongst its internal systems. This paper is, to the best of our knowledge, the first attempt to separately identify the internal (within a single firm) and global (across multiple firms) correlation of cyber-risks and to estimate their combined effect on the presence of cyber-insurance market. Moreover it contains as well the first empirical approach to measure correlation in cyber-insurance.

The remainder of this paper is structured as follows. Section 2 elaborates on the source of correlation of IT risks and explains how different classes of risk vary in terms of relative importance of internal and global risk correlation. Section 3 proposes a comprehensive equilibrium model for the cyber-insurance market. The model captures specific features of information assets and includes both types of risk correlation as exogenous parameters. A simulation experiment in the same section demonstrates under which configurations of internal and global correlation a cyber-insurance market may thrive. The second main contribution of this paper is discussed in Section

4, where we present a method to empirically estimate the size of correlation from distributed honeynet data. We give broad estimates for global and internal correlation, compare different models of correlation structure, and address requirements for future data collection to yield more valid and reliable results. The concluding Section 5 discusses the lessons learnt on methodological, technical, managerial and policy dimensions.

2 The Correlated Nature of IT Security Risks

Due to significant homogeneity and presence of dependencies in computer systems their failure is highly correlated. Recent spate of Internet worms like *MS-Blaster* and *Sasser* have highlighted this very threat. These worms exploited vulnerabilities present in ubiquitous Microsoft Windows operating system to infect millions of computers worldwide. Computer viruses like worms are also highly contagious. Using email to spread, *Mydoom* virus compiled for Win32 platform – generic for Windows operating system – was able to infect an estimated million computers worldwide within 5 days of its release [49]. Although worms and viruses receive maximum media attention, other factors that can cause significant economic damage to a firm's information system include, insider attacks, spam, configuration errors, hardware failure, software bugs, and theft among others [20].

2.1 Classes of Cyber-Risk and Correlation

While individual firms care about correlated failure of systems only within their own network, the insurance companies are concerned about global correlation in their entire risk portfolio because that affects the risk premium they charge individual firms. Interestingly, different classes of cyber-risks exhibit different correlation properties (see Table 1).

Table 1: Examples for different kinds of cyber-risk correlation

	Global correlation ρ_G	
Internal correlation ρ_I	Low	High
High	Insider attack	Worms and viruses
Low	Hardware failure	Spyware/phishing

The failure of a computer within a firm due to hardware problem is likely neither influenced by, nor is it expected to influence failure of other computers in the same firm or other firms, unless defective computers belong to same faulty production batch. Hardware failures can therefore be considered to exhibit low intra-firm correlation (henceforth ρ_I) and low global correlation (henceforth ρ_G). Insider attacks exhibit high ρ_I but low ρ_G because an insider who is abusing his privileges, like admin password, can affect almost all computers within his administrative domain but cannot compromise computers outside his domain [43]. In contrast, software attacks involving user interaction, such as phishing or spyware, have high ρ_G and low ρ_I because a few careless employees in many different firms may respond to a phishing email or install a new game at work thereby infecting or compromising their system. However, all such employees are likely not clustered within a single firm. Typically, worms and viruses exhibit both high ρ_I and ρ_G because they are seldom contained within a single network.

The research in network security area is striving to develop techniques to contain spread of worms and viruses by automatic generation and distribution of attack signatures [24, 45, 27]. As

these techniques make use of the concurrence of malicious traffic to identify pattern and extract signatures, global correlation may be reduced by the maturing of those technologies, but it is unlikely to vanish completely. On the other hand, internal correlation is unlikely to reduce much as the local response time required to contain a worm outbreak is too short [47, 23]. O'Donnell et al. [29] and Chen et al. [10] propose using software diversity to limit internal correlation.

2.2 Implications for Cyber-Insurance Policy Design

Reasoning about correlation also sheds new light on existing cyber-insurance products. The leading providers of cyber-insurance in the market today are AIG and Lloyd's of London. Table 2 gives a snapshot of policies on offer from AIG's *NetAdvanatage* suite of cyber-insurance products.

Table 2: Different Cyber-Insurance policies from AIG's *NetAdvantage* suite

Coverage	Product variation						
	1	2	3	4	5	6	7
Assets							
Information asset coverage					×	×	×
Network business interruption					×	×	×
Follow-up costs							
Criminal reward fund					×	×	×
Crisis communication fund					×	×	×
Malicious action							
Physical theft of data on hardware			×	×		×	×
Identify theft			×	×	×	×	×
Cyber-extortion			×	×	×	×	×
Cyber-terrorism	×	×	×	×	×	×	×
Liability							
Network security liability			×	×		×	×
Internet professional liability		×		×			×
Web content liability	×	×	×	×		×	×
Punitive, exemplary & multiple damages	×	×	×	×		×	×

Source: AIG, [26]

Majuca et al. [26] justify multiple policies from AIG as a means of product differentiation to serve different market segments. We concur with them, however, we suspect the rationale for offering multiple policies in the market may not always be to serve market segments but to sometimes also proactively segment the market into as many independent risk classes as possible. As seen from the Table 2, some policies are indeed independent of some others. Moreover, it is particularly interesting that coverage for asset losses due to generic cyber-risks are always bundled with funds covering extra expenses. The former risk classes are presumably exposed to high global correlation, whereas the latter are not (criminal rewards are paid only once and crisis communication is dispensable, yet counter-productive, if the whole industry is affected). This kind of bundling makes sense in terms of risk diversification and in terms of hiding high safety loadings for correlated risks in the composite premium.

3 Modeling the Market for Cyber-Insurance

The objective of this section is to theoretically analyze the interplay between the two types of cyber-risk correlation and its effect on the market for cyber-insurance. We present a formal model, consisting of supply- (Sect. 3.1) and demand-side (3.2) of a cyber-insurance market and

the equilibrium conditions (3.3). Inference from the model is drawn using Monte Carlo simulation methods (3.4).

3.1 Supply-Side: Two-Step Risk Arrival with Correlation

In this paper, we propose to address the particularities of cyber-risks in a two-step risk arrival process. The first step models the aggregation of cyber-risks within a single firm's network represented by n computers. The second step aggregates the risks in the portfolio of an insurer issuing coverage to k similar firms. We allow for correlation on both steps, whereas the extent of correlation may vary between the portfolio level (global correlation ρ_G) and the firm level (internal correlation ρ_I).

3.1.1 Intra-Firm Risk Correlation

The failure of computers within a firm due to a security incident[1] can be considered as a collection of correlated Bernoulli trials such that each computer failure is a coin toss, the outcome of which depends on outcome of other trials. Computers on a firm's internal network that have same configuration have a uniform correlation structure i.e. correlation of failure for any two computers within the group is same [10]. We chose to model computer failure within a firm using the Beta-Binomial (BB) distribution, which has been used in computer science literature to model correlated failure of backup systems [2] and to model failure across multiple versions of software [28]. Other approaches to model dependent Bernoulli trials include correlation with a latent random trial [4]. The Beta-Binomial distribution is computed by randomizing the parameter p (probability of failure) of the Binomial distribution by Beta distribution. This lends Bayesian subjectivity to the correlation of individual Bernoulli trials, which can be estimated by security analysts based on the technical and managerial set up in place within a firm.

The probability that $X \sim \mathrm{BB}(n, \pi, \rho_I)$ computers fail in an incident is given by

$$P(X = x|n, \pi, \rho_I) = \frac{\mathrm{B}(n - x + \beta, x + \alpha)}{(n+1)\ \mathrm{B}(n - x + 1, x + 1)\ \mathrm{B}(\alpha, \beta)} \tag{1}$$

$$\text{where } \alpha = \pi \cdot \left(\frac{1}{\rho_I} - 1\right) \ , \quad \beta = (1 - \pi) \cdot \left(\frac{1}{\rho_I} - 1\right) \text{ and } \quad \mathrm{B}(a, b) = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a + b)} \ .$$

Γ denotes the Gamma function, B is the Beta function, and the parameters include the total number of computers on the firm's network n, the (unconditional) probability of failure π, and the correlation measure ρ_I in the range 0 (no correlation) to 1 (perfect dependence).

The main focus of information security research and practice has been on reducing the mean failure rate π. However, it is important to observe that a cautiously managed homogeneous deployment of systems may exhibit low mean yet high correlation in failure. In section 4.3, we provide empirical evidence to support our claim. Though approaches like intrusion detection (IDS) are useful in proactively learning about attacks and isolating vulnerable components on the network in order to reduce chances of cascading or correlated failure [31, 39], frequent false alarms and exceedingly fast rates of infection have marred their success in practice [16].

3.1.2 Global Risk Correlation

As mentioned above, the insurer has k firms in its risk portfolio. The losses and thus claims at firms are correlated due to presence of global correlation ρ_G. We model the distribution of these correlated risks in the overall portfolio using copulas [14]. Copulas are sophisticated statistical tools to model dependence of arbitrary probability distributions. In this paper, we use the t-copula because of its property to model correlation of extreme events [11].

A k-dimensional copula C is a k-dimensional distribution function on the unit space $[0, 1]^k$ with uniform marginal distributions. Any copula C can be used to join k distributions with marginal

[1] A security incident can be defined as a collection of similar attacks spaced together in time [22].

distribution functions $F_1, \ldots, F_k$ to a multivariate distribution F as follows (Sklar's Theorem, see [14]):

$$F(x_1, \ldots, x_k) = C(F_1(x_1), \ldots, F_k(x_k)) \ . \tag{2}$$

In our model, $F_1 \ldots F_k$ are the marginal loss distribution functions of k firms in the insurers portfolio, which happen to be Beta-Binomials with parameters n, π, and ρ_I. (In a generalized case for real-world applications these parameters could vary between firms. We refrain from doing this since we lack reasonable priors). The individual loss distributions are tied together via the t-copula C_{ν,ρ_G} with density function

$$c_{\nu,\rho_G}(\mathbf{u}) = \frac{d^k_{\nu,\rho_G}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_k))}{\prod_{i=1}^k d_\nu(t_\nu^{-1}(u_i))}, \quad \mathbf{u} \in (0,1)^k \quad , \tag{3}$$

where

- d_ν is the density function of a univariate t-distribution with ν degrees of freedom,
- t_ν^{-1} is the quantile function of a univariate t-distribution with ν degrees of freedom, and
- d^k_{ν,ρ_G} is the k-dimensional joint density of a multivariate t-distribution with ν degrees of freedom and pairwise correlation ρ_G.

The probability density function of a univariate t-distribution is defined as

$$d_\nu(x) = \frac{1}{\sqrt{\nu} \cdot \mathrm{B}(\frac{\nu}{2}, \frac{1}{2})} \cdot \left(\frac{\nu}{\nu + x^2}\right)^{(1+\nu)/2} \ . \tag{4}$$

Once suitable large sets of quantitative data on cyber-losses become available, the choice of the t-copula can be reevaluated in comparison with other classes of copulas. The gist of our results, namely the relative contribution of ρ_I and ρ_G to the existence of a cyber-insurance market, does not primarily depend on the type of copula used to model the global dependence structure.

For the simulation of random variables in our complete risk arrival model with its k-dimensional multivariate Beta-Binomial distribution $\mathrm{BB}(n, \pi, \rho_I)$ and dependence structure C_{ν,ρ_G}, we employ the multivariate normal variance mixture method described in [11].

3.2 Demand-Side: Information Security Risk Management

The supply-side model presented in the previous section allows an insurer to calculate appropriate premiums for cyber-risks with a given correlation profile (ρ_I, ρ_G). A demand-side model is needed to analyze when and whether it is optimal for firms to buy insurance coverage at a given premium. In the following, we introduce a stylized model of the business value of information technology, and then discuss operable compensation schemes for cyber-insurance contracts with regard to the intangible nature of information assets and the difficulty to value and substantiate claims.

3.2.1 Modeling Information Assets

The efficacy of a firm's information system is determined by its ability to store, process and retrieve information in an efficient manner. While some industries like e-commerce depend completely on their information systems, other industries depend on them to a varying degree to carry out their business. Failure of information system due to an attack or malfunction can severely limit certain business functions that depend on information storage, processing or retrieval.[2] Therefore, most systems are designed to incorporate some level of redundancy or fault-tolerance at both communication and storage level. In a typical network setting, clients store information on servers which distribute it among other servers for consistency, load-balancing and fault-tolerance. Performance and security are generally competing goals when dealing with information [50]. No redundancy

[2]Even if fall-back plans exist, continuing core business without IT is accompanied by productivity losses.

implies higher performance and low security, while backups and consistency-checks enhance security at the cost of lowering performance [44]. Numerous threshold schemes for the design of storage systems have been proposed [37]. These schemes have three parameters: n, m and p (where $n \geq m \geq p$). We assume that the information asset of a firm is divided among n nodes on its network. Due to presence of some redundancy in the network the entire information can be recreated with help of any m nodes. Assuming that some dependencies exist among them, at least p nodes need to be compromised to breach any useful information (where p can also be equal to 1 in case of no dependency).

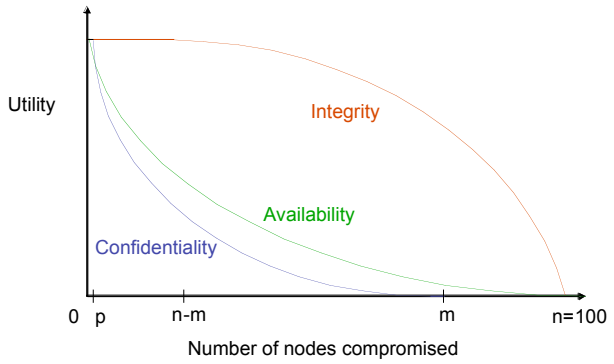

Figure 1: The fall in utility as a function of nodes compromised

Under this setup we observe the impact of node failure on the firm for each of the three common protection goals (Figure 1):

Confidentiality: To steal complete information an adversary needs to compromise at least m nodes. It can steal some information if the number of nodes breached is $\geq p$.

Integrity: Information can be restored if number of nodes compromised is no greater than $n-m$.

Availability: Due to dependencies and interconnection of nodes on the network, the failure of one node affects other nodes. The degrading effect can be high for nodes which have high dependencies like print servers, file servers, routers etc, while a stand alone desktop has only minimal effect.

From the above shown relationship between the number of failed nodes and the enforcement of security properties, specific loss functions $\ell(x)$ can be derived. A loss function maps the physical state (number of node failures) to disutility a firm faces due to that physical loss. Due to the very nature of information assets it becomes extremely difficult to objectively quantify confidentiality, integrity or availability of information and the loss caused to the firm due to breach in any/all of them. For instance, breach of two megabytes out of ten megabytes of trade secret does not necessarily translate into a 20 % breach of confidentiality.

There are some proposals for indirect measures of losses in the literature: Cavusoglu et al. [8] estimated loss in market value of traded firms due to announcement of security breaches. A similar study by Campbell et al. [7] reports higher losses in market value for security breaches where confidential data has been exposed, which is consistent with our understanding of a steeper decline in utility for this protection goal (see Figure 1). Goldfarb [18] estimated loss in market share of firms due to Denial of Service (DoS) attacks, while not explicitly sizing the attack itself.

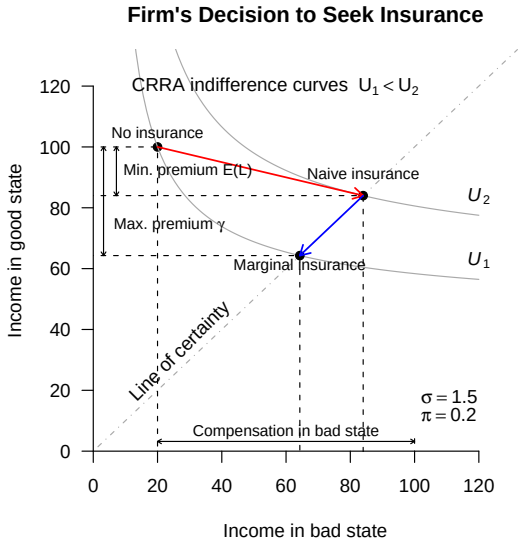

Figure 2: Demand-side model as stylized decision problem: determine the maximum gross premium γ where a risk-averse firm is indifferent between buying insurance or not. Higher premiums will thwart a functioning insurance market.

3.2.2 Firm's Decision to Seek Insurance

In our preliminary analysis we assume a linear loss function $\ell(x) \sim x$ (i.e. the dollar loss faced by a firm is linear in number of its computers affected) and a CRRA[3] utility function. A situation equivalent to risk aversion is also obtained when losses are a function of downtime and failed entities are being serviced in a maintenance queue, which is a plausible assumption for recovery processes after computer incidents [10].

A risk-averse utility function is one where a firm prefers low variance of income even when expected income is smaller than in a high variance alternative scenario. Then the decision of a firm to seek insurance for a given premium can be illustrated in a stylized two-state model, as depicted in Figure 2 (cf. [48], among others). Consider a firm with initial wealth $I_0 = 100$, which it will retain in the *good state*. In the *bad state*, however, the firm looses $q = \frac{4}{5}$ of its initial wealth. Let, bad state and good state occur with probability π and $1 - \pi$, respectively. The firm's pay-out in case of "no insurance" yields a (risk-weighted) utility U_1, which fixes the set of pay-out combinations of equal utility on an *indifference curve*. This means that all points in the set are equally preferable for the firm. If the firm was offered an insurance policy at the marginal cost of $E(L) = I_0 \cdot q \cdot \pi$, it could realize a superior utility level $U_2 > U_1$ on the *line of certainty*, where all pay-outs are independent of the state ("naive insurance" point). The limit case where the pay-out structure with insurance yields the same utility as no insurance is marked as "marginal insurance" point. It sets the upper bound for the premium and a comparison with this maximum premium allows us to decide whether firms seek insurance or not. In our simulation study we extend this model to a higher number of $(n + 1)$ states with a probability distribution function depending on the correlation parameter ρ_I (see Eq. 1), and we allow for partial insurance.

In a competitive insurance market, firms pay a premium that is marginally greater than the expected loss in order to avoid exposure to the risk. Due to the unique correlation structure of cyber-risks it is not certain that the premiums are always economically reasonable. In the next section we investigate how our models for supply-side (Sect. 3.1) and demand-side (this section) interact and identify cases where cyber-insurance is practical.

[3]Constant Relative Risk Aversion, see [36]

3.3 Market Equilibrium Conditions

In this section, we explore the conditions that need to be fulfilled for a market in cyber-insurance to thrive. As mentioned before, the economic loss due to breach/failure of an information system is difficult to calculate and substantiate. However, for insurers to come up with practical policies it is essential that the risk be objectively and unambiguously defined, therefore, we believe if claims are linearly dependent on the number of system failure then a policy can be unambiguously offered and objectively monitored. Based on this simple setup we explore the interaction between the demand side and the supply side of cyber-insurance.

Given an insurance premium of γ per node, the firm chooses the fraction λ^* as the amount of insurance coverage bought for each node on its network, which maximizes its expected utility. In the limit case $\lambda^* = 0$, the firm decides to buy no insurance at all and bear all risks internally (self-insurance, see [13]). The firm thus pays a net premium of $\lambda^* \cdot \gamma \cdot n$ to the insurance company, and in case of loss due to failure of x computers it receives a compensation of $x \cdot \lambda^*$. However, premiums are not determined exogenously, they depend on the expected expenditure of insurance companies to settle all claims in a given period. The insurers' costs C in a single period can be expressed as a sum of three components

$$C = E(L) + A + i \cdot c \quad . \tag{5}$$

Where,

- $E(L)$ is the expected loss amount, with L being a random variable.
- A is the sum of all administrative costs, which we assume to be negligible.
- c is the safety capital required to settle all claims if the realization of L turns out to the ϵ-worst case (ϵ is the probability of ruin for the insurer).
- i is the interest rate to be paid for the safety capital c. The rate should reflect the risk associated with the business in general and the choice of ϵ in particular.

Parameters ϵ and i are exogenous in our model, and we use values of $\epsilon = 0.005$ and $i = 0.1$ (similar to [4]).

Since $E(L)$ solely covers the average case, the importance of safety capital to avoid ruin of the insurance company is evident. Determining the right amount of c, however, is difficult because it depends on the tails of the loss distribution L. L itself is a sum of k correlated random variables modeling the loss amount of each individual firm in the insurer's portfolio, which is again a sum of the correlated random variables modeling the risk arrival process at each individual node in the firm. The shape of the p.d.f. after each of the convolution steps depends on the amount of correlation, so both ρ_I and ρ_G appear in the calculation of premium γ, which makes the derivation of L in closed form intractable. Consequently, we resort to Monte Carlo simulation methods to determine the regions in the $\rho_I - \rho_G$ plane, where a sound business model to offer cyber-insurance at reasonable premiums exists. This is equivalent to identifying regions with non-negative consumer and supplier surplus, therefore yielding positive welfare effects.

3.4 Simulation Results

We first calculated the premium that the insurers need to charge firms to insure risks with certain correlation properties. Then, taking the premium as given, we calculated the firm's utility both with and without insurance to determine when a firm would opt for insurance. The plots in Figure 3.4 show the premiums, and indicate which regions satisfy the conditions for insurance market to exist. The results presented in this section are based on a total of 800 million simulation trials. The upper pair of plots shows firm's decisions for a risk with probability of failure of 1 % ($\pi = 0.01$), whereas the lower pair assume that losses occur on average in every 10^{th} period ($\pi = 0.1$). The left and right plots differ in the shape parameter of the t-coplua for global correlation ν. The left side assumes heavy tails (i.e., stronger dependence in the extreme outcomes), whereas

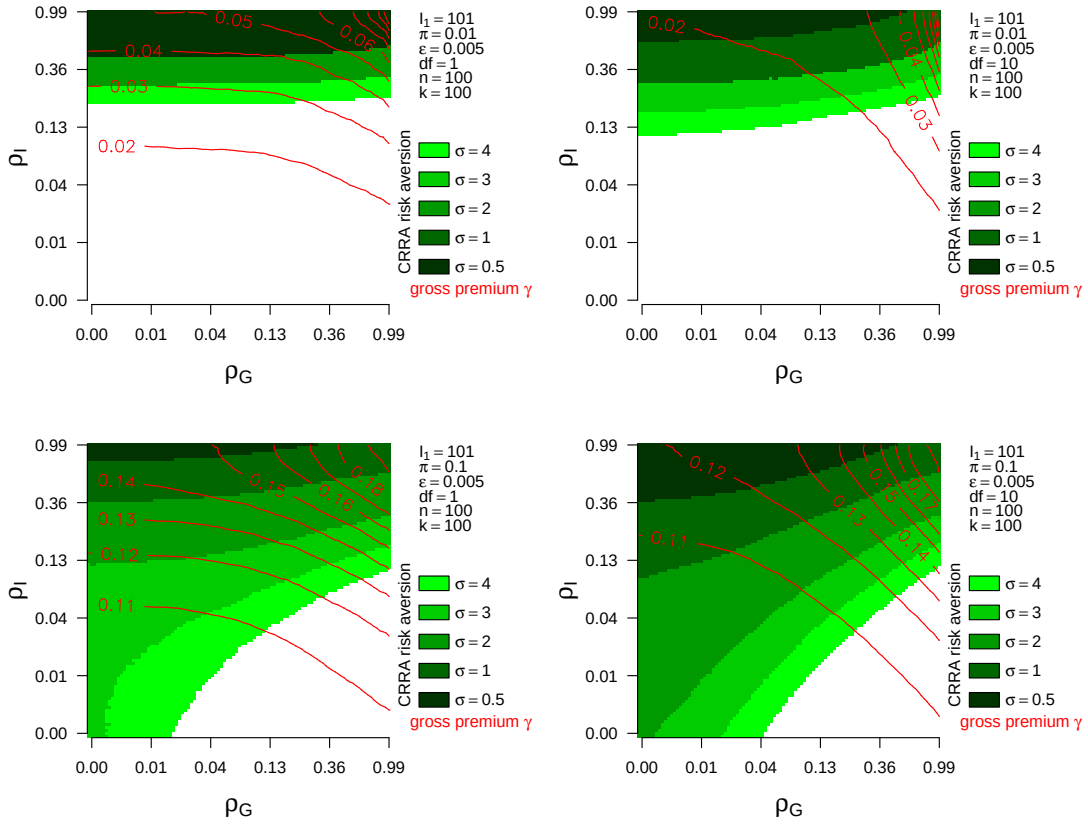

Figure 3: Red lines indicate the minimum gross insurance premium γ to cover a normalized risk of par value 1 for varying level of ρ_I and ρ_G. White areas are "uninsurable", green areas indicate regions where cyber-insurance is practical. π = prob. of computer failure; ϵ = prob. of ruin for insurer; I_1 = initial wealth of firm; df = shape of the t-copula ν; n = no. of computers per firm; k = no. of firms in insurer's portfolio; σ = coefficient of risk aversion. Results obtained from Monte Carlo simulation with 20,000 trials per parameter setting.

the right side assumes more moderate tails (i.e., the correlation is located in the center of the distribution, more similar to what a standard Pearson correlation for Gaussian random variables would imply). We make this distinctions because financial market and insurance research has observed that correlation in the tails shows up in a number of risk classes where more and reliable data exists [14]. The overlaying red lines indicate the marginal premium to cover a risk of par value 1. Therefore each of our "unit firms" would have to pay 100 times that premium to fully cover their entire network of $n = 100$ nodes. Dark areas indicate regions in the $\rho_G - \rho_I$ plane, where firms (of varying risk aversion, see legend) would decide to buy insurance, i.e., where a cyber-insurance market can exists. Empirical research on investment decisions has reported realistic values for the risk aversion coefficient between $1 < \sigma < 3$. Note that both axes are scaled on a square-root scale to allow for a better resolution in regions of small correlation.

Regarding the results, we notice that with increase in risk-aversion firms prefer insurance. However, they prefer not to insure risks if both ρ_I and probability of failure are low. This is so, because firms already achieve a kind of risk balancing within their own network and thus do not need to buy external risk balancing. Insurers, on the other hand, demand higher premium in presence of high global correlation ρ_G, which is required to balance a clustered portfolio. Therefore, we see that only firms with higher risk-aversion demand insurance when ρ_I is low and premium is high. Finally, the insurable region deteriorates for small shape parameters (df) of the t-copula, which

reflects a stronger dependency in the tails of the distribution. Since the entire joint distribution determines insurability, empirical research is needed to find the most appropriate copula and the parameterization for different classes of cyber-risks.

Lack of empirical data regarding cyber-insurance claims has often been cited as an impediment to analyzing cyber-risks. Indeed it is a "new" kind of risk, but if we understand the underlying risk generating process then we can estimate the risk correlation and thus aggregation. In the following section, we use real world data on network exploits to estimate the correlation in risk arrival both at internal as well as global level.

4 Empirical Estimation of Correlation in Risk-Arrival due to Network Exploits

The existence of correlation in cyber-risks is taken as a plausible presumption in the literature though the evidence is merely anecdotal. In this section, we use quantitative longitudinal data on attack intensity to obtain rough estimates for the range of realistic correlation parameters. Although attack statistics are not likely to provide accurate information about loss amounts, they can serve as operable proxies for loss events. In this research, we are not primarily interested in the absolute amounts but rather on the existence of a correlation structure, which is assumed to originate from the global activity of network exploits like worms and viruses. Therefore, if the hypothesis of global correlation holds for this particular risk class, then it should be measurable in the distribution of attacks over time across locations.

4.1 Description of Data

We use honeypot data to measure attack activity. Honeypots are dedicated hosts placed on the Internet which simulate the interaction of vulnerable systems and thereby pretend to be promising targets for all kinds of malicious activity. Actually, the honeypot software keeps track of all network traffic and thus serves as monitoring device for attack activities and exploit strategies. The literature distinguishes between low and high interaction honeypots [35], which differ in the degree of reactivity of the simulated system. High interaction honeypots emulate an entire operating system and are reported to deceive even human attackers working interactively, whereas low interaction honeypots merely react to the first communication attempts and therefore serve primarily as trap for automated attacks. As the latter collect more standardized and better quantifiable data, we use data from the *Leurre.com* [34] honeynet project that runs dozens of low interaction honeypot sensors deployed at partner organizations worldwide.[4] It has been stated in previous research that data collected from this kind of honeynet is suitable for monitoring the activity of malicious software automatically attacking vulnerable systems and/or propagating through the attacked hosts [33]. Therefore it should serve as a good indicator for the correlation structure behind our worms and viruses risk class (where we assume both high internal and global correlation, see Sect. 2.1 above).

The data is principally structured as event series, where records of type $(t, \mathcal{L}, \mathcal{S}, h)$ denote that sensor location $\mathcal{L}$ has recognized $h > 0$ hits of port sequence $\mathcal{S}$ at time t. Port sequences consist of one or more ports in the TCP/IP protocol being targeted in a row from a unique source as identified by the IP address and session[5]. Port sequences have been reported as a simple and quite reliable indicator for identifying attack types [35], though there might remain some ambiguity for short port sequences on popular ports (e.g., port 80 for HTTP or port 22 for ssh). Time t is measured in units of calendar days according to GMT. The raw data contains 183,000 events from 35 sensor locations[6] in the time period between February 2003 and September 2005. Hits h are the

[4]The raw data and the complete list of partners is kept confidential. Only affiliates to partner organization have access to aggregated data for research purpose.

[5]An IP address may reappear in different sessions because multiple hosts behind a NAT (Network Address Translation) router can attack using the same IP address

[6]Most partners run more than one sensor in their sub-net.

absolute number of incoming traffic from unique hosts. For certain analyses we use *attacks* instead of hits, where an attack denotes a nonzero number of hits per unique combination of $(t, \mathcal{L}, \mathcal{S})$.[7] Further, the *intensity* is defined as the average number of hits per attack.

Immediate data analysis was impeded by some inconsistency of data across time. For instance, partners joined the honeynet sequentially. Since in the aggregate data available from *Leurre.com* there was no indicator available to signal the exact uptime of individual sensor locations, the challenge in data selection lied in approximately identifying sensor downtimes, marking these periods as missing values and correctly accounting for them in the analysis to diminish the risk of spurious correlation due to a misinterpretation of sensor downtimes. Therefore, we applied the following data cleaning steps to generate our final data set for the subsequent correlation estimations:

1. Visually identified the most dense time period of 15 months between June 2004 and August 2005 and excluded all data outside this window (only 7 % of the events were dropped).
2. The remaining data was aggregated by total hits per sensor location and month. Sensor/month pairs with no hits were removed.
3. For each sensor the average hit rate per month was computed (based on the months not removed from the sample so far).
4. Those sensor/month pairs where the hit rate was below the threshold of 30 % of the sensor's individual monthly average hit rate were removed. This measure was introduced to eliminate months with partly up- and downtimes (e.g. some sensors may have gone online or offline in the middle of a month).
5. Finally, data for sensors with less than 10 months was removed (corresponds to 2/3 of the 15-months window).

The so-reduced set of about 70 % of the raw events comprises 13 sensors with relatively dense and homogeneous data. The remaining sensor locations are distributed across 4 continents (Europe being slightly over-represented) and include partners in the IT industry, research institutions and telecommunication providers. Figure 4 shows an example of the resulting data matrix for one port sequence with daily resolution in time. The attack density histograms are scaled to the individual sensor location's maximum hit rate. The distribution of hits across sensor locations is given on the right-hand axis. Plain white blocks result from sensor/month pairs that were excluded in the data cleaning phase. Note that the downtime-removal is independent of the port sequence, which contributes to carefully preventing spurious correlation.

4.2 Estimation of Global Correlation

We estimate global correlation for each port sequence independently to reduce the effect of different underlying risk-arrival processes appearing as a mixture distribution in the data. While analyzing the state of (in)security of millions of hosts present on the Internet, Burch et al. [6] found that firewall behavior varied significantly in allowing/disallowing a combination of ports. Based on their findings we believe that network exploits using different combinations of port sequences will have varying level of correlation in their ability to successfully target vulnerable hosts.

Directly calibrating the t-copula based global correlation model of Sect. 3.1.2 is not possible since the data does not indicate how many nodes of an organization were attacked at any time (we only observe attacks at the honeypots). Therefore we have to resort to simpler risk-arrival models instead. In this section, we will investigate both Beta-Binomial model (similar to the one presented earlier for internal correlation, cf. Sect. 3.1.1), and one-factor latent risk model as formulated in [4] to model global correlation. This dual approach not only allows for an assessment of the amount of correlation but also sheds light to the question which arrival process suits best to model global malicious traffic.

[7] Since event records with $h = 0$ are omitted by the data source, attacks directly map to events in the raw data.

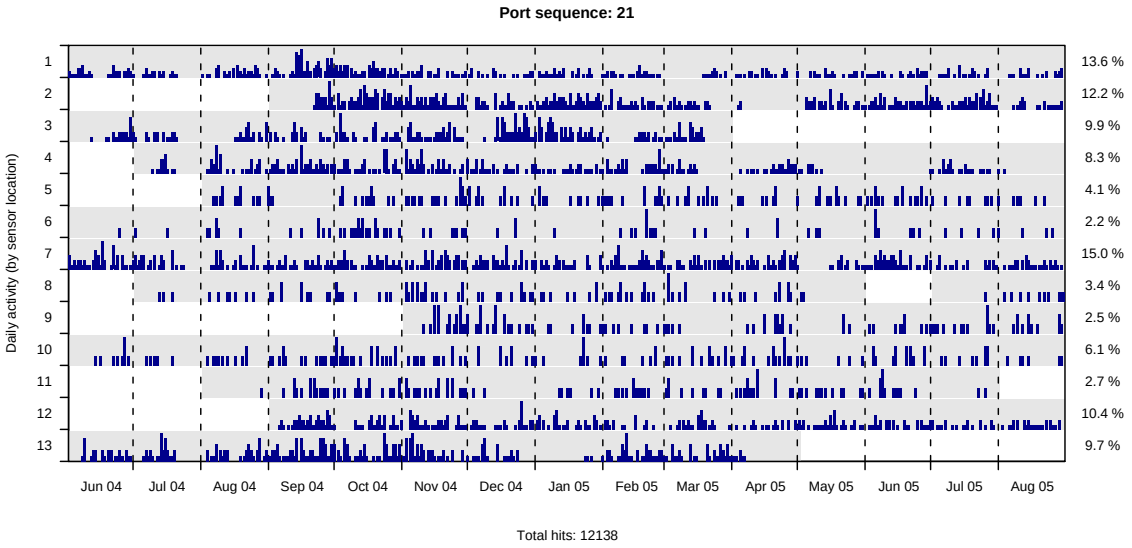

Figure 4: Attack data matrix for port 21 (FTP). Hit rate per sensor location across time.

4.2.1 Beta-Binomial Model

In Beta-Binomial (BB) model, the number of attacked sensors x_t registering at least one hit of the port sequence under investigation at day t is modeled as realization of the random variable $X \sim \mathrm{BB}(n_t, \pi, \rho_{\mathrm{BB}})$ following a Beta-Binomial distribution where n_t the total number of sensors active on day t (see Eq. 1 for the Beta-Binomial distribution function). Different points in time are assumed to yield independent realizations of X, which concurs with the notion of a stationary multivariate Bernoulli process ($t_{\max} = 457$ data points).

This distribution model belongs to the class of Generalized Additive Models for Location, Scale and Shape (GAMLSS) and can be fitted to data using maximum penalized likelihood[8] [38]. Note that censored (i.e., missing) data is implicitly accounted for by letting n_t depend on t. Table 3 reports the resulting estimates for different port sequences $\mathcal{S}$. Port sequences with fewer than 50 attacks were excluded from the analysis because the likelihood function fails to converge for very small π (due to lack of variance in the data). However, this implies a small risk that this exclusion systematically sorts out port sequences with low correlation, which has to be born in mind when interpreting the results.

According to this model, we found evidence for the existence of global attack correlation in 19 out of 27 port sequences ($\approx 70\,\%$, using the likelihood ratio test as a decision criterion with probability of false acceptance $p_\alpha < .01$). The overall amount of correlation remains very small with a median ρ_{BB} of 0.03. Single ports, however, yield point estimates for the correlation coefficient above 0.1, such as port 22 (ssh) and port 23 (telnet). According to analytical results in the literature and in the previous section, those levels of correlation can already impede the practicability of a cyber-insurance market (keeping in mind that the assumptions of the risk model and the length of an insurance period differ). Nevertheless, please consider these estimates as preliminary indications that are subject to numerous possible biases due to rigid model assumptions.

4.2.2 One-factor Latent Risk Model

To validate the results and strengthen the evidence on the existence of global correlation, we fit a second possible model for global attack dependence. In a one-factor model, the probability of attack for each node is influenced by a latent two-state variable, which can be interpreted as the

[8] We use a logistic link function for π and a square-root link for ρ_{BB} to keep the parameters in reasonable ranges.

Table 3: Estimates for global correlation in the Beta-Binomial model

		Correlation ρ_{BB}			Activity	
Port sequence	$\hat{\pi}$ (std. err.)	point est.	95 % conf. interval	AIC	attacks	intensity
21|	0.43 (0.008)	0.045	0.03 – 0.06	1919***	2.2 K	5.6
22|	0.50 (0.011)	0.124	0.10 – 0.15	2133***	2.6 K	7.0
23|	0.07 (0.005)	0.105	0.08 – 0.14	1116***	350	4.9
25|	0.31 (0.008)	0.038	0.02 – 0.06	1820***	1.6 K	4.4
80|	0.63 (0.008)	0.045	0.03 – 0.06	1920***	3.2 K	16.0
111|	0.10 (0.005)	0.024	0.01 – 0.04	1262***	508	3.6
135|	0.72 (0.007)	0.031	0.02 – 0.04	1795***	3.7 K	122.9
139|	0.59 (0.007)	0.002	0.00 – 0.02	1797	3.0 K	22.5
443|	0.18 (0.007)	0.047	0.03 – 0.07	1624***	914	4.2
445|	0.72 (0.008)	0.057	0.04 – 0.07	1865***	3.7 K	105.9
1080|	0.16 (0.007)	0.094	0.07 – 0.12	1633***	823	4.1
1433|	0.72 (0.007)	0.038	0.03 – 0.05	1828***	3.7 K	25.6
3072|	0.08 (0.005)	0.068	0.05 – 0.09	1211***	421	1.2
3128|	0.06 (0.004)	0.028	0.01 – 0.05	1033***	315	2.1
3389|	0.12 (0.005)	0.030	0.02 – 0.05	1380***	613	3.8
4128|	0.03 (0.003)	0.080	0.05 – 0.11	651***	142	4.0
4899|	0.63 (0.009)	0.068	0.05 – 0.09	1966***	3.2 K	10.4
80|57|	0.03 (0.003)	0.006	0.00 – 0.03	736	176	2.9
135|139|	0.06 (0.003)	0.007	0.00 – 0.03	973	284	5.7
135|1433|	0.02 (0.002)	0.016	0.00 – 0.05	533*	101	3.1
139|80|	0.13 (0.005)	0.042	0.03 – 0.06	1422***	637	5.0
445|139|	0.25 (0.007)	0.016	0.01 – 0.03	1687*	1.3 K	5.2
80|57|21|	0.02 (0.002)	0.012	0.00 – 0.03	502	92	3.0
135|445|80|	0.01 (0.002)	0.013	0.00 – 0.04	414	69	2.3
135|445|139|	0.06 (0.004)	0.021	0.01 – 0.04	992**	286	3.5
139|445|80|	0.14 (0.005)	0.030	0.01 – 0.05	1443***	690	2.5
135|445|139|80|	0.01 (0.002)	0.013	0.00 – 0.04	371	58	2.4

Significance codes for likelihood ratio test (LRT) between Beta-Binomial model and uncorrelated Binomial model: $*$ $= p_\alpha < .05$, $**$ $= p_\alpha < .01$, $*** = p_\alpha < .001$; residual deg. of freedom: 455

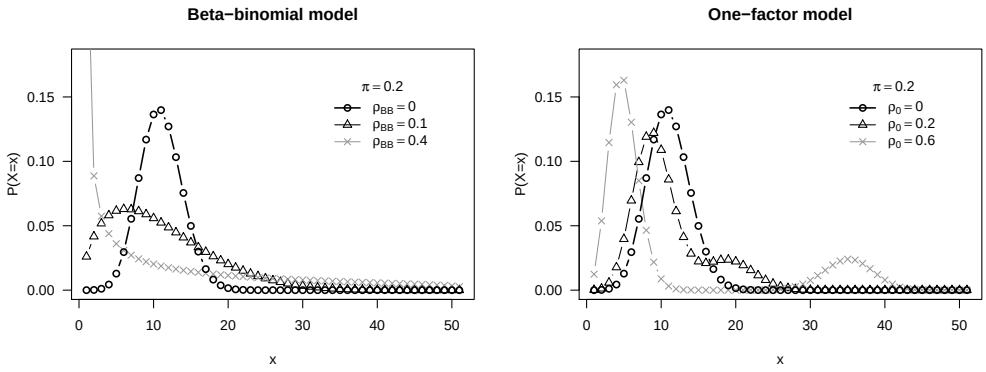

Figure 5: Theoretical probability density functions of correlation models for varying ρ_{BB} and ρ_0.

activity state of a global attack source. Hence, the number x_t of attacked nodes at time t is a realization of the random variable X, where

$$X_t(n_t, p, \rho_0) = \sum_{\mathcal{L}=1}^{n_t} R_{\mathcal{L}} \tag{6}$$

with $R_{\mathcal{L}}$ being Bernoulli variables for each sensor location $\mathcal{L}$ with total probability of attack $P(R_{\mathcal{L}} = 1) = \pi$. All $R_{\mathcal{L}}$ correlate with a latent systemic risk variable R_0, where

$$P(R_0 = 1) = P(R_{\mathcal{L}} = 1) = \pi \qquad \forall\ \mathcal{L} \in \{1, \ldots, n_t\}. \tag{7}$$

We further assume a Pearson product-moment correlation $\rho_0 \in [0, 1]$ between R_0 and each $R_{\mathcal{L}}$, so that

$$\rho_0 = \frac{E\left[(R_0 - E(R_0)) \cdot (R_{\mathcal{L}} - E(R_{\mathcal{L}}))\right]}{SD(R_0) \cdot SD(R_{\mathcal{L}})}. \tag{8}$$

As shown in [4], the distribution function of X can be expressed as a mixture of two Binomial distributions:

$$P(X_t = x | n_t, \pi, \rho_0) = \pi \cdot \text{BN}(x, n_t, \pi + (1 - \pi) \cdot \rho_0) + (1 - \pi) \cdot \text{BN}(x, n_t, \pi \cdot (1 - \rho_0)) \tag{9}$$

$$\text{with} \qquad \text{BN}(k, n, p) = \binom{n}{k} p^k (1 - p)^{n-k}\ .$$

As in the previous model, we assume a stationary process with independent realizations of X across time. Note that the total attack probability π is equivalent to the same parameter in the Beta-Binomial model, whereas ρ_0 is not directly comparable to ρ_{BB}. A comparison of the probability density function for attack counts between the Beta-Binomial and the latent factor risk model is depicted in Figure 5 for fixed $n = 50$ and $\pi = 0.2$.

Fitting the one-factor model to data requires an iterative approach, because the (assumed) latent risk factor R_0 cannot be observed directly. We employ the expectation maximization (EM) algorithm [12], which alternates until convergence between the *E-step*, where a prior for the latent variable is updated, and the *M-step* that finds the most likely model parameters based on data and the current assumption for the latent variable. In the E-step, we use the Bayes theorem to compute a vector p_t of the probabilities that $R_0 = 1$ at time t:

$$p_t = \frac{\pi \cdot \text{BN}(x_t, n_t, \pi + (1 - \pi)) \cdot \rho_0}{\text{BN}(x_t, n_t, \pi + (1 - \pi) \cdot \rho_0) + \text{BN}(x_t, n_t, \pi \cdot (1 - \rho_0))} \tag{10}$$

This, however, does not yet assure that the constraint in Eq. 7 holds after every E-step. In estimation experiments on simulated data we realized that the convergence properties can be improved by leveling the average of p_t to π. In order to avoid the values of p_t to move out of the bounds for probabilities $[0, 1]$, we implement the correction step as an additive shift of the logistic transform of p_t. Hence,

$$\hat{p_t} = f^{-1}\left(f(p_t) + d\right) \quad \text{with} \quad f(u) = \log\frac{u}{1-u} \quad \text{and} \quad f^{-1}(v) = \frac{e^v}{e^v + 1}\ . \tag{11}$$

The correction offset d is determined numerically in each iteration so that $\sum_{t=1}^{t_{\max}} \hat{p}_t = t_{\max} \cdot \pi$. In the M-step, a typical maximum likelihood approach is pursued to compute the new parameters:

$$(\hat{\pi}, \hat{\rho_0}) = \arg\max \sum_{t=1}^{t_{\max}} \log\left[\ \hat{p}_t \cdot \mathrm{BN}(x_t, n_t, \pi + (1-\pi)\cdot\rho_0) + (1-\hat{p}_t)\cdot \mathrm{BN}(x_t, n_t, \pi\cdot(1-\rho_0))\ \right] \tag{12}$$

The estimates of the one-factor latent risk model are reported in Table 4. In contrast to the previous model, the EM algorithm converges also for port sequences with few observations. We still do not report sequences with less than 50 observed attacks because an interpretation would be misleading. However, as computing the asymptotic distributions of the parameter estimates is not as straight-forward, no confidence intervals are given and we had to resort to a simulation method to obtain critical values for the hypothesis test that true $\rho_0 > 0$. $p_\alpha^{(S1)}$ is computed from EM estimates on $N = 1000$ random data sets generated strictly from the model assumptions with $\pi = \hat{\pi}$ and $\rho_0 = 0$. The reported value is the probability that random data without correlation yields an estimate $\tilde{\rho_0}$ equal or greater to the one reported from actual data. However, to exclude the possibility that positive correlation coefficients occur as an artifact of the lack of fit of the stylized model, we ran a second simulation experiment (again, $N = 1000$ for each port sequence) where the random data was generated with individual $\pi_{\mathcal{L}}$ for each sensor location – thus reflecting the differences in attack exposure between sensor locations (see Fig. 4) – and with exactly the same structure of missing data as in the original data set. Hence, $p_\alpha^{(S2)}$ is a more critical metric that captures part of the modeling error as well. It thus can be seen as a robustness check.

Using the most critical criterion, i.e., highly significant LRT, and both $p_\alpha^{(S1)}$ and $p_\alpha^{(S2)} < 0.01$, 19 out of 35 port sequences ($\approx 54\,\%$) provide evidence for attack correlation. The median $\rho_0 = 0.18$ is somewhat higher than the median for ρ_{BB} with single peaks range up to about 0.30.

4.2.3 Comparison of Models for Global Correlation

With both models estimated on the same data, we are able to compare the results and gain better insight into the structure of daily event series of attacks on the Internet. A graphical comparison is displayed in the scatterplot of Figure 6. The almost perfect correspondence of total attack probability π between both models is not really surprising (left graph). It is more noteworthy that also the correlation estimates ρ_{BB} and ρ_0, albeit on different scales, show a clear linear dependence. This finding supports the evidence for correlation in global automated network attacks across time and renders it less likely that the correlation is spurious due to random artifacts of notoriously suboptimal model assumptions.

Though both models are suboptimal because of a reduction to two parameters only, one might ask, which of the two alternative models suits better to explain the correlation structure of the data. Since both models are not nested, and completely different estimation methods were employed, a rigorous comparison is nontrivial. For a first impression, we looked at individual port sequences and plotted their probability density functions as shown in the leftmost plot of Figure 7. We use kernel smoothers to hide the discontinuities due to varying n_t. The dark solid line shows the observed attack density. It is evident that a Binomial fit (light solid line) underestimates the tails of the actual distribution, which is a clear sign for the existence of some amount of correlation. Next, compare the Beta-Binomial best fit (dotted line) with the latent factor risk model. Both allow for fat tails, but the latent factor models appears to be superior, because it does not shift the

Table 4: Estimates for global correlation in the latent factor model

		Correlation estimate				Activity	
Port sequence	$\hat{\pi}$	$\hat{\rho}_0$	$p_\alpha^{(S1)}$	$p_\alpha^{(S2)}$	AIC	attacks	intensity
21|	0.42	0.26	0.000	0.000	1708***	2.2 K	5.6
22|	0.48	0.33	0.000	0.000	1885***	2.6 K	7.0
23|	0.05	0.25	0.000	0.000	1012***	350	4.9
25|	0.29	0.23	0.000	0.000	1670***	1.6 K	4.4
57|	0.02	0.06	0.966	0.662	538*	114	3.6
80|	0.62	0.25	0.000	0.000	1722***	3.2 K	16.0
111|	0.09	0.17	0.002	0.000	1190***	508	3.6
135|	0.70	0.27	0.000	0.000	1619***	3.7 K	122.9
139|	0.58	0.17	0.795	0.000	1698***	3.0 K	22.5
443|	0.17	0.23	0.000	0.000	1482***	914	4.2
445|	0.71	0.28	0.000	0.000	1647***	3.7 K	105.9
1080|	0.13	0.28	0.000	0.000	1454***	823	4.1
1433|	0.69	0.28	0.000	0.000	1628***	3.7 K	25.6
3072|	0.07	0.22	0.000	0.000	1113***	421	1.2
3128|	0.05	0.18	0.000	0.000	965***	315	2.1
3389|	0.11	0.20	0.000	0.000	1285***	613	3.8
4128|	0.02	0.22	0.000	0.000	588***	142	4.0
4899|	0.62	0.30	0.000	0.000	1710***	3.2 K	10.4
8080|	0.09	0.12	0.707	0.000	1151***	497	3.1
28934|	0.02	0.03	0.998	0.038	531	104	1.3
80|57|	0.03	0.12	0.164	0.146	713***	176	2.9
135|139|	0.05	0.14	0.065	0.000	935***	284	5.7
135|445|	0.21	0.12	0.981	0.000	1463***	1.1 K	7.7
135|1433|	0.02	0.12	0.041	0.001	515***	101	3.1
139|80|	0.11	0.21	0.000	0.000	1311***	637	5.0
139|445|	0.45	0.16	0.857	0.000	1684***	2.3 K	8.2
139|1433|	0.02	0.07	0.925	0.716	464*	86	1.7
445|80|	0.05	0.10	0.882	0.214	843***	242	2.7
445|139|	0.24	0.19	0.034	0.000	1582***	1.3 K	5.2
80|57|21|	0.02	0.13	0.011	0.006	480***	92	3.0
135|445|80|	0.01	0.12	0.022	0.007	398***	69	2.3
135|445|139|	0.05	0.18	0.000	0.000	928***	286	3.5
139|445|80|	0.12	0.19	0.000	0.000	1350***	690	2.5
57|1433|445|139|	0.03	0.11	0.230	0.136	665***	160	3.0
135|445|139|80|	0.01	0.12	0.022	0.008	357***	58	2.4

Significance codes for likelihood ratio test (LRT) between latent factor model and Binomial model (i.e, $\rho_0 = 0$): $*$ $= p_\alpha < .05$, $**$ $= p_\alpha < .01$, $*** = p_\alpha < .001$; residual deg. of freedom: 455

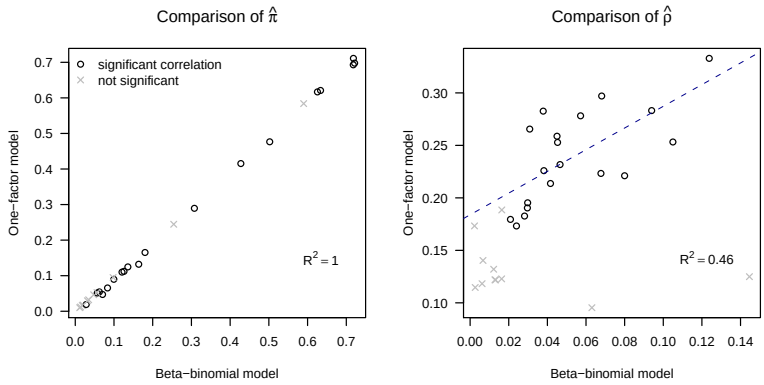

Figure 6: Comparison of parameter estimates between Beta-Binomial model and latent factor model. Each point represents one port sequence. The regression line is computed only on port sequences with significant correlation (LRT $p_\alpha < 0.01$) in both models.

probability mass to small realizations of X, as in the Beta-Binomial model. Therefor we can state that, at least for this port sequence, the assumption of one latent factor being the main source for correlation is more plausible than Bayesian uncertainty about π, as implied in the Beta-Binomial model. A similar tendency could also be observed in other port sequences, however this kind of visualization becomes less clear for extreme (high or low) total attack probabilities π.

But is one factor enough or can be do even better by allowing for more factors? To explore this question, we computed the entire correlation matrix between sensor locations using pairwise case exclusion to account for missing values. The correlation matrix was then fed into a principal component analysis (PCA) to evaluate the dimensionality of the data. The eigenvalues, as a measure of explained variance per factor, suggest five factors with eigenvalue > 1. However, the first of them is absolutely dominant (see middle plot of Fig 7). Therefore, the gain in explanatory power by adding more factors would be limited – the one-factor model does already quite well. Finally, this assertion gets even more support by the very good correspondence between the most likely realization of the (assumed) latent factor, as given from the last vector $\hat{p}_t$ of the EM algorithm, and the first principal component extracted by the PCA (right plot of Fig. 7).

Future work should relax the stationarity assumption and analyze autocorrelation in the data and the latent factor. This could provide valuable insights for insurance business because after a system is hardened against a particular vulnerability, the persistence of attack traffic to exploit that hole can be ignored. Another important goal would be to improve the data quality, both in terms of consistency (across time and sensors) and quality (using worm fingerprints instead of port sequences to identify separate risk classes).

4.3 Estimation of Internal Correlation

Unaware of internal network configuration at individual sensors domains, it is not straightforward for us to estimate the internal correlation of failure. However under two key assumptions, stated below, we are able to specify a model for internal correlation that we estimate.

1. Every computer that is successfully infected by a network exploit turns into an attacker itself and targets other computers.

2. With high probability an attacking computer targets other computers on its internal network.

The first assumption is especially true of worms, as worms by definition seek, infect and recreate themselves at vulnerable computers. Network exploits that are mounted by a human attacker

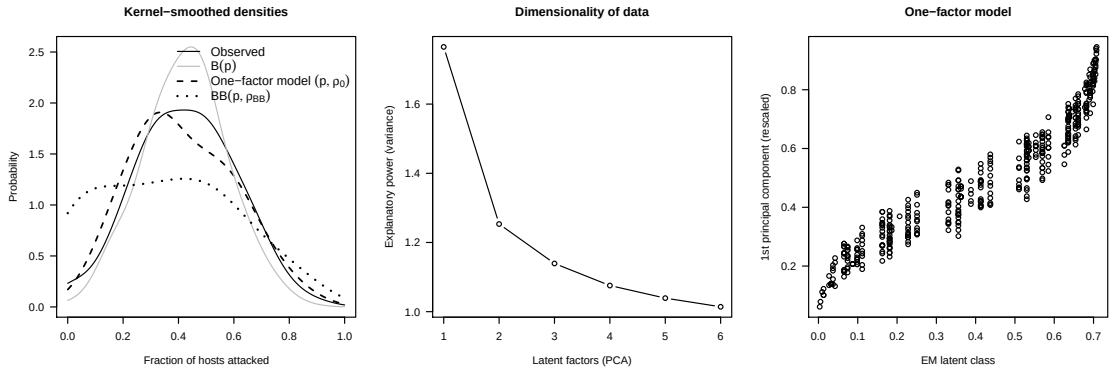

Figure 7: Drill-down to port sequence 21 (FTP): Observed and estimated probability density functions (left); eigenvalues of pairwise correlation matrix as measure of the dimensionality (middle); correspondence of estimated latent risk factor $\hat{p}_t$ after convergence of the EM algorithm with the first extracted factor of a principal component analysis (PCA, right)

himself or through a command and control channel like a botnet also satisfy this assumption to certain extent, because attackers gain is monotonic in the number of computers it can infect. The second assumption pertains to the scanning approach that a worm or an attacker employs to seek vulnerable systems. Since the publication of article on Warhol worms in 2002 by Staniford et al. [47] most worms have used subnet permutation scanning. In that article, authors hypothesized that subnet permutation scanning is superior to random scanning of IP address space in seeking potential targets, and since then we have seen almost all worms using some form of intelligent scanning. Human attackers generally scan for other vulnerable hosts on the network using an array of attack tools [40, 32]. Therefore, our second assumption is reasonably correct, however, if the internal network at a sensor domain is sufficiently partitioned via use of VLANs, internal firewall, access control lists, system and software diversity, such that infected hosts on the network are unobserved by the honeypot then, our estimate for correlation could be biased. The direction of bias will be clear after the model specification.

Under the above stated assumptions, the honeypots at a sensor location are able to observe all infected computers on their internal network $\mathcal{L}$, and can thus count the number of infected computers $x_{\mathcal{L},t}$ at time t. In the framework of Beta-Binomial model (cf. Sect. 3.1.1) of computer failure on the internal network, we are able to observe realizations x of number of computer failure per domain over a period of time. Since the size of host population at each sensor domain is not known[9] to us, we resort to host discovery approaches of actively scanning the entire host space to identify active hosts at each domain. The results of the estimation are reported in Table 5. The three rightmost columns show summary statistics on the domain's network size, where column "active" reports the number of unique hosts responding to our probes. This values has been used as a fixed $n_{\mathcal{L}}$ of the Beta-Binomial model. Column "attacking" indicates the total number of unique internal attack sources observed in the entire time frame. A subset of these sources is counted as $x_{\mathcal{L},t}$ for each day. Finally, "IP range" indicates the theoretical size of the internal network as obtained from the subnet mask. Note that we report the residual degrees of freedom for each domain since it depends on the individual uptime (N.B. the number of cases in each fitted model is given as $df + 2$).

Only five (out of 13) domains reported a considerable number of attacks from within their own network, which may either be attributed to differences in the quality of network administration or merely the fact than in some locations the sensors are placed outside the site's firewalls. For all of these five locations, a Beta-Binomial model with nonzero correlation fitted the data significantly

[9]Size of network host space is not a valid proxy for number of hosts.

Table 5: Estimates for internal correlation in the Beta-Binomial model

Domain		Correlation ρ_I		AIC / LRT	Network size (# of hosts)		
number	$\hat{\pi}$	point est.	95 % conf. int.	(residual df)	active	attacking	IP range
1.	0.00	0.0045	$0.00 - 0.01$	$123\,(302)^{\ *}$	63	9	256
2.	0.06	0.1050	$0.09 - 0.12$	$2260\,(363)^{\ ***}$	149	789	65.5 K
3.	0.00	0.0040	$0.00 - 0.00$	$1926\,(395)^{\ ***}$	2.7 K	275	65.5 K
4.	0.01	0.0166	$0.01 - 0.02$	$4925\,(455)^{\ ***}$	8.9 K	8.1 K	65.5 K
5.	0.00	0.0000	$0.00 - 0.00$	$730\,(455)^{\ ***}$	65.0 K	105	65.5 K

Significance codes for likelihood ratio test (LRT) between Beta-Binomial model and uncorrelated Binomial model: $*\ = p_\alpha < .05$, $**\ = p_\alpha < .01$, $*** = p_\alpha < .001$; Basis: daily data of all port sequences; domains were sorted by active hosts and labeled with ordinal numbers for confidentiality reasons

better than an independent Binomial process[10]. However, the resulting correlation coefficients are much smaller than in the corresponding Beta-Binomial model for global correlation (cf. Table 5).[11] It is particularly noteworthy that the size of internal correlation differs significantly between networks. We interpret this as a joint outcome of a) differences in network structure and management quality and b) biases in the estimation of the total number of hosts in each domain[12]. Future research on more reliable data should aim to disentangle the (interesting) substantial component (a) from measurement error (b).

As mentioned before, we are unaware of internal network management at any of the sensor locations, therefore, our estimates for ρ_I and π are only indicative and should not be considered definitive or conclusive. However, our approach is sound and can be used by a network administrator with contextual knowledge to determine correlation on his network. Specifically, given time series (or discrete event) data about node failure on a network the internal correlation ρ_I can be determined. Ideally, a network administrator would like to know correlation properties of different services running on the network like FTP, RPC, various messaging and P2P file sharing services, such that he can specifically allow or disallow certain service(s) based on the risk of correlation. At the same, employing the IDS, VLAN, access control and internal firewalls to limit risk correlation for all services.

4.4 Validity and Robustness

Our analyses of correlation parameters ρ_I and ρ_G is an important first step in identifying risk classes. Honeypots have proved to be successful in analyzing network-based exploits [34]. Therefore, the rate and severity of attacks observed by honeypots can serve as a reasonable proxy for rate and severity of computer failure on a network. However, as the state of computers on a network is often too dynamic – changes when new software are installed and services enabled – the relatively stationary state of a honeypot may not be truly reflective of the computers it is representing. Therefore, the honeypots may misjudge attacker behavior. Furthermore, the limited interaction capability of honeypots in our data set may bias our results toward low interaction (low level) attacks. Specifically and less importantly, in the estimation of ρ_I we may have mistaken the

[10] Here we interpret the outcome of the LRT but we ran further simulation experiments with random uncorrelated data to gain confidence and assure the correctness of our estimation procedure.

[11] Note that the two approaches of measuring correlation are exact dual of each other, in case of ρ_I we observe sources of attack, whereas, for ρ_G we count attacked sensors.

[12] Domains 2 and 5 are visibly prone to such errors, as fewer active than attacking hosts might signal changes in the network structure between data observation and size estimation, and 65+ K active hosts may result from firewalls responding to our probes. We decided not to exclude these cases to raise awareness for possible pitfalls in the proposed measurement method.

size of the networks if they have not remained constant over time. In spite of all the limitations, our estimation results for ρ_I and ρ_G are reasonable and indicative of presence of correlation at both global as well as intra-firm level.

5 Discussion

5.1 Summary of Results

In this paper, we have introduced a new classification for correlation properties of cyber-risks based on a twin-tier approach. We have shown how the two-step risk arrival process for cyber-risks can be incorporated in an economic model that takes into account the specific properties of both information assets and IT risks, namely systemic interdependence of loss events within and between firms. This model has been employed in simulation analyses to infer parameter constellations where a market for cyber-insurance can exist in theory and where it cannot.

Our simulation results indicate that cyber-insurance is best suited for classes of risk with high internal and low global correlation. This is so, because low internal correlation allows firms to realize self-insurance in their own network and thus limits demand for cyber-risk transfer. High global correlation, in turn, causes imperfect risk-pooling in the insurers' portfolios. Consequently, insurers have to add high safety loadings to the premiums and thus limit the supply for cyber-insurance.

We used honeynet data from *Leurre.com* project to estimate correlation in network exploits. For the estimation of global correlation parameter, ρ_G, we analyzed correlation of attacks across multiple globally dispersed honeypots, whereas internal correlation parameter, ρ_I, was determined by correlating instances of infected computers on the internal network of the honeypot(s). We found evidence for correlation in both dimensions, whereas the results for ρ_G turned out to be more robust and thus more valid than the results for ρ_I, which is largely attributed to the nature of our empirical data. We acknowledge that few assumptions and some data limitations may potentially bias our results. However, our estimation technique is sound and extensible.

5.2 Implications

Technical, managerial and policy approaches could be developed that can favorably alter the inherent correlation structure of the market. On the technical side, a stronger emphasis on diversity of system platforms might be an appropriate measure to counter both internal [10] and global [4] correlation. Techniques for automatic worm signature generation and distribution should be perfected, while at the same time, the current practice of unreserved auto-updates of system or application software should be reconsidered (see also [3]). On the managerial level, the recent trend to standardization, through outsourcing or other means, may create latent liabilities that have not yet appeared on the horizon of risk management and thus are not reported on the balance sheet. Policy makers can address correlation via diversity in several ways. They have indirect control of the market structure in software markets via competition policy, and/or by making cyber-insurance compulsory for certain businesses. A direct stimulus with less regulatory burden can also be given by assigning diversity a higher priority in public procurement. The exact measures and its likely outcomes, however, are to be evaluated in more targeted research and on the basis of more appropriate empirical data. This leads us to methodological implications. Apart from the frequently complained lack or unavailability of data [15], we have experienced in our research that the (restricted) data sources at our disposal lack basic statistical requirements, such as rigorous and clear standards as well as consistency over time and comparability between locations. As reliable longitudinal data on all levels of regard (network traffic, attacks, failure, losses) is indispensable for sound management of cyber-risks, we see great prospect for interdisciplinary collaboration between the network monitoring community and econometricians to define ample standards, with the envisaged types of analysis in mind, similar to those of macro-economic time series.

5.3 Directions for Future Research

Lack of research in this area could be a reason for why cyber-insurance market has not matured yet. *Mi2g*, a reputed security trend analysis and consulting firm, estimates global loss due to cyber security incidents in upwards of US $200 billion[13], while the current cyber-insurance market is worth only about US $2 billion [26]. We believe that a more detailed analysis of security outcomes following the correlation among component factors, as we describe, will be helpful in preparing market friendly coverage policies. In future work, we plan to study different loss functions, payout choices and their consequent impact on design of cyber-insurance policies.

We propose to partner with insurers in estimating the impact of technical and business processes on the level of internal as well as global correlation. The partnership is likely to yield more refined metric for correlation and their direct correspondence to systems in place like software diversity, network management and access control. Going further, insurance companies can demand critical evaluation of technologies and processes, which in turn would yield new insights for more secure and less correlated environments. Finally, the role of insurers in the assumed markets for security vulnerabilities [41, 5] has to be modeled.

Acknowledgements

This work was supported in part by grant no. CNS-0433540 from the National Science Foundation.

The authors gratefully acknowledge the support from researchers at Eurecom, France, who granted access to their fabulous *Leurre.com* database.

[13]FAQ: "SIPS and EVEDA" at `http://www.mi2g.com/cgi/mi2g/press/faq.pdf`

References

[1] A. Arora, D. Hall, C. A. Pinto, D. Ramsey, and R. Telang. Measuring the risk-based value of IT security solutions. *IEEE IT Professional Magazine*, 6(6):35–42, 2004.

[2] M. Bakkaloglu, J. Wylie, C. Wang, and G. Ganger. On correlated failures in survivable storage systems, 2002. Technical Report CMU-CS-02-129, Carnegie Mellon University, School of Computer Science.

[3] S. Beattie et al. Timing the application of security patches for optimal uptime. In *Proceedings of LISA 2002:* 16th *Systems Administration Conference*, pages 233–242, Berkeley, CA, 2002. USENIX Association.

[4] R. Böhme. Cyber-insurance revisited. In *Workshop on the Economics of Information Security (WEIS)*, Harvard University, Cambridge, MA, 2005. `http://infosecon.net/workshop/pdf/15.pdf`.

[5] R. Böhme. A comparison of market approaches to software vulnerability disclosure. In G. Müller, editor, *Proc. of ETRICS*, LNCS 3995, pages 298–311, Berlin Heidelberg, 2006. Springer Verlag.

[6] H. Burch and D. Song. A security study of the internet: An analysis of firewall behavior and anonymous DNS, 2004. Technical Report CMU-CS-04-141, Carnegie Mellon University, School of Computer Science.

[7] K. Campbell, L. A. Gordon, M. P. Loeb, and L. Zhou. The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security*, 11(3):431–448, 2003.

[8] H. Cavusoglu, B. Mishra, and S. Raghunathan. The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce*, 9(1):69–105, 2004.

[9] H. Cavusoglu, B. Mishra, and S. Raghunathan. A model for evaluating IT security investments. *Communications of the ACM*, 47(7):87–92, 2004.

[10] P.-Y. Chen, G. Kataria, and R. Krishnan. Software diversity for information security. In *Workshop on the Economics of Information Security (WEIS)*, Harvard University, Cambridge, MA, 2005. `http://infosecon.net/workshop/pdf/47.pdf`.

[11] S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review*, 71(1):111–129, 2005.

[12] A. Dempster, N. Larid, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[13] I. Ehrlich and G. S. Becker. Market insurance, self-insurance, and self-protection. *Journal of Political Economy*, 80(4):623–648, 1972.

[14] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin Heidelberg, second edition, 1999.

[15] E. Gal-Or and A. Ghose. The economic incentives for sharing security information. *Information Systems Research*, 16(2):186–208, 2005.

[16] Gartner Inc. Hype cycle for information security, May 2003. `http://www.gartner.com/5_about/press_releases/pr11june2003c.jsp`.

[17] D. Geer et al. CyberInsecurity – The cost of monopoly, 2003. `http://www.ccianet.org/papers/cyberinsecurity.pdf`.

[18] A. Goldfarb. Why do denial of service attacks reduce future visits? Switching costs vs. changing preferences. In *Workshop on the Economics of Information Security (WEIS)*, Harvard University, Cambridge, MA, 2005.

[19] L. A. Gordon and M. P. Loeb. The economics of information security investment. *ACM Transactions on Information and System Security*, 5(4):438–457, 2002.

[20] L. A. Gordon, M. P. Loeb, W. Lucyshyn, and R. Richardson. 10^{th} annual CSI/FBI computer crime and security survey, 2005. `http://www.gocsi.com`.

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

[22] J. D. Howard. *An Analysis Of Security Incidents On The Internet: 1989–1995*. PhD thesis, Carnegie Mellon University, 1997. `http://www.cert.org/research/JHThesis/Start.html`.

[23] P. Jungck and S. S. Y. Shim. Issues in high-speed internet security. *IEEE Computer*, pages 36–42, July 2004.

[24] C. Kreibich and J. Crowcroft. Honeycomb – Creating intrusion detection signatures using honeypots. In *Proceedings of the Second Workshop on Hot Topics in Networks (HotNets-II)*, November 2003.

[25] H. Kunreuther and G. Heal. Interdependent security. *Journal of Risk and Uncertainty*, 26(2/3):231–249, 2003.

[26] R. P. Majuca, W. Yurcik, and J. P. Kesan. The evolution of cyberinsurance. In *ACM Computing Research Repository (CoRR)*, Technical Report cs.CR/0601020, 2006.

[27] J. Newsome, B. Karp, and D. Song. Polygraph: Automatic signature generation for polymorphic worms. In *Proceedings of the IEEE Security and Privacy Symposium*, May 2005.

[28] V. F. Nicola and A. Goyal. Modeling of correlated failures and community error recovery in multiversion software. *IEEE Transactions on Software Engineering*, 16(3):350–359, 1990.

[29] A. J. O'Donnell and H. Sethu. On achieving software diversity for improved network security using distributed coloring algorithms. In *ACM Conference on Computer and Communications Security*, pages 121–131, 2004.

[30] H. Ogut, N. Menon, and S. Ragunathan. Cyber insurance and IT security investment: Impact of independent risk. In *Workshop on the Economics of Information Security (WEIS)*, Harvard University, Cambridge, MA, 2005. `http://infosecon.net/workshop/pdf/56.pdf`.

[31] V. Paxson. Bro: A system for detecting network intruders in real-time. In *Proceedings of the 7th Usenix Security Symposium*, January 1998.

[32] C. Phillips and L. Swiler. A graph-based system for network vulnerability analysis. In *Proceedings of the 1998 Workshop on New Security Paradigms*, pages 71–79, 1998.

[33] F. Pouget, M. Dacier, H. Debar, and V. H. Pham. Honeynets: Foundations for the development of early warning information systems. In *NATO Advanced Research Workshop*, Gdansk, 2004. `http://www.eurecom.fr/~pouget/papiers/NATO_book.pdf`.

[34] F. Pouget, M. Dacier, and V. H. Pham. Leurre.com: On the advantages of deploying a large scale distributed honeynet platform. In *Proc. of E-Crime and Computer Conference (ECCE)*, Monaco, March 29–30 2005. `http://www.honeynet.org/papers/individual/ECCE_pouget_dacier_pham.pdf`.

[35] F. Pouget and T. Holz. A pointillist approach for comparing honeypots. In K. Julisch and C. Krügel, editors, *Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA 2005)*, LNCS 3448, pages 51–68, Berlin Heidelberg, 2005. Springer Verlag.

[36] J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136, 1964.

[37] M. O. Rabin. Efficient dispersal of information for security, load balancing and fault tolerance. *Journal of the ACM*, 32(2):335–348, 1989.

[38] B. Rigby and M. Stasinopoulos. Generalized additive models for location, scale and shape. *Applied Statistics*, 54:507–554, 2005.

[39] M. Roesch. Snort – Lightweight intrusion detection for networks. In *Proceedings of the 13th Systems Administration Conference (LISA'99)*, November 1999.

[40] N. V. Scanner. `http://www.nessus.org/`.

[41] S. E. Schechter. *Computer Security Strength & Risk: A Quantitative Approach*. PhD thesis, Harvard University, Cambridge, MA, 2004.

[42] S. E. Schechter and M. D. Smith. How much security is enough? A risk management approach to computer security. In R. N. Wright, editor, *Financial Cryptography (7th Int'l Conf.)*, LNCS 2742, pages 73–87, Berlin Heidelberg, 2003. Springer Verlag.

[43] E. E. Schultz. A framework for understanding and predicting insider attacks. In *Proc. of Compsec*, pages 526–531, London, UK, October 2002.

[44] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[45] S. Singh, C. Estan, G. Varghese, and S. Savage. Automated worm fingerprinting. In *Proceedings of the 6th ACM/USENIX Symposium on Operating System Design and Implementation (OSDI)*, December 2004.

[46] K. J. Soo Hoo. *How Much Is Enough? A Risk-Management Approach To Computer Security*. PhD thesis, Stanford University, CA, 2000. `http://cisac.stanford.edu/publications/11900/`.

[47] S. Staniford, V. Paxson, and N. Weaver. How to 0wn the Internet in your spare time. In *Proceedings of the 11th Usenix Security Symposium*, August 2002.

[48] H. R. Varian. *Intermediate Microeconomics – A Modern Approach*. W. W. Norton & Company, New York, 5th edition, 1999.

[49] Wikipedia. `http://en.wikipedia.org/wiki/MyDoom`, last accessed: March 14, 2006.

[50] J. J. Wylie et al. Survivable information storage systems. *IEEE Computer*, 33(8):61–68, 2000.