Locating Defense Positions for Thwarting the Propagation of Topological Worms

Sheng Wen, Wei Zhou, Yini Wang, Wanlei Zhou, *Senior Member, IEEE,* and Yang Xiang, *Member, IEEE*

***Abstract*—A common view for the preferable positions of thwarting worm propagation is at the highly connected nodes. However, in certain conditions, such as when some popular users (highly connected nodes in the network) have more vigilance on the malicious codes, this may not always be the truth. In this letter, we propose a measure of betweenness and closeness to locate the most suitable positions for slowing down the worm propagation. This work provides practical values to the defense of topological worms.**

***Index Terms*—Network security, worms, propagation.**

I. Introduction

TOPOLOGICAL worms, such as email worms and social network worms, pose a critical security threats to the Internet [1]. Firstly, they rely on the information contained in the victim machine to locate new targets. This intelligent mechanism allows for a far more ***efficient*** propagation than scanning worms that make a large number of wild guesses for every successful infection. Instead, they can infect on most attempts and thus, achieve a ***rapid*** spreading speed. Secondly, by using social engineering techniques on modern topological worms, most internet users can possibly fail to recognize malicious codes and get infected, which results in a ***wide*** range of propagation.

In order to eradicate topological worms, as well as to control and limit the impact of their outbreak, previous works [2] [3] [4] presented certain strategies to immunize a group of users in the network to prevent topological worms from propagating to a large scale. However, how to choose the appropriate size and membership of this subset to constrain topological worm spreading remains a difficult question. A common view for the preferable positions of defense is at the highly-connected users [2] [3] or those with most active neighbors [4]. Indeed, popular users in a scale-free network and their intuitively short paths to other nodes in a strongly clustered small world [5] [6] greatly facilitate the propagation of an infection over the whole network, particularly at their early stage. However, is the viewpoint always true for slowing down the topological worm propagation by quarantining those positions?

Considering the network sketched in Fig. 1(a), two large groups (group 1 and 2) are bridged by connections among just a few members. In this case, nodes A, B, C and D have a larger degree of edges than node E. If they have low probability to infect each other directly, node E will become a key point for propagating worms from one group to another. Consequently, E has a high priority for being quarantined even if it has a lower degree. In Fig. 1(b), node A and B have fewer connections than node C that has the largest degree. However, A and B have the shortest paths to all the others (hops in this case) and can access all the nodes in the network more quickly than anyone else, so they are more attractive to attackers for spreading topological worms. As a result, quarantining a highly-connected node may not be a prior selection for preventing worm propagation.

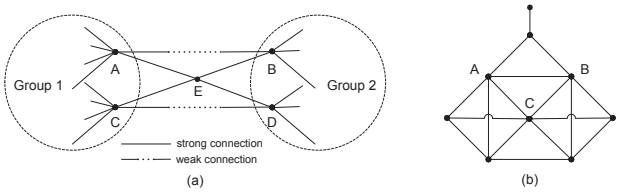

Fig. 1. Exceptional topologies. (a) Node E controls the propagation between group 1 and group 2, so it is more important than node A, B, C and D; (b) Node A and B have smaller length of average short paths to every node, so they are more attractive to attackers than node C is.

This letter shows our new finding on locating the preferable defense positions for restraining the propagation of topological worms. We find that the preferable positions may not always be at those popular users if they are more watchful on the malicious information. Validation based on simulation studies reveals that the proposed measure can help find the most suitable positions for defense.

II. Representation of Worm Propagation

To evaluate the effect on thwarting worm propagation by quarantining certain nodes, we propose a model to represent the spreading procedure of topological worms. Firstly, an infectious node can propagate worms, and a susceptible user can be infected and become a new infectious node. This means topological worms spread from node *i* to node *j* via one or more intermediate nodes. We assume *k* to be the number of spreading hops between node *i* and node *j*, and use $X_i(k)$ to represent the infection state of node *i* after *k* hops. If node *i* is infected after *k* hops, $X_i(k) = 1$, otherwise $X_i(k) = 0$.

Secondly, we use an *n* by *n* square matrix *T* with element p_{ij} to indicate a network consisting of *N* nodes, as in

$$T = \begin{pmatrix} p_{11} & \cdots & \cdots \\ \cdots & p_{ij} & \cdots \\ \cdots & \cdots & p_{NN} \end{pmatrix} p_{ij} \in [0,1] \quad p_{ij} = 0 (i = j) \qquad (1)$$

If node *i* is susceptible, it can be compromised by any of its infected neighbors. The element p_{ij} in matrix *T* denotes the

Manuscript received December 3, 2011. The associate editor coordinating the review of this letter and approving it for publication was C. Mitchell.

S. Wen and W. Zhou are with the School of Information Science and Engineering, Central South University, Changsha, P.R. China, 410083. They are currently visiting the School of Information Technology, Deakin University, 3125, Melbourne, VIC, Australia (e-mail: {wsheng, weiz}@deakin.edu.au).

Y. Wang, W. L. Zhou, and Y. Xiang are with the School of Information Technology, Deakin University, 3125, Melbourne, VIC, Australia (e-mail: {yiniwang, wanlei, yang}@deakin.edu.au).

Digital Object Identifier 10.1109/LCOMM.2012.030512.112452

propagation probability from node *i* to node *j*, so we have

$$p_{ij} = P\left(X_j(k) = 1 \mid X_i(k-1) = 1, X_j(k-1) = 0\right) \quad (2)$$

Thirdly, $X_i(k) = 1$ happens when either $X_i(k-1) = 1$ or $X_i(k-1) = 0$ and node *i* is infected by their neighbors (N_i) at the *k*-th hop. Therefore, we derive the infection probability of node *i* as in

$$\begin{aligned} &P(X_i(k) = 1) \\ &= P(X_i(k-1) = 1) + P(X_i(k-1) = 0)\cdot \\ &\quad P(X_i(k) = 1 | X_i(k-1) = 0) \\ &= P(X_i(k-1) = 1) + P(X_i(k-1) = 0)\cdot \\ &\quad P(X_i(k) = 1, \exists j \in N_i \ X_j(k-1) = 1 | X_i(k-1) = 0) \\ &= P(X_i(k-1) = 1) + P(X_i(k-1) = 0)\cdot \\ &\left\{1 - \prod_{j=1}^{N}\left[1 - P(X_i(k) = 1, X_j(k-1) = 1 | X_i(k-1) = 0)\right]\right\} \end{aligned} \quad (3)$$

We assume the events $X_i(k-1) = 0$ and $X_j(k-1) = 1$ are independent (see details of independence condition in [7]), so we have

$$\begin{aligned} &P(X_i(k) = 1) \\ &= P(X_i(k-1) = 1) + P(X_i(k-1) = 0)\cdot \\ &\quad \left\{1 - \prod_{j=1}^{N}\left[1 - P\big(X_i(k) = 1 | X_i(k-1) = 0, X_j(k-1) = 1\big)\cdot \right.\right. \\ &\quad \left.\left. P\big(X_j(k-1) = 1\big)\right]\right\} \\ &= P(X_i(k-1) = 1) + P(X_i(k-1) = 0)\cdot \\ &\quad \left\{1 - \prod_{j=1}^{N}\left[1 - p_{ji} P\big(X_j(k-1) = 1\big)\right]\right\} \end{aligned} \quad (4)$$

Finally, we can estimate the number of infected nodes after worm spreads *k* hops in the network *n*(*k*) by (5)

$$n(k) = E\left[\sum_{i=1}^{N} X_i(k)\right] = \sum_{i=1}^{N} E[X_i(k)] = \sum_{i=1}^{N} P(X_i(k) = 1) \quad (5)$$

We can also examine worm spreading speed by investigating $v(k) = n(k) - n(k-1)$.

III. Prior Positions For Defense

Inspired by the discussion on Fig. 1, the probable prior defense nodes should satisfy two necessary properties: 1) once the node is well protected, the number of infected nodes *n*(*k*) will be limited to a suppressed scale which is less than the case of deploying defense on other nodes; 2)the spreading speed is largely determined by the propagation scale in the early stage [8]; once the node is well protected, we expect the spreading speed $v(k)$ within a few *k* hops will be slower than the case of deploying defense in other nodes.

Correspondingly, we explore two concepts of nodes: betweenness and closeness, which have been studied in many disciplines [9]. Traditional consideration on betweenness is based on the dynamic process of starting from one node to another node, and then to another node, such as '*drunk man*' and node betweenness. The propagation of topological worms is more like a rumor diffusion process, starting from one node to *n* nodes and then to *m* nodes ($m \gg n$). In this letter, in order to analyze worm propagation, we have:

TABLE I
The Distribution of Nodes with Higher Degree, Betweenness and Lower Closeness Values

Degree	0-10	10-20	20-30	30-40	40-50	50-60	60-70
N_d	28	51	15	3	2	0	1
N_b	92	5	2	1	0	0	0
N_c	94	2	2	1	0	0	1

N_d: the number of nodes with higher degrees in each degree range
N_b: the number of nodes with higher betweenness values in each degree range
N_c: the number of nodes with lower closeness values in each degree range

Definition 1. *the betweenness $b_i(k)$ of a node i is equal to the probability that a worm spreading procedure starting at s and ending at t passes through node i along paths within k hops, averaged over all s and t.*

This measure is appropriate to a network in which worms spread at random until they infect all the nodes in the network, and it includes contributions from many paths that are not optimal in any sense. Based on this definition, we introduce *n*-order matrix $T^{(k)}$ and *n*-order betweenness vector $B^{(k)}$ as in

$$T^{(k)} = \begin{pmatrix} p_{11}^{(k)} & \cdots & \cdots \\ \cdots & p_{ij}^{(k)} & \cdots \\ \cdots & \cdots & p_{NN}^{(k)} \end{pmatrix} p_{ij}^{(k)} = 1 - \prod_{h=1}^{N}\left(1 - p_{ih}^{(k-1)} p_{hj}\right) \quad (6)$$

$$B^{(k)} = \langle b_1(k) \cdots b_i(k) \cdots b_N(k) \rangle \quad (7)$$

Each element in $T^{(k)}$ means the propagation probability from node *i* to node *j* via *k* hops and $p_{ij}^{(0)} = p_{ij}$. We use $\beta_{ij}(k)$ to denote the propagation probability from node *i* to node *j* within *k* hops ($\beta_{ij}(0) = p_{ij}$) and the betweenness can be calculated as in

$$\beta_{ij}(k) = 1 - \prod_{h=1}^{k}(1 - p_{ij}^{(h)}) \quad (8)$$

$$b_i(k) = \frac{1}{(k+1)N^2} \sum_{x=0}^{k} \sum_{h=1}^{N} \sum_{r=1}^{N} \Big(\beta_{hi}(x)\beta_{ir}(k-x)\Big) \quad (9)$$

Definition 2. *the closeness $c_i(k)$ is the mean geodesic distance (propagation probability as the edge weight) between a node i and all other nodes reachable from it within k hops.*

Actually, closeness can be regarded as a measure of how long it will take worm copies to spread from a given node to others. We introduce an *n*-order closeness vector $C^{(k)}$ and $\theta_{ij}(k)$ as the probabilistic average distance between node *i* and node *j* within *k* hops. Thus, the closeness can be calculated as in

$$C^{(k)} = \langle c_1(k) \cdots c_i(k) \cdots c_N(k) \rangle \quad (10)$$

$$\theta_{ij}(k) = \sum_{h=1}^{k}(p_{ij}^{(h)} h) / \sum_{h=1}^{k} p_{ij}^{(h)} \quad (11)$$

$$c_i(k) = \frac{1}{N} \sum_{h=1}^{N} \theta_{ih}(h) \quad (12)$$

It can be observed in the definitions that the nodes with high betweenness and low closeness satisfy the necessary properties for deploying prior defense.

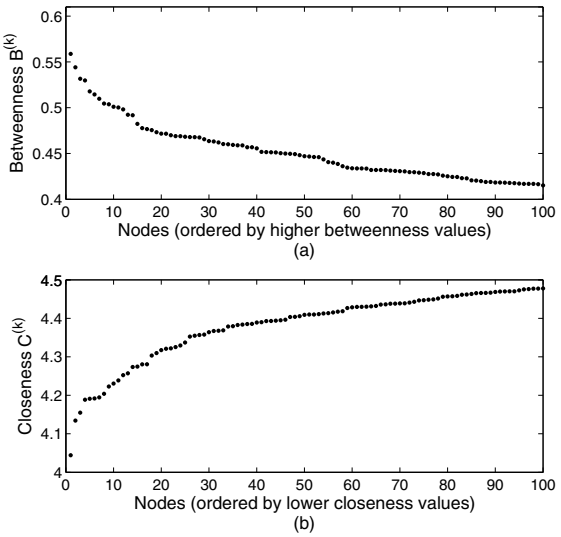

Fig. 2. Betweenness and closeness values of nodes: a) the highest 100 betweenness values; b) the lowest 100 closeness values.

IV. Evaluation

In this section, we evaluate our viewpoint for thwarting the propagation of topological worms. The implementation is in C++ and Matlab7. The random numbers in the experiments are produced by the C++ TR1 library extensions. The topology adopted for evaluation is a representative small topology (1000 nodes) based on the analysis of a realistic email network [5]. We run each simulation 100 times for average results.

To evaluate which kind of nodes should be preferentially protected, we create two different scenarios based on Internet users' vigilance of the malicious codes:

Scenario 1: some popular users are watchful on the malicious codes. The possible reason may be that these popular users have more experience on using Internet services including security issues. As an example, we assume 5% highly connected nodes have only 10% probability to be infected compared with other nodes which have an infection probability of 50%. We choose the top 100 nodes which have higher betweenness values than others in Fig. 2(a). We also compare the distribution of nodes with higher betweenness values and higher degrees. As shown in Table I, we can see that quite a number of nodes with higher betweenness values (e.g. 92 nodes have 0-10 degrees) may **not** be the nodes with higher degrees (e.g. 51 nodes have 10-20 degrees). Moreover, in Fig. 3(a), when we quarantine different kinds of nodes respectively, the defense on the nodes with higher betweenness outperforms all the other strategies in suppressing the propagation scale and speed. This means the nodes with higher betweenness values should be preferentially quarantined under this condition.

Scenario 2: we still assume some popular users are more watchful on the malicious codes. We choose the top 100 nodes which have lower closeness values in Fig. 2(b). We also compare the distribution of nodes with lower closeness values and higher degrees. We find in Table I that the nodes with lower closeness values also greatly **deviate** from the nodes with higher degrees and are mainly grouped in the range of lower degrees (e.g. 94 nodes have 0-10 degrees). Let the worm

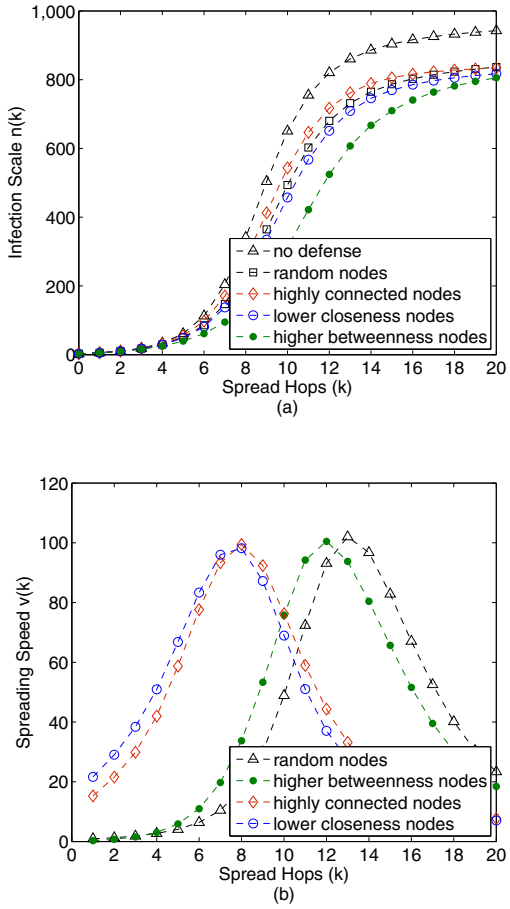

Fig. 3. Propagation scale: (a) some popular users are watchful; (b) spread starts at different positions.

propagation start at the nodes with higher degrees, higher betweenness values and lower closeness values respectively. We expect our closeness measure can help find certain nodes, from which the spreading speed is most rapid in the early stage of the worm propagation. As shown in Fig. 3(b), the nodes with lower closeness values can spread worms swiftly and are more likely to be selected by attackers for the propagation of topological worms. Therefore, this kind of nodes should be preferentially quarantined.

Through the above analysis, we have:

Remark 1. Counter-intuitively, the popular Internet users may not always be the preferable positions for thwarting the propagation of topological worms.

Remark 2. Under certain conditions, the nodes with higher betweenness or lower closeness values may be more suitable for slowing down the spread of topological worms.

V. Conclusion

This letter presents a novel investigation on locating the best positions for thwarting the propagation of topological worms. Our work can help efficiently suppress the infected scale of the network and decrease the spreading speed of topological worms. Extended work of this letter will include the estimation on the most suitable quarantine size in the network.

REFERENCES

[1] M. Fossi and J. Blackbird, "Symantec Internet security threat report 2010," Symantec Corporation, Tech. Rep., Mar. 2011.

[2] C. C. Zou, D. Towsley, and W. Gong, "Modeling and simulation study of the propagation and defense of Internet e-mail worms," *IEEE Trans. Dependable and Secure Computing*, vol. 4, no. 2, pp. 105–118, 2007.

[3] X. Fan and Y. Xiang, "Modeling the propagation of peer-to-peer worms," *Future Gener. Comput. Syst.*, vol. 26, pp. 1433–1443, 2010.

[4] G. Yan, G. Chen, S. Eidenbenz, and N. Li, "Malware propagation in online social networks: nature, dynamics, and defense implications," in *Proc. 2011 ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS'11.

[5] H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," *Phys. Rev. E*, vol. 66, no. 3, Sep. 2002.

[6] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 2007 ACM/USENIX Internet Measurement Conference*.

[7] Z. Chen and C. Ji, "Spatial-temporal modeling of malware propagation in networks." *IEEE Trans. Neural Networks*, vol. 16, pp. 1291–1303, 2005.

[8] C. C. Zou, W. Gong, D. Towsley, and L. Gao, "The monitoring and early detection of Internet worms," *IEEE/ACM Trans. Networking*, vol. 13, pp. 961–974, 2005.

[9] M. Newman, "A measure of betweenness centrality based on random walks," *Social Networks*, vol. 27, no. 1, pp. 39–54, 2005.