

**Balancing a Balanced Budget:  
making intelligent security budget allocations**

**Adam Anthony  
University of Maryland, Baltimore County**

**December 5, 2005**

**Abstract**

The problem area for this project is that of making budget-aware decisions on computer security resource allocation. One rising theme in computer security is the use of economic models to address important problems. What researchers are discovering is that economic decisions are a contributing cause of many computer security failures. This paper will start with a short discussion of economic pressures in computer security, followed by a survey of related work, including the work of Aspnes, et al. It expands on the work of Aspnes, Chang and Yampolskiy who researched the application of what they call the sum of squares partition problem to finding an approximate solution for an optimal inoculation strategy against computer viruses in a network. For their research, an optimal strategy was one with a minimum possible cost. The research of this paper will build upon the model found in Aspnes, et. al. First, the project team will aim to determine the effect of removing simplifying assumptions in order to apply the model to a realistic business model. Second, the research will turn to investigating a variation of the model in which all components are similar with the addition of a Budget total  $B$  allocated for expenditure on security for a network. Results include three metrics for network level security as a function of spending distribution, and a condition for identifying the optimal node for spending a unit of security resources.

**Keywords**

Computer Security, Budget, game theory, graph algorithm, deployment, metric, cost effectiveness

## **Section 1: Introduction**

One rising theme in computer security is the use of economic models to address important problems. What researchers are discovering is that economic decisions are a contributing cause of many computer security shortcomings. The model defined in this paper is used to develop three metrics that help evaluate the cost-effectiveness of a security investment, as well as showing a process for determining the best node for which to spend a security resource. The paper will start with a short discussion of economic pressures in computer security and the motivation for this paper, followed by a survey of related work, including the work of (Aspnes, 2004). My research expands on the work of Aspnes, Chang and Yampolskiy who researched the application of what they call the sum of squares partition problem to finding an approximate solution for an optimal inoculation strategy against computer viruses in a network. For their research, an optimal strategy was one with a minimum possible cost. This research builds upon the model found in Aspnes, *et. al.* First, the project discusses the effect of removing simplifying assumptions in order to apply the model to a realistic business model. Then, the research will turn to investigating a variation of the model that reflects the consequences of a security investment as the probability the machine will be successfully attacked. It can be shown that such a model would revert a mixed strategy for virus inoculation, where each node has an associated  $p_i(\vec{a}) = [0,1]^n$  representing the probability that a node  $i$  will be successfully attacked.

### **1.1 Motivation**

Computer security failures are not always a result of failures in technology.

Facilities exist with which we can satisfactorily secure all of our important data.

Therefore, it is not so much a problem in the resources, but in the allocation and prioritization of computer security resources that results in major losses at the corporate and personal level. Upon reading various articles on virus protection, an interesting theme arose: a secure network does not require that every single computer on that network needs to be 100% secure. The significance of this work is that it accepts the economic pressures that often lead to insecure networks and aims to provide a method for evaluating how secure a network could possibly be with a set expense limit. If the method is found to be useful, it could raise the effectiveness of every dollar spent on computer security. Data derived from this model could also be used to convince executives that the amount of money allocated for network security is too high or too low.

## **1.2 Previous Work**

(Hamilton, 2002) is a brief survey discussing the potential problems of applying straight game theory to information security. It lists seven issues to be addressed that are different from a traditional game (i.e. chess) which are: limited examples to draw from, players making multiple simultaneous moves, opponents under no time control constraints, opponents may have different end goals, the set of known legal moves may change during the game, opponent resources and end goals may change during the game, and timing for move and state updates is not well defined. While many of these are valid concerns, it has not been concluded yet whether it is acceptable to eliminate them with a few simplifying assumptions. The game developed here strives to be simple, avoiding

many of these problems.

(Kephart, 1993) tries to adapt a well-studied example of mathematical epidemiology, originally created for biological epidemics, to the spread of computer viruses. This is an oft-repeated approach that does not involve game theory necessarily, but it is worth mentioning as a different perspective on the problem.

(Nikoletseas, 2003) is not entirely related because it does not propose an inoculation strategy. However, it does investigate how attacks propagate and it is the only article that handles the concept of a node having a probability of failure. They suggest in the conclusion that their model could be used for a game-theoretic approach, though no such approach was discussed further. (Wang, 2000) does a better job than the previous one in modeling network topologies. This paper outlines the effects of viral infection on clustered and hierarchical topologies and investigates the results of inoculating only a subset of the nodes in a graph. These papers support my assumption that an optimally secure network may not have the same level of security at each node.

(Aspnes, 2004) produced their model using an undirected graph  $G=(V,E)$  where  $V$  is a set of hosts and  $E$  is a set of connections between them. There is a uniform cost  $C$  per host for all hosts that is the cost of installing anti-virus software. There is also a uniform cost of  $L$  per host for all hosts that is the loss associated with a viral infection on that host. In this model,  $C$  and  $L$  are constant. They show that a solution to the sum of squares partition problem will yield an approximate solution to finding the optimal inoculation strategy (which would be a strategy incurring the least of total cost plus total loss). This paper out-shines all the other papers listed above because of its graceful,

simple, yet strong conclusion. For this reason, I chose to extend portions of their work to my own project. My project addresses proposed further work of researching the implications of non-constant, non-static C and L.

There have also been numerous studies proposing actual systems for securing a network, such as (Xiong, 2004) which proposed an email quarantine system for containing email worms, and (Borders, 2004) which analyzed outgoing network communications to try to detect malicious anomalies.

(Geer, 2003) have the opinion, in regards to proposed defense strategies, that “If you haven't calibrated the model with measurement, only one thing is certain: You will either overspend or under-protect.” What they mean by this is that a model is worthless if you can't conclusively say 'Model A will provide you with a more secure network than model B.' In all of the previous work listed, the final conclusion seems to be the same: the [insert model here] can be applied to reduce the number of attacks on a network. Each study also seems to emphasize the fact that their system can do something different than other systems. However, there's little proof provided that this difference is better, except for some logical hand waving. The conclusion that a security technology reduces attacks is trivial. This is what security technologies are supposed to do. What we need to be able to say is that adopting this new technology will make your network safer by a measurable amount. That is where this research improves upon previous ideas. The result of this research shows a method that allows researchers to say “After allocating this [resource/protection mechanism/model] in my network, there will be X fewer attacks.”

## **Section 2: Assumptions and limitations of (Aspnes, 2004)**

Initial impressions of the work in *Inoculation Strategies for Victims of Viruses and the Sum-of-Squares Partition Problem* (Aspnes, 2004) indicated that the model was easily expanded to notions of allocating security decisions based on a budget. Indeed, the work in (Aspnes, 2004) could be used for simple budget decisions. Because the efforts in their research were to find an optimal balance between cost of anti-virus software and losses incurred from a viral infection, the research aimed to minimize the combined costs, thereby spending the least amount of money possible. These results are important because it shows that there is a need for balance in information security budgets. On one side, their model can be used to show that a business is overspending on anti-virus software. On the other side, it can be used to show that they are not spending enough. Either of the afore-mentioned scenarios would result in higher costs for the company in the long run.

(Aspnes, 2004,18) listed several simplifying assumptions, including “every infected node infects all unprotected neighbors; the costs of installing the anti-virus software and becoming infected are known and equal for all nodes; the virus imposes no costs on protected nodes; and nodes can observe which of the other nodes intend to install the anti-virus software and adjust their own strategies in response.” Some of these assumptions are so far from the truth that the model has little to no applicability.

The biggest limitation is that the cost of securing a computer and the loss incurred from a viral infection are known and constant. The most obvious reason that this is a

limitation is because the loss is nearly impossible to predict. This is because the losses cannot possibly be known until after the fact, and may not even be realized until much later. For an example, consider a simple home user's computer that becomes infected with a virus. In some cases, the virus is particularly harmful and erases the entire drive. In other cases, the recovery cost is smaller, perhaps only requiring the user to download a hot fix application that cleans the virus from every file. Both of these losses can clearly be ranked, and are clearly different. Furthermore, a user who keeps only games and pictures has a lower loss for a total drive failure than the user who keeps games, pictures and banking records.

In reference to attack countermeasure costs, the real life costs of installing security measures are not static for each node on the network. Keeping a single node or group of nodes secure requires constant re-evaluation of what products are installed, how well they are doing, and whether or not they are being used properly. Also, at any instant, the cost of protecting each node is not constant. Some nodes require more attention (i.e a file or web server) than others (i.e. a user-level machine) in order to remain safe. There are other, more variable costs as well that differ from the base cost of purchasing and installing products. One hidden cost is the trade off for installing a security product – a computer may be more secure, but now it is totally unusable. For example, removing a computer from the network will keep it secure from electronic attacks, but at the trade off of having a disconnected computer.

This research will not try to tackle the difficult task of defining and proving the validity of different deployment costs  $C$  and attack losses  $L$ . Instead, the focus will be to

produce a variable model that will easily adapt to any definition of the measurement of C and L. The following section will show how the undirected graph model from (Aspnes, 2004) can be improved upon to accommodate a variable C and L.

### **Section 3: A new model**

Aspnes, Chang and Yampolskiy make a convincing point that the installation of security countermeasures as a broad-scale protection scheme may not be the most cost-effective choice. The new model that follows incorporates the effect of spending on other machines for the protection of a given node. This new perspective shows the negative effect that overspending on one node and under-spending on a second connected node has on the former. It is argued that this perceived effect shows evidence that a single unit of security resources can have a variable effectiveness.

#### **3.1 Formal Description**

The model has a game-theoretic flavor to it. There are two players, a system administrator and a lone hacker. In a single round, the administrator allocates a portion of a finite budget B to each node in his network. The hacker then chooses a node at random and attacks. Depending on how much money was spent on the chosen node, the hacker will have a higher or lower probability of a successful attack. If he succeeds, he is now able to attack every connected node. Each node in the system has a variable allocation of a part of B, and therefore a variable probability that the hacker will succeed at the second-level nodes. The hacker continues this pattern until he fails on every path, or the entire network is compromised. The hacker's abstract goal is to infect as many computers as possible, ideally the entire network. The system administrator's goal is to minimize the

number of infected nodes, ideally with zero nodes being infected.

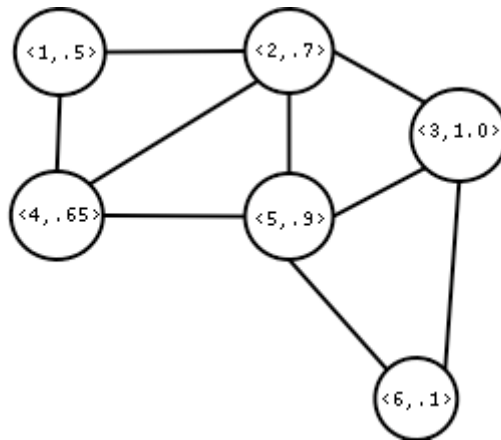
### **3.1 Extending (Aspnes, 2004) to accommodate budget**

The model in (Aspnes, 2004) has some aspects that are appealing. First, it is simply defined. A basic bi-directional graph is used to define hosts and the connections between them. Second, the process of what it means to 'secure' a node is hidden. Finally, it uses a worst-case assumption that a virus will spread to any unprotected computer that it can touch. The notion of adding budget into the model works as follows.

In (Aspnes, 2004) they characterize their model as a normal undirected graph  $G = (V, E)$ . Each node has a strategy  $0 < a_i < 1$  which describes the probability that a node  $i$  will install anti-virus software (thereby being impervious to viral attacks). Then nodes deciding to install the software (as well as their edges) are removed, yielding an attack graph  $G_a$ , which is the collection of nodes and edges remaining that are susceptible to viral infection. The game they define is a single round game where an attacker chooses a single random node to attack. If the attack succeeds, the infection will spread to all nodes connected to the initial breached node in  $G_a$ . Before the game starts, each node is allowed to choose, based on their  $a_i$  value, whether to install the software or not, also considering the decisions of neighboring nodes. Later in their paper, it is decided that allowing such a probability adds a level of complexity and only pure strategies, that is,  $a_i \in [0, 1]$ , are allowed. More on this can be found in (Aspnes, 2004, 3-5).

Regarding anti-virus software as the only protection tool necessary for securing a node is unrealistic. Firewalls, anti-spam tools, anti-spyware tools, intrusion detection systems, and user training are a few examples of the diverse number of resources that

professionals feel are necessary for securing a node. Therefore, I further theorize that instead of defining  $C$  as the one-time cost of installing anti-virus software, one can instead define  $C_i$  as the sum of costs that represent the items in any pre-determined set of security resources for node  $i$ . Then, instead of  $a_i$  being the probability that a user installs the anti-virus software, define  $a_i$  to be a probability, based on a function of  $C_i$ , that an attack on a node will succeed. That is, a probability based on how much money was spent securing a specific node that an attack will succeed. Additionally, a parameter  $B$  can be included which is the maximum amount of resources allowed for improving the security of the network. Assuming an administrator wants to spend his entire budget,  $B = \sum C_i$ . The remainder of the basic model remains the same. Below is an example network topology. Each ordered pair represents  $\langle n, a_i \rangle$  where  $n$  is the node number.



**Figure 1: An example topology with edges and  $a_i$  values**

This model is a clear adaptation of the original model, and two properties adapt:

The cost of a mixed strategy  $\vec{a} \in [0,1]^n$  to node  $i$  is

- **Equation 1:**  $\text{cost}_i(\vec{a}) = C_i + L \cdot p_i(\vec{a})$

where  $p_i(\vec{a})$  is the probability of node  $i$  being infected given  $\vec{a}$

- The total social cost of a strategy is the sum of the cost at each node.

Defining  $p_i(\vec{a})$  is somewhat complicated. The probability represented by  $a_i$  is not enough. Since the attacker chooses an entry point at random, this probability must also be accounted for. For pure strategies, (strategies where either enough money is spent for  $a_i$  to be very close to 0 or no money is spent) calculating  $p_i(\vec{a})$  is merely  $a_i \cdot k_i/n$  where  $k_i$  is the size of the connected component of  $G_a$  containing  $i$ . If the strategy permits partial spending on a node's security, finding  $p_i(\vec{a})$  is more complicated. First define  $R_{ijp}(\vec{a})$  to be the probability that a node  $i$  will be reached and successfully attacked from the initially infected node  $j$  over path  $p$ . Then  $p_i(\vec{a}) = \max_j(a_j/n, a_j/n \cdot \sum R_{ijp}(\vec{a}))$  for all  $j$ , where the summation is over all paths  $p$ . Equation 2 shows how to compute  $R_{ijp}$ .

**Equation 2:**

$$R_{ijp} = \left( \prod_p a_p \right)$$

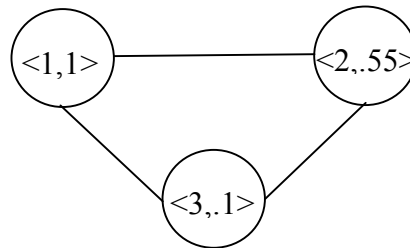
where  $p$  is a set of nodes on a path from  $i$  to  $j$

- \*note :
1. Node  $j$  not included in any set  $p$ ,
  2. No path from  $i$  to  $j$  is cyclical

Characterizations of the model concerning Nash equilibria are no longer correct. A logical reason is because in Aspnes' model, each node represents a player. In our model, we have a single player who makes decisions about protecting nodes. A rational reason is that their characterizations are simply defined using the constant  $C$  and  $L$ , which are not assumed for this paper.

### 3.2 Example

Assume that the probability of an attack succeeding on node  $i$  ( $a_i$ ) can be measured as  $a_i = 1 - \frac{C_i}{C_i + 1}$  where  $C_i$  is the amount of resources spent on node  $i$ . Note this assumption is not proven, but is a logical definition because the growth of this function is fast at first, and  $\lim_{C_i \rightarrow \infty} a_i = 0$ . It is permissible to have a different function for each node, but complicates the example. Let there be 3 nodes in a graph  $G = (V, E)$  where  $V = \{1, 2, 3\}$  and  $E = \{(1, 2), (1, 3), (2, 3)\}$ . We'll measure all  $C_i$ 's and  $L_i$ 's in hundreds of units (Dollars, if the reader requires a tangible measurement). Let  $B = 2.1$ , and let  $S$  be the spending set,  $S = \{C_1, C_2, C_3\} = \{0, 1.2, 9\}$ . This makes  $\vec{a} = \{1, .55, .1\}$ . Finally, let  $LR$  be the loss/recovery set,  $LR = \{L_1, L_2, L_3\} = \{.25, 1, 50\}$ . The graph of this situation is in figure 2.



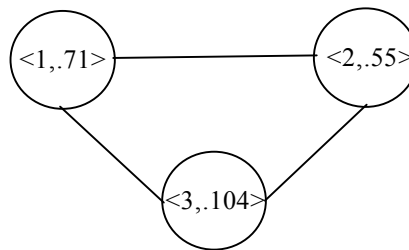
**Figure 1: Example graph**

A qualitative analysis of this graph reveals some interesting features. First, a large amount of money has been spent on protecting node 3 – 900 units. This seems reasonable, since node 3 is the most costly to lose. However, because the security spending on node 1 is so low,

$$\begin{aligned}
p_3(<0,.55,.1>) &= \max\left(\frac{.1}{3}, \frac{.55}{3} * (.1 + .1), \frac{1}{3} (.1 + .55 * .1)\right) \\
&= \max(0.033, 0.037, 0.052) \\
&= 0.052
\end{aligned}$$

What has happened here is that the insecurity of the other two machines on the network has actually caused the most valuable machine to become less secure than the security spending seemed to promise. This is because there are two paths from either node 1 or node 2 by which the node can attack node 3. This sum of attacks has caused the probability of a security breach on node 3 to increase.

Now adjust S to try to fix this problem. Assuming there is no more money to spend, let's reduce our spending on node 3 to 8.6 and allocate .4 units to securing node 1, so  $S = \{C_1, C_2, C_3\} = \{.4, 1.2, 8.6\}$  and  $\bar{a} = \{.71, .55, .104\}$ . Now our graph looks like:



**Figure 3: Same example graph with a new  $\bar{a}_i$**

And we get:

$$\begin{aligned}
p_3(<.71,.55,.104>) &= \max\left(\frac{.104}{3}, \frac{.55}{3} * (.71 * .104 + .104), \frac{.71}{3} (.104 + .55 * .104)\right) \\
&= \max(0.0347, 0.033, 0.038) \\
&= 0.038
\end{aligned}$$

What we are seeing here is the causal relationship between spending security dollars on different machines. The .4 units that were reallocated to node 1 had a higher effectiveness in protecting node 1, node 2 and node 3 than they did at node 3. The probability function

may be assumed, but the relationships between the probable failures of nodes are clearly portrayed. This example demonstrates the false sense of security that many people feel when they spend large amounts of money and resources securing their most important machines (such as a database server) but neglect the health of other machines.

#### **Section 4: Practical Use of the model**

This model, while having many interesting features that deserve further investigation, can have an immediate impact on motivating system administrators to think critically about their spending decisions. The following two sections show contrasting thought patterns that a security expert might have. The examples and the model above support the second pattern.

##### **Section 4.1 Typical decision-making process**

First, a network administrator looks at a node and says “X is what is at risk if the node is compromised, and I would need to allocate Y resources to recover from this error.” This definition of  $X + Y = L$  incorporates not only the losses but also the recovery. Both must be measured to reflect the situation where a loss is irrecoverable ( $X = \text{infinity}$ ).

Then, he looks at each node and says “The amount of resources I need to allocate for this machine to be considered safe is C.” Note that it is logical that if all  $c_i > L$ , no resources will be allocated for protecting the machine, and if  $C \leq L$  the wisest decision for that machine would be to allocate C. Finally, if  $C > L$  and some  $c_i \leq L$  then the network administrator has a difficult choice to make. Are the partial protections for node i worth reducing the probability of an attack? The function defining  $a_i$  above would suggest that it is good to spend at least a little bit of money, as early on in security purchasing, gains are

the largest. The task of discovering a probability curve that accurately reflects security spending for a host machine is discussed in the future work section. Ultimately, the trivial decision of how many  $c_i$ 's to purchase will depend on this probability function. The example above, however, shows that neglecting worthless machines is not necessarily a good idea. The following section describes the phenomena seen in the example using an analogy.

#### **4.2 Networking security and the good sibling's paradox**

A new perspective on computer security can be likened to what I call the “good sibling's paradox”. Here is the scenario. A husband and wife have two children. The younger child is well behaved, always does what she is told, and is generally delightful. The older one is a handful of trouble, is constantly misbehaving, does not help out around the house and frequently is insubordinate and shouts at his parents. Many siblings are able to understand the sibling's paradox. For even though the younger child has been taught from birth that being well-behaved is virtuous, she sees time and time again the bad child getting more attention and more praise for occasional good behavior than she receives for performing the same good acts on a daily basis. The confusion goes further when, frustrated over the lack of attention, the good child begins misbehaving in order to attract the attention that she needs. Then the parents focus on the good child, and the bad child starts acting worse than usual because there is no watchful eye. This spiral of misbehavior is difficult to stop.

Enter the life of a security administrator. Securing a network is a process, not a solution. It is a constant uphill battle that necessitates re-evaluation at every turn. With

this perspective, an administrator can feel like a frustrated parent. One section of his network starts 'misbehaving' and he must focus all of his attention on repairing it. But as he is doing this, he finds that the 'good children' in his network begin to regress for the lack of attention. A master security administrator has the natural talent to balance this directed awareness so that all parts of the network remain more or less secure. But an inexperienced administrator finds himself focusing too much attention on certain machines, and the rest of his machines suffer for the misdirected attention. When he turns to those other machines, the machine he cared for so diligently falls behind. The model above, if applied, can show the administrator where his attention will be most effective. Since under spending on a node can affect the security of the nodes getting the most attention, the informed administrator can use this info to pre-empt a situation that could soon be out of his control.

## **Section 5: Results**

The compound probability model shown above has many interesting features. Thorough, in-depth investigation of these features and their implications are left for further work. The results of this work are a new metric for network security and a strategy for re-allocating budget.

### **5.1 New metrics**

The model described here can also be used as a network security metric, provided a proper set of probability functions can be specified. The simplest way would be to define a probability threshold 't' for the network. Three metrics can be derived from the model using this threshold. The first is the number of nodes whose  $p_i(\vec{a}) < t$ . This is a

useful metric for a network administrator, so that he can say “The number of nodes with less than 1% probability of failure is greater than it was before I installed this product.”

The second metric available is a measurement of attack penetration. For each node whose individual probability is above  $t$ , how many other nodes will be infected before the probability of infecting one more node is less than  $t$ ? The third is similar to the first. It measures a network’s risk factor as  $\sum L$  where  $L = \{L_i \mid p_i(\vec{a}) > t\}$ . What this measures is the sum of losses expected from nodes whose probability of failure is higher than the acceptable threshold.

## 5.2 Budget Allocation Strategy

The cost effectiveness of a single security unit spent on a node is  $\Sigma(p_i(\vec{a}') - p_i(\vec{a})) * L_i$  where  $p_i(\vec{a}')$  is the new value for node  $i$ . The optimal spending of a unit of security spending is for the node yielding a maximum such summation.

Unfortunately, this is easier said than done. Calculating  $p_i(\vec{a})$  is a computationally intensive task, and to re-calculate it  $|V|$  times could be taxing.

Determining the best location to spend a single unit of security resources could logically lead to a method for allocating an entire budget. At the very least, one could iteratively re-compute the best location for spending one unit at a time until the entire budget is spent. Methods considered for solving this problem include dynamic programming solutions for calculating  $p_i(\vec{a})$  to reduce computations, distributed implementations that compute multiple  $p_i$ ’s in parallel (since they are independent of one another), and implementations utilizing distributed artificial intelligence agents. The final implementation of this solution is left for future work.

## **Section 6: Future Work**

The most important area of future work is the determination of valid functions of resource spending which accurately depict a probability of attack. Much work is being done in this area including research in applying Quality of Service methods (Geer, 2003), Risk Analysis (Geer, 2003), and measurements of difficulty in accomplishing an attack (Schudel, 2001). Progress in any of these areas will lead to progress in this system.

Other future work includes more research into applying this system to network security metrics, and the development of an algorithm for quickly determining the best way to spend a security resource. More time could also be spent further analyzing and defining all aspects of this system, exploring further game-theoretic aspects of the model and coding an implementation of the model.

### **7.1 Conclusion**

The ability to understand the effect of security resource spending on the actual security of a network is a difficult open problem whose solution would have major economic impact for the entire world. The work demonstrated here is an early design of a method that would allow new methods of host-based security measurement to be adapted to a flexible and descriptive resource-allocation model. Progress in this research area will give security professionals an important tool for protecting their complex systems.

## Sources Cited

- Aspnes, James, Chang, Kevin, and Yampolskiy, Aleksandr. *Inoculation Strategies for Victims of Viruses and the Sum-of-Squares Partition Problem*. Yale University Technical Report TR-1295. July 2004.
- Borders, Kevin, Prakash, Atul. Web Tap: Detecting Covert Web Traffic. Proceedings of the 11<sup>th</sup> ACM conference on Computer and Communication Security. Washington, D.C., 2004, pages 110 – 120.
- Geer, David, K. Soo Hoo, and A. Jacquity. Information Security: Why the future belongs to the quants. IEEE Security & Privacy. July – Aug. 2003, pages 24 – 32.
- Hamilton, Samuel, Miller, Wendy, Ott, Allen, Saydjari, O.S. Challenges in applying game theory to the domain of *information warfare*. 4<sup>th</sup> Information Survivability Workshop. 2002.
- Kephart, J.O. Chess, D.M. And White, S.R. *Computers and epidemiology*. IEEE Spectrum, pages 20-26. 1993.
- Nikolietseas, Sotiris, Prasinos, Grigorios, Spirakis, Paul and Zaroliagis, Christos. *Attack Propagation in Networks*, Theory of Computing Systems, Volume 36, Issue 5, September 2003, Pages 553 – 574.
- Schudel, Greg and Wood, Bradley. *Adversary work factor as a metric for information assurance*. Proceedings of the 2000 workshop on New security paradigms. Ballycotton, County Cork, Ireland 2001.
- Wang, Chenxi, Knight, John C. and Elder, Matthew C. *On computer viral infection and the effect of immunization*. Annual Computer Security Applications Conference. Pages 246 – 256, 2000.
- Xiong, Xintao. ACT: Attachment Chain Tracing Scheme for email virus detection and control. Proceedings of the 2004 ACM workshop on rapid malware. Washington D.C., 2004, pages 11 – 22.