IS MICROSOFT A THREAT TO NATIONAL SECURITY? THE EFFECT OF TECHNOLOGY MONOCULTURES ON CRITICAL INFRASTRUCTURE

Recent research has posited that technology monopolies are harmful, and specifically that monocultures like Microsoft's Windows operating system pose a national and global security threat. While the debate has been heated on both sides a quantitative analysis has not been endeavored on the topic. In this analysis the monoculture hypothesis is tested to determine the effect that diversity has on network resilience. Simulations are run to examine at what level of market share a single technology system could cause a catastrophic failure in a network. The paper also discusses the results in the context competition, anti-trust, and national security.

Introduction

A report delivered to the Computer and Communications Industry Association entitled *CyberInsecurity – The Cost of Monopoly* (Geer et. al 2003) makes a case for how Microsoft's market dominance poses a security threat, and in turn has sparked a heated debate on the topic in the media, industry and academia. Despite the controversy on the topic there has not been a quantitative analysis of the effects of diversity on resilience to address the monoculture hypothesis. There has, though, been a flurry of new findings regarding the structure of large complex networks and specifically their resiliency. This research has shown that several critical technological networks are scale free structures with power law connectivity distributions, such as the Internet at the autonomous system level and the router level (Faloutsos et al 1999), the World Wide Web (Barabasi et al 1999, Huberman and Adamic 1999), and physical SDH telecommunications networks (Spencer and Sacks 2003). Power law connectivity structures essentially illustrate the fact that a small minority of the nodes in the network has the vast majority of connections

while the vast majority of nodes in the network have only a few connections. Studies have examined the vulnerability of scale free networks finding that they are resilient to random attacks, but highly susceptible to targeted attacks (Albert et al 2000). Researchers have also examined the immunization of scale free networks including the Internet, finding that targeted immunization strategies work well, while random strategies fail to eradicate a virus below an epidemic threshold (Pastor-Satorras and Vespignani 2001, Pastor-Satorras and Vespignani 2002, Dezsos and Barabasi 2002).

These studies, though, have analyzed scale free networks assuming that all nodes are homogenously susceptible to attack or infection. Often times in real world networks only subsets of nodes are susceptible to attack or infection in a heterogeneous population of nodes. One example of such a scenario is Internet worms, which are designed to attack/exploit only specific operating systems or platforms. If a node is not running the operating system or platform that worm is designed to exploit then it cannot affect it. This quality of computer networks brings the issue back around to the monoculture hypothesis and the question: what level of diversity in node types is needed so that any one malicious attack won't cause a catastrophic failure of the network? Further, does Microsoft's market share in operating systems and servers constitute a national security threat? To test the monoculture hypothesis some background on homogeneity, what causes it, and the existing case against Microsoft is useful before describing the methodology and reporting results.

Economic Drivers of Homogeneity

A single vendor dominates many real work technology systems and networks, and this phenomenon often results in highly homogenous networks. For example Cisco systems accounts for 85.5% of the global router market and Microsoft accounts for 97.34% of the global operating system connected to the Internet (Collins 2002, Onestat 2003). In homogenous technology networks the vast majority of nodes connected to the network all run on the same platform or are built by the same vendor. Building networks where all nodes are homogeneous introduces several economic efficiencies to the network; technicians only need to be trained on one system, upgrades can be done universally at one time, and systems can be bought in larger bulk allowing per unit cost savings. To quote Ausmith (2003) "Diversity has a cost. Enterprises have standardized on specific computer hardware and software to reduce procurement, operation, and maintenace costs" (p.17).

This explains why it would economical for one network to be homogenous, but the Internet connects many networks together. Why then are so many networks homogenous across the spectrum of the Internet. One explanation is an adaptation of Arthur's (1989) theory of increasing returns. The fundamental premise behind increasing returns is that in a competitive market environment as technologies are adopted there is cumulating increasing returns as more individuals select one technology over another. In simpler terms it is the tendency for technologies that are ahead to get further ahead, and

technologies that loose advantage, loose further advantage (Arthur 1996). The process by which a particular technology gets ahead is not always clear:

> …the economy, over time, can become locked-in by "random" historical events to a technological path that is not necessarily efficient, not possible to predict from usual knowledge of supply and demand functions, and not easy to change by standard tax or subsidy policies (Arthur 1989, p.106).

Mathematically Bianconi and Barabasi (2001) describe these winner-take-all networks. They provide a variation of the scale free model that includes a fitness parameter in addition to the proclivity of new nodes to attach to already highly connected nodes that results in the power law descriptions described in the introduction. According to Arthur's work fitness and preferential attachment could result from a wide variety of factors, not just competitiveness of a node. This implies that winner-take-all and scale free networks might be theoretical efficient but possibly not operationally optimal, since the technology locked into the network could not be the most competitive possibility. The possibility of this phenomenon could exacerbate the increasing homogeneity of networks.

Microsoft as a National Security Threat

The statistics above stated that Microsoft comprises 97% of operating systems connected to the Internet. This near monopoly of operating systems have led to statements such as Geer's (2003) that, "The identicality and flaw density in the Microsoft

Windows monoculture present clear dangers to national security" (p.14). Further Geer delineates that cascading attacks that spread from one user to another across a network by infecting the same computing platform as the greatest risk. Lastly, "the only answer to the problem is platform diversity" (Geer 2003, p.17). These are all strong statements and call for quantitative analysis and a methodological approach to analyze appropriately.

Methodology

The methodology of the approach for this analysis begins with a simulated scale free network of 12,000 nodes, in which all nodes a homogenous. Next, 1% of these nodes are randomly changed into a different node type (vendor, application, etc.) and the network is then heterogeneous. At this point a malicious attack will be introduced to the population that only affects the new node type, at this step, that is only 1% of the population. After the malicious attack has spread through the susceptible population the number of disconnected nodes and the number of nodes with only one connection is calculated. Next, 2% of the nodes are randomly seeded as the new node population and the process is repeated. This process is duplicated at 1% intervals until all the nodes in the network consist of the new node type. The results of this procedure are illustrated in figure 22.

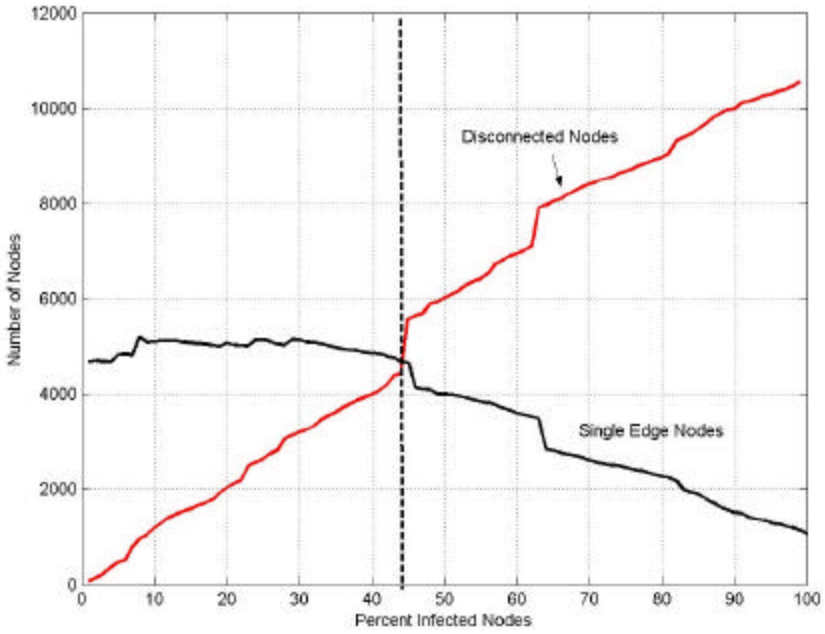

Figure 22. Number of Disconnected Nodes Versus Size of Susceptible Population with Random Seeding

Results

The number of disconnected nodes in the network increases linearly until 43% of nodes are susceptible at which time there is an almost vertical jump in disconnected nodes. The large jump in disconnected nodes coincides with the crossover with number of single edge nodes that experiences a corresponding steep drop. The combination of steep numerical shift and cross over is indicative of a catastrophic failure in the network, since nearly half of the nodes in the network are now disconnected and incapable of communication. When 43% of the population becomes any single one node type a single

malicious attach can cause traumatic damage to the total network. At the 43% threshold there are more disconnected nodes than single edge nodes and large parts of the networks can no longer communicate with each other. The same break point at 43% can be observed when the number of disconnected edges is plotted as shown in figure 23, further supporting the rejection of the hypothesis.

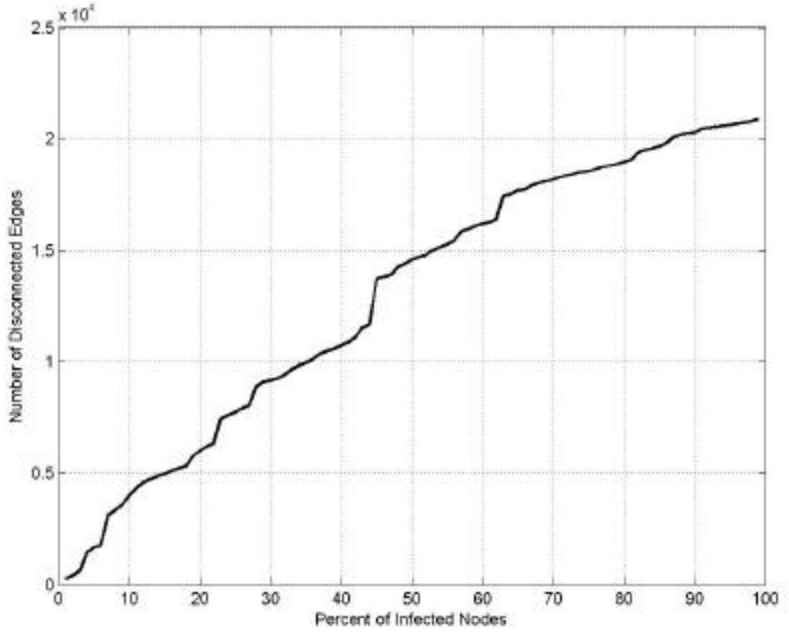

Figure 23. Number of Disconnected Edges Versus Size of Susceptible Population for Random Seeding

Another way to view the degradation of the network is by examining the data by segregating nodes by their number of edges. This was done by constructing a histogram with midpoints at 25, 50, 100, 250, 500 edges per node. The results of this approach are presented in figure 24.

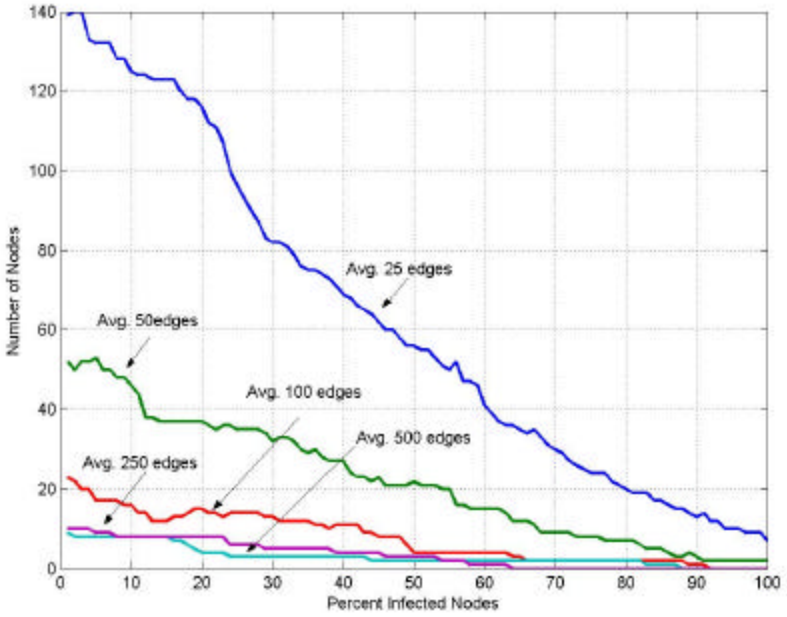

Figure 24. Node Failures Segregated by Number of Edges for Random Seeding

The histogram approach illustrates that nodes with 25 average edges decrease most rapidly overall, while nodes with an average of 500 edges experience a sharp drop after 20% of nodes are infected. The effect on nodes with a large number of edges is particularly important because of the heavy dependence on the nodes to connect the poorly connected nodes that constitute the majority of the network (Albert and Barabasi 2002). The results suggest that nodes with higher connectivity are more resilient to increasingly susceptible node populations.

Thus far the local properties of the network have been examined, the number of disconnected nodes and edges in the network. As the nodes or edges are removed, the connectivity of the network changes, and this is not accounted for in just the number of nodes and edges that are disconnected.

One way to find the integrity of a network is to measure its diameter, as the nodes or edges are removed the diameter of a network changes. The task of computing changing diameter of a network is computationally expensive. This problem is even more acute for sparsely connected networks consisting of a very large number of nodes, such as scale free networks. Recently researchers have used spectral analysis to study the dynamic behavior of networks (Biggs, 1993; Alon, 1986). Some researchers (Seary and Richards, 1997) have used discrete Laplacian to compute the steepest descent or gradient to measure the global connectivity of networks. When either nodes or edges are disconnected, the graph shows the steepest gradient in the direction where there is maximum number of nodes/edge removals. A global property that will capture this behavior is the eigenvalue of a network. Using an adjacency matrix, one may compute the discrete Laplacian as follows: Let **A** represent an adjacency matrix of a graph G. Let **D** represent the diagonal matrix formed from node degree of graph G,. i.e. **D** = diag(**C**), where, diag refers to diagonal matrix operation on **C**, the column matrix of node degree of computed from matrix **A**. Then the discrete Laplacian **L** is given by: **L** = **D** – **A**. Computing eigenvalues of **L** and studying changes in the eigenvalues of the discrete Laplacian will indicate the dynamic behavior of network G. For random removal of nodes and or edges, the changes in eigenvalues may be used as a substitute for diameter of sparse networks with very large number of nodes. One such attempt is shown in figure 25. Incidentally the jump in the eigenvalue matches with the crossover point at 43% shown in figure 22. This result indicates that a global failure as well as local failure occurs when 43% of the nodes are susceptible. This finding reinforces 47% as a point of

catastrophic failure since at this point not only are nodes locally disconnecting but also globally, not allowing communication across the network to its far-flung constituents.

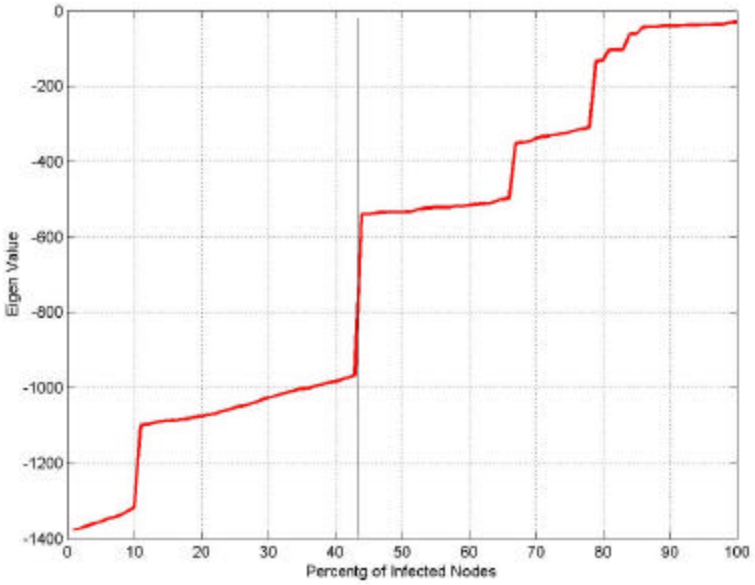

Figure 25: Eigenvalues Versus Size of Susceptible Population with Random Seeding

Targeted Seeding of Heterogeneous Networks

Another variation on this procedure is to use a targeted seeding of the prey population instead of a random seeding. It is important to include this approach since it provides a static comparison to the random approach, which will vary slightly every time it is run. Specifically starting with the most connected node as the first member of the new species, and the second most connected, third most connected, and so on until the least connected node is turned into the new species. The number of disconnected nodes and the number of single edge nodes is calculated for every 1% interval. The results of this targeted approach are shown in figure 26.

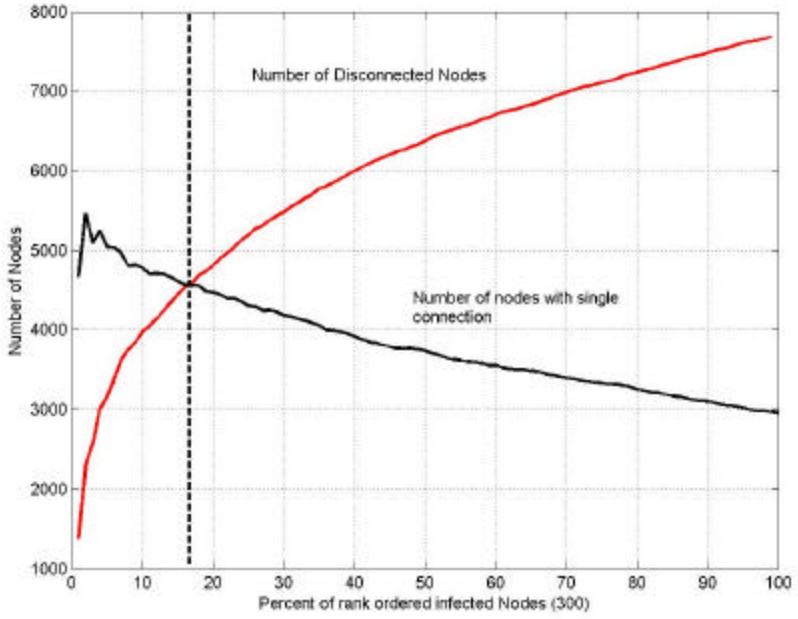

Figure 26. Number of Disconnected Nodes Versus Size of Susceptible Population with Targeted Seeding

The results of the targeted approach illustrate a far more rapid degradation of the network, i.e. an exponential increase in the number of the disconnected nodes versus the linear degradation seen in the random seeding. Also the crossover point with the number of single edge nodes occurs far earlier around 17%, instead of 43%, coinciding with the inflection point of the disconnected nodes curve. While the simulation is not realistic it does demonstrate the point that when a node type clusters amongst the more connected nodes in a network it dramatically increases the vulnerability of that network. The corollary to this result would be the more diversity in the core of the network the more robust the network. Viewing the node degree histogram plot used with the random

seeding approach reinforces this corollary. The same distribution of average connectivity at 25, 50, 100, 250, 500 midpoints was again used, and is displayed in figure 27.

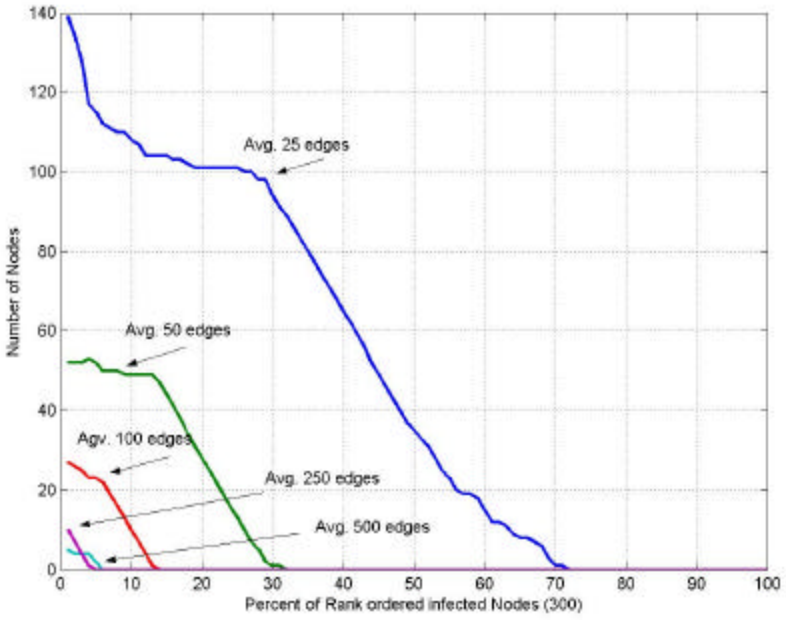

Figure 27. Node Failures Segregated by Number of Edges for Targeted Seeding

The shape of disconnected node arc for the 25, 50, and 100 edges are all very similar, while the 250 and 500 degrades down to zero almost immediately. The rapid failures of highly connected nodes are directly correlated with the targeted seeding strategy. In the targeted seeding strategy the most connected nodes are made susceptible first, so their rapid failure would be expected. The rapid failure of the network in general is linked to the failure of the most connected nodes first. This is reinforced when the number of disconnected edges is presented, as presented in figure 28.

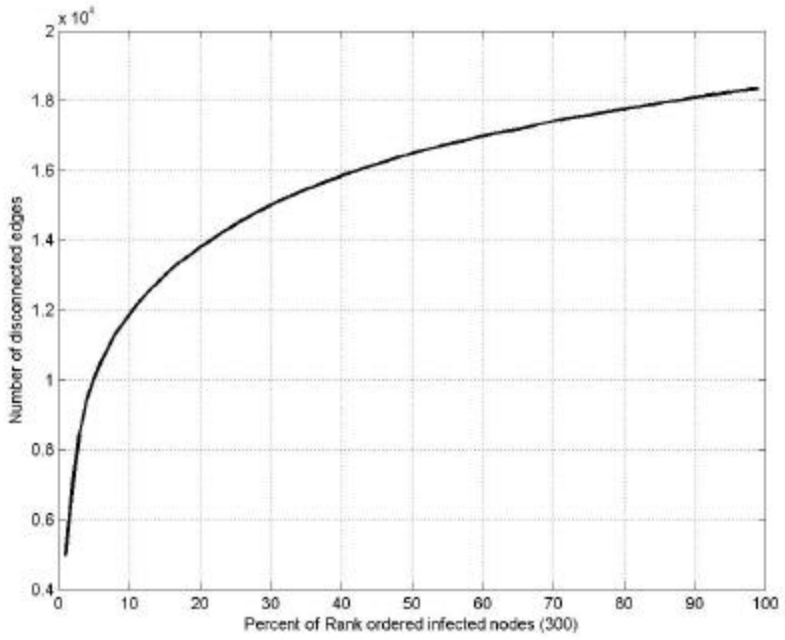

Figure 28. Number of Disconnected Edges Versus Size of Susceptible Population for Targeted Seeding

The targeted seeding results in an exponential loss of edges in the network, as opposed to the linear loss seen in the random scenario. This finding reinforces that the greater homogeneity in a highly connected core the more vulnerable a network is.

Implications for the Monoculture Hypothesis

The results of the simulations point out that at certain thresholds, in this case 47%, the pervasiveness of a single platform can lead to catastrophic network failures. Further when homogenous platforms are concentrated in the core of a network the threshold to failure decreases precipitously. Both findings are important but must be placed in context to answer the questions stated at the outset of this research. The first caveat that must be

made is that while vendors can have a monopoly in their market segment the Internet itself will never be completely homogenous. Even Microsoft's near complete dominance of the end user operating systems is only one part of the Internet. End user systems must still connect to routers to guide their traffic and servers to obtain their requested data. Thus, even though Microsoft has a 97% market share on operating systems connected to the Internet it dies not have a 97% market share of the Internet as whole. The Internet by its nature connects a wide variety of applications and platforms together in a mostly seamless web. As a result these simulations are crude, they do no delineate nodes as being routers, servers, or end user operating systems. This fact, though, can inform how one interprets the results from these simulations. As stated previously the Internet at multiple levels matches the topology used in this simulation, with that in mind the results can be placed in context.

The first question posed by this research was, do technology monocultures produce security threats and vulnerabilities. In a macro sense the results produced by this research support the monoculture hypothesis. When over 47% of the network consisted of a single node type (culture) that had a vulnerable exploit the network was susceptible to catastrophic failures. In this case catastrophic failure means that large parts of the network can no longer communicate with each other because of balkanization. Further when homogeneity is preferentially introduced into the most connected components at the core of the network catastrophic failures occur more quickly. This insight allows an interesting perspective on the question of Microsoft's market dominance leading to a national security threat, the second question posed by this research.

Microsoft's near monopoly is in end user operating systems and there has been a long litany of worms and virus's that have exploited vulnerabilities in the system. Many of these have been the cascading failures described by Geer (2003) that spread from user to user implementing identical operating platforms. Despite 97% of these platforms running Microsoft Windows operating systems there has never been a catastrophic failure of the Internet as a result of any worm or virus. Catastrophic failure meaning that none of these malicious attacks have resulted in a balkanization of the network where large chunks of the network could no longer communicate with each other or a complete shut down of the Internet. Why then the discrepancy with the numbers seen in this study illustrating a catastrophic failure at 47%.

The answer lies in the connectivity of end user operating systems. Machines running Microsoft Windows are almost entirely implemented by end users who rarely have more than one connection to their machine. Microsoft Windows machines are at the edge of the network and not in the core of the network. No matter how many Windows operating systems are infected or fail the core of the network will still run even if there is no one left to actually send traffic.

Windows vulnerabilities, though, are not the only Microsoft products to be the target of malicious attack. NIMDA and Code Red both attacked Microsoft IIS servers and the Sapphire worm attacked Microsoft SQL server machines. In all three instances there was much damage done but none resulted in a catastrophic failure to the Internet. If

the path of logic presented above is followed it is because Microsoft total share of the server market has varied between only 20-23%, not close to the 47% that would induce a catastrophic failure (Onestat 2003). Thus if catastrophic failure of the network is the threshold by which national security threats are defined Microsoft would not qualify, simply because their monoculture is not at the core of the network. Other definitions of what constitutes national security could lead to other conclusions. Further, there is no doubt that the number of hijacked Microsoft windows machines that have resulted from vulnerability exploits pose a serious threat as a means of attack on core critical infrastructure.

These conclusions point to the inevitable question are there monocultures at the core of the network. Current reports state that 86.5% of the router market belongs to Cisco Systems and 60% of core Internet routers are Cisco (Collins 2002). Cisco has had its own problem with vulnerabilities and security flaws, most notably the release of code by a group of Italian teenagers to exploit nine known Cisco vulnerabilities[1] (Reardon 2004). Cisco also had a new vulnerability found in their carrier class core routers in 2003 (Reardon 2004). While these vulnerabilities are problematic they are fortunately not prone to what Geer labels "cascading" vulnerabilities. Perhaps a more accurate term is self-replicating, but the premise is the same, Cisco vulnerabilities do not allow the spread of an exploit from machine to the next. Typically Cisco vulnerabilities are exploited through denial of service attacks that take down a predetermined node or subset of nodes, but do not spread throughout the network. As a result no Cisco vulnerabilities have

[1] It should be stated that the nine vulnerabilities were pre-existing and Cisco had previously addressed them with software upgrades and work arounds.

resulted in catastrophic failures even though the numbers found in this research point to the theoretical possibility.

Conclusions and Policy Implications

The findings of this research illustrate the security problems of reliance on a single network application. When 43% of nodes in a network were susceptible to attack the results indicated the presence of a catastrophic event from the failure of those nodes. When susceptible nodes were targeted in the highly connected core of the network the threshold for a catastrophic failure was even lower, only 17%.

The results should not been viewed as hard numbers, by which decisions can be based. The approach is useful when questions arise if the market share of a particular company in a networked environment is causing possible security vulnerabilities on a macro scale. Results could vary widely by network topology, type of attack, and vulnerability. The particular network used for these simulations is indicative of a large number of critical global infrastructures, including the Internet at the router and autonomous system level and the World Wide Web. While the specific cases of Microsoft and Cisco outlined in this analysis do not point to immediate catastrophic threats the potential is there. As such it is useful to examine what role public policy could play in remedying such vulnerabilities.

The similarity of the network tested in this research to those critical infrastructures raises the question if monopoly or quasi monopoly conditions can contribute to national security vulnerabilities. Further, if such vulnerabilities bear out in the future one must ask, “Is antitrust policy an appropriate response?”. Antitrust regulation falls under the Sherman Act and specifically, the Supreme Court stated in its *Professional Engineers Case*, 435 U.s. 679, 695 (1978):

> The Sherman Act reflects a legislative judgment that ultimately competition will produce not only lower prices, but also better goods and services. "The heart of our national economic policy long has been faith in the value of competition." Standard Oil Co. v. FTC, 340 U.S. 231, 248. The assumption that competition is the best method of allocating resources in a free market recognizes that all elements of a bargain - quality, service, safety, and durability - and not just the immediate cost, are favorably affected by the free opportunity to select among alternative offers.

In this case the safety clause would be the part of legislation that specifically applies to network-based applications. When there is a lack of competition and a resulting lack of diversity this research points to a pronounced negative safety externality. The interdependent and interconnected nature of network-based applications exacerbates the safety issue by affecting constituents who do not use the product or application. In the case of this research those constituencies are represented by the disconnected nodes and were the non-susceptible nodes in each simulation.

The problem of a growing lack of diversity in networks can be exacerbated by technology lock-in, "A customer experiences "lock-in" when switching costs exceed the potential incremental value of alternative suppliers' products over its current supplier's product (Lookabough and Sicker 2003)". Lock-in has been a common technology strategy in recent years directed towards growing market share and producing stable revenue growth. When a lock-in strategy aggregates across any one platform the result can be a non-optimal market creating interdependent security vulnerabilities and negative externalities. Lookabough and Sicker (2003) state that, "security induced lock-in has resulted in convergence to a stable equilibrium that is not the globally optimal one" and they point towards antitrust policy as one possible policy remedy.

Utilizing national security as a basis to enforce antitrust policy to promote competition has historical grounding. During World War II "national security was a potent weapon in the battle for both public investment and antitrust enforcement (Hart 1998)." Interestingly US policy to date has gone in the opposite direction with legislative efforts. The Critical Infrastructure Security Act of 2001 (S 1456 IS) states that, "antitrust laws inhibit some companies from partnering with other industry members, including competitors, to develop cooperative infrastructure security strategies." While closer cooperation among industry actors is vital it is important not to shelve antitrust policy options in the process. This research highlights the benefits of diversity in increasing the robustness of a network and there is a direct link between competition and increased diversity. Antitrust is one policy tool that has been successfully used by the US

government to promote national security in the past and it is possible that is could serve that purpose again.

While antitrust has a precedent for use, it is most likely perceived as an unpopular heavy-handed approach to the problem, as seen in the legislation previously cited. Other more subtle approaches could be the use of Federal procurement policy to encourage or mandate minimal levels of diversity in government networks and computing environments. Spreading this type of initiative to the private sector could include Federal research and development funding for firms developing technology to compete with current dominant technologies. Other innovative policy solutions are possible that could increase competition in critical technology areas effectively increasing diversity across critical infrastructures.

Works Cited

Albert R. and Barabási, A., 2002, Statistical mechanics of complex networks, *Reviews of Modern Physics* 74: 47-97.

Albert R., Jeong H., and Barabási A.L., 2000, Attack and error tolerance in complex networks. *Nature* 406: 378.

Alon N, 1986, "Eigenvalues and Expanders" Combinatrica 6:2, 73-88.

Arthur B, 1989, "Competing Technologies, Increasing Returns and Lock-in by Historical Events," *Economic Journal,* 99, 106-131

Arthur B, 1996, "Increasing Returns and the New World of Business" *Harvard Business Review,* July-Aug

Arthur WB, 1999, "Complexity and the economy" *Science* 284: 107-109

Aucsmith D, 2003, "Diversity has a cost" *IEEE Security & Policy* 3(3): 17

Barabasi A, Albert A, 1999, "Emergence of scaling in random networks" *Science* Oct: 509-512

Bianconi G, Barabasi AL, 2001, "Bose-Einstein condensation in complex networks" *Physical Review Letters* 86(24): 5632-5635

Biggs N, 1993, *Algebraic Graph Theory* New York: Cambridge University Press.

Collins J, 2002, "Dominant Cisco grows router market share" *Personal Computer World* http://www.pcw.co.uk/News/1131853

Dezsos, Z., Barabasi, A.L., 2002, *Halting viruses in scale-free networks.* Physical Review E 65: 055103 (R).

Faloutsos C, Faloutsos P, Faloutsos M, 1999, "On power-law relationships of the Internet Topology" *Computer Communication Review* 29: 251-260

Geer DE, 2003, "Monopoly considered harmful" *IEEE Security & Policy* 3(3): 14-17

Hart DM, 1998, *Forged Consensus: Science, Technology, and Economic Policy in the United States, 1921-1953,* Princeton, N.J.: Princeton University Press

Huberman B, Adamic L, 1999, "Growth dynamics of the World Wide Web" *Nature* 401:131-134

Lookabough T, Sicker DC, 2003, "Security and Lock-in" Economics and Information Security Workshop, University of Maryland – http://www.cpppe.umd.edu/rhsmith3/papers/Final_session8_lookabaugh.sicker.pdf

Onestat.com 2003, http://www.onestat.com/html/aboutus_pressbox24.html

Pastor-Satorras, R., and Vespignani, A., 2002, Immunization of complex networks. *Physical Review* E 65: 036104-1.

Pastor-Satorras, R., Vespignani, A., 2001, Epidemic dynamics and endemic states in complex networks. *Physical Review* E 63: 066117.

Reardon M, 2004, "Code attacks Cisco vulnerabilities" http://www.msnbc.msn.com/id/4631851/

Seary A, Richards, W, 1997, "The Physics of Networks" INSNA Sunbelt XVII, San Diego, February 13-17, 1997

Spencer J, and Sacks L, 2003, "On Power-Laws in SDH Transport Networks" IEEE ICC 2003, May 2003, Anchorage, Alaska, USA