The Impact of Network Topology on the Spread of Anti-Virus Countermeasures

Li-Chiou Chen
lichiou@andrew.cmu.edu
Kathleen M. Carley
KathleenCarley@cmu.edu

Carnegie Mellon University

Abstract

Are there properties of networks that affect the spread of anti-virus countermeasures? Is countermeasure spreading more effective through one network than through another? In this paper, we investigate this problem by simulating the impact of network topology on the effectiveness of anti-virus countermeasures. We simulate the spread of countermeasures based on the idea that computer viruses and countermeasures spread through two separate complex networks -- the virus-spreading network and the countermeasure-spreading network. This problem can be though of having countermeasures act as competing species against computer viruses. We find that certain properties of networks determine if countermeasures can spread faster than computer viruses and influence the size of the virus infection.

Contact:
Li-Chiou Chen
Dept. of Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, PA 15213

Tel: 1-412-268-5550
Fax: 1-412-268-6938
Email: lichiou@andrew.cmu.edu

Key Words: Computer viruses, network topology, spread of epidemics, computer simulation.

Support: This work was supported in part by the NSF/ITR and the Pennsylvania Infrastructure Technology Alliance, a partnership of Carnegie Mellon, Lehigh University, and the Commonwealth of Pennsylvania's Department of Economic and Community Development. Additional support was provided by ICES (the Institute for Complex Engineered Systems) and CASOS – the center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University (http://www.casos.ece.cmu.edu). The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation, the Commonwealth of Pennsylvania or the U.S. government.

The Impact of Network Topology on the Spread of Anti-Virus Countermeasures

Li-Chiou Chen and Kathleen M. Carley

INTRODUCTION

Studies [6][7][8][9] have shown that the spread of epidemics and the spread of computer viruses are dramatically affected by the topology of the underlying networks that the epidemics/viruses utilize to spread. Based on these studies, we are interested in a further problem. What are the properties of networks that influence the spread of anti-virus countermeasures while computer viruses are spreading? In this paper, we investigate this problem by simulating the impact of network topology on the effectiveness of anti-virus countermeasures.

The spread of computer viruses is a non-linear dynamic system which is similar to the spread of epidemics in human populations [5][9]. The Susceptible-Infected-Removed (SIR) model has been widely used to model the spread of epidemics and to study immunization strategies [2]. The SIR model is a "population-level" description of the epidemic diffusion process. The problem with the SIR model is that it only describes the state changes of the population over time. Implicitly it assumes that the population is well-mixed. Namely, everyone is connected to everyone else. This is usually not the case in either human or computer networks. Hence, the SIR model requires increasing the number of model variables to account for variations in network structure.

We propose an anti-virus strategy called Countermeasure Competing (CMC). The detail description of the model for CMC is in [4]. Since this paper focuses on the impact of network topology, we only describe the hypothesis that CMC is built upon. By revising the SIR model, CMC considers both the network for spreading viruses (the virus-spreading network, G_c) and the network for spreading countermeasures (the countermeasure-spreading network, G_v). CMC is based on the hypothesis that countermeasures against computer viruses spread through a countermeasure-spreading network. The spread of countermeasures is similar to the spread of computer viruses, but, unlike computer viruses propagate themselves, countermeasures act to suppress the spread of computer viruses. This can be thought as having two viruses spreading at the same time: a good one and a bad one. The properties of the networks that influence the spread of a good one over a bad one enable the overall system to be less likely to be affected by the bad one. In the real world, both the virus-spreading network and the countermeasure-spreading network can represent either physical networks (connecting computers/programs) or social networks (connecting people/groups). Whether each of them is a social network or a physical network depends on the vulnerability/information that the virus/countermeasure utilizes in order to spread.

THE SIMULATION FOR THE SPREAD OF COUNTERMEASURES

The simulation is designed to be flexible enough so that it can examine the effectiveness of CMC by varying the countermeasure-spreading network using Monte-Carlo sampling techniques. Using computer simulation, we conduct virtual experiments to simulate CMC using six different network topologies. These six network topologies include two empirical networks and four theoretical networks. The two empirical ones are the TWL network, which has 106 nodes and is inferred from the TWL virus reporting records [4], and an Internet autonomous system network topology[1], which has 11,716 nodes. The four theoretical ones include a scale-free network[2], a lattice[3], a random network, and a fully connected network.

For each virtual experiment, we fix G_v to be one of the empirical networks and vary G_c to be either the

[1] Available at "http://moat.nlanr.net/AS/ ", downloaded on August 2001.

[2] All scale-free networks are generated based on the algorithm in [3].

[3] We use the Small-World network algorithm in [26] to generate the lattice (with reconnecting probability=0), and the random network (with the reconnecting probability =1).

empirical one or one of the four theoretical networks. Both G_v and G_c in a same experiment have the same number of nodes because we assume that each node in G_v maps to one node in G_c. The probability that each node would adopt the countermeasures received is set to 0.1. The performance of CMC is measured in the relative size of the virus infection (*RS*). *RS* is calculated as the size of the virus infection[4] based on CMC relative to the size of the virus infection without any anti-virus strategy. That is, CMC is more effective if *RS* is smaller. *RS* for each experiment is calculated as the average of 10^5 runs. We then correlate *RS* with various network measures. For each countermeasure-spreading network, we calculate five network measures: epidemic threshold[5][8], density[6], average path length[7], and clustering coefficient[8][11] and degree centralization[9][10].

THE IMPACT OF NETWORK TOPOLOGY

Table I: Correlations between the network measures for countermeasure-spreading networks (G_c) and the relative size of the virus infection (*RS*)

	The ratio of countermeasure-spreading rate to virus-spreading rate						
	0	0.5	1	2	4	6	12
Epidemic threshold	0	0.65	0.84	0.93	0.94	0.92	0.92
Density	0	-0.98	-0.86	-0.71	-0.58	-0.51	-0.49
Average path length	0	0.24	0.30	0.36	0.47	0.56	0.64
Clustering coefficient	0	-0.83	-0.82	-0.68	-0.52	-0.42	-0.36
Degree centralization	0	-0.75	-0.55	-0.25	-0.18	-0.18	-0.22

Among the measures we calculate, epidemic threshold has the highest positive correlation to *RS* when the countermeasure-spreading rate[10] is larger than the virus-spreading rate[11]. Epidemic threshold is defined as the minimal epidemic spreading rate that an epidemic can prevail [1]. In a complex network, epidemic threshold varies with the edge distribution of networks. Applying this property on countermeasure spreading, we find that the countermeasure-spreading network with a lower epidemic threshold is more effective to reduce the size of the virus infection than the ones with higher epidemic thresholds. In addition, density has a negative correlation with *RS*. This result implies that CMC is more effective if the connectivity of G_c is larger. Moreover, the effectiveness of CMC increases with clustering coefficient (negatively correlated to *RS*), and decreases with average path length (positively correlated to *RS*). This result implies that countermeasures spread faster when the cliquishness of G_c increases, and they spread slower when average path length increases. This result confirms the finding in [11] about epidemic spreading on a network with the Small-World property. Finally, we find that the effectiveness of CMC increases when the degree centralization of a network increases. However, the correlation is smaller

[4] The term is from "the size of the epidemic", which are commonly used in epidemiological literature. "The size of the virus infection" here refers to the fraction of nodes which have infected a computer virus over time.

[5] When an epidemic spreads on a complex network, the epidemic threshold can be estimated by $\rho_{threshold} = \frac{< e >}{< e^2 >}$ where <e> denotes the average number of edges and <e^2> denotes the average square of edges [8].

[6] Density measures the connectivity of a network, which is defined as the number of edges of a network divided by the largest possible number of edges of this network [10].

[7] Average path length is defined as the average of the shortest path length between any two nodes in a network.

[8] Clustering coefficient measures the cliquishness of a network. Node clustering coefficient is defined as the connectivity of the neighbors of a node. Clustering coefficient is the average of node clustering coefficients in a network [11].

[9] Degree centralization measures the differences of the connectivity among nodes, which takes the average of the difference of individual node connectivity and the average node connectivity [10].

[10] The countermeasure-spreading rate represents the probability that a countermeasure spreads from one node that has received countermeasures to a neighboring node during each time period.

[11] The virus-spreading rate represents the probability that a virus spreads from one node that has infected by a virus to a neighboring node during each time period.

comparing to other properties.

In summary, we find that CMC is more effective when the countermeasure-spreading network is highly connected (as such FULL) or highly centralized (with a higher epidemic threshold, a higher cliquishness, or a shorter average path length), in spite of the low probability of adopting countermeasures (=0.1). The influence of the network topology on the effectiveness of countermeasure spreading is two folds. If G_c and G_v have a same property, this property influences the spread of viruses as the same way as the spread of countermeasures. In this case, the effectiveness of countermeasure spreading increases when the ratio of the countermeasure-spreading rate to the virus-spreading rate increases. However, if the properties of G_c are different from those of G_v, the effectiveness depends on the ratio between the properties of these two networks. Hence, to suppress the spread of computer viruses, G_c needs to have the properties that are able to spread countermeasures faster than viruses.

CONCLUSIONS

We find that certain properties of networks determine if countermeasures can spread faster than computer viruses and influence the size of the virus infection. For example, a network with a lower epidemic threshold has this property. A network that has a few nodes with high connectivity also has this property. Based on this result, it will be effective to spread countermeasures on a network of major response centers or anti-virus companies to their large customer base even when the probability of decision makers to adopt countermeasures is only slightly larger than 0.1. Future work could be done based on our study. Our study simulates the impact of network topology on the spread of countermeasures and viruses through two separate complex networks. The same idea can be applied to other problems where network topology plays a role in two competing contagious agents, such as the effect of spreading rumors on the diffusion of correct information.

REFERENCES

[1] R. M. Anderson and R. M. May, *Infectious Diseases in Humans*: Oxford University Press, 1992.

[2] N. J. T. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, 2nd ed. New York: Oxford University Press, 1975.

[3] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, pp. 509-512, 1999.

[4] L.-C. Chen and K. M. Carley. "The Impact of Countermeasure Spreading on the Propagation of Computer Viruses," unpublished working paper, Carnegie Mellon University, 2003.

[5] J. O. Kephart and S. R. White, "Measuring and Modeling Computer Virus Prevalence," presented at IEEE Computer Security Symposium on research in Security and Privacy, Oakland, California, 1993.

[6] J. O. Kephart, "How Topology Affects Population Dynamics," in *Artificial Life III*, C. G. Langton, Ed. Reading, MA: Addison-Wesley, 1994.

[7] R. M. May and A. L. Lloyd, "Infection Dynamics on Scale-free Networks," *Physical Review E*, vol. 64, 2001.

[8] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, "Epidemic Outbreaks in Complex Heterogeneous Networks," *The European Physical Journal B*, pp. 521-529, 2002.

[9] R. Pastor-Satorras and A. Vespignani, "Epidemics and Immunization in Scale-free Networks," in *Handbook of Graphs and Networks: From the Genome to the Internet*, S. B. a. H. G. Schuster, Ed. Berlin: Wiley-VCH, 2002.

[10] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.

[11] D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small-World' Networks," Nature, vol. 393, 1998.