Bounding Virus Proliferation in P2P Networks with a Diverse-Parameter Trust Management Scheme

Lin Cai and Roberto Rojas-Cessa

***Abstract*—Peer-to-peer (P2P) networking has the potential of providing wide channels for file exchange. At the same time, P2P is prone to the proliferation of viruses. Peer trust reputation can be used to prevent virus dissemination. However, when viruses have infectious properties, peer reputation may not be enough to limit their proliferation. This letter shows the inefficiency of simple peer reputation to bound epidemics in an infectious environment and introduces a trust management algorithm that uses the combination of trust values of peers and infection values of both peers and content to bound the proliferation of viruses in P2P networks. The proposed trust management scheme can bound virus proliferation to a small number of peers without inhibiting file-download activity.**

***Index Terms*—Malware, P2P, peer-to-peer networks, trust management, virus proliferation.**

I. Introduction

In P2P networks, peers provide resources, including bandwidth, storage space, and computation power. Therefore, the potential for information distribution of P2P networks is being considered for deployment of massive applications such as IPTV [1], [2], where video sources rely on intermediate peers for further distribution of content. Pragmatic uses of information exchange in a (technology-wise) successful manner have been recently shown for file sharing [3], [4].

P2P networks promote file sharing and files are prone to carrying viruses. Therefore, this combination also facilitates virus propagation as peered users not only store the received content of the received information but also execute it. This pre-disposition to accept an unknown file leaves an open door for viruses[1]. Furthermore, Internet advertisement of popular downloads can inform users of available popular files and raise interest in those files, thus creating a complete incubation environment for viruses.

Several studies about virus proliferation have been presented [5]-[7]. They consider network topologies and features that describe the proliferation profile of viruses. Analysis of models of virus proliferation is beyond the scope of this letter. Viruses usually have specific destructive objectives, they might aim to affect the host computer, to retrieve financial information illegally, or to disrupt communications (e.g., denial of service). Depending on the characteristics, viruses in a host may or may not affect other stored files.

The authors are with the Networking Research Laboratory, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA. Email: {lc76, rojas}@njit.edu. Roberto Rojas-Cessa is the corresponding author.

[1]Any threatening code is called virus, including malware, intrusion software, or worms that aim to disrupt the communications purposes of a host or a network.

Trust management schemes are used to decrease the number of misbehaving hosts and their effect [8]-[10]. A dynamic trust management scheme based on localized trust evaluation and on a warning dissemination to prevent others from downloading a file from a suspicious peer was proposed [11]. The scheme aims to limit the proliferation of viruses using trust values on peers. Although the authors didn't assign a name to the scheme in the paper, we call this scheme Dynamic Threshold Management (DTM) in the remainder of this letter for brevity.

In general, viruses may piggyback themselves onto files stored in the infected host, infecting them. This letter shows that file infectious underscores trust management schemes based on peer reputation. This letter proposes a new trust management scheme that uses the combination of peer trust values and a warning messaging system combined with file reputation and peer infectious values. The proposed scheme is simulated under an infectious environment. The results show that the proposed scheme is efficient for epidemic control in P2P networks.

The remainder of this letter is organized as follows. Section II introduces the proposed scheme, the terms and the parameters for evaluation of peer trust, and the operation of the proposed management scheme in a P2P network. Section III shows the performance of the proposed scheme. Section IV presents our conclusions.

II. Combined Reputation-based Trust Model

In a P2P network with N peers, each peer has a file reputation table and a peer trust table. A file reputation table, which holds file identifications of known infected files in the network, is used to evaluate the infectious risk associated with a file. It is assumed that file identities remain unchanged for long periods of time and that viruses don't change intentionally the file identities. Although there are several alternatives to assign file identification (e.g., file name or file length), this is out of the scope of this letter. The trust table, which stores peer trust and infectious values, is used by the peer to select a downloading peer source. The trust table at peer i has up to $2(N-1)$ entries. The trust value at peer i about peer j, is denoted as $T_v(i,j)$, where $T_v(i,j) \in [0,1]$. A trust value at peer A about peer B represents the degree of A's expectations of downloading a clean file from B. The trust value for peer j is calculated as the total number of downloads of uninfected files from peer j over the total number of downloads. Any peer trusted by any other peer is called a trustee and any peer that trusts a trustee is called a truster. A peer has higher on its immediate trustee than on trustees of trustees. Both the distance of the trustee and the relationship type define

the term *social distance*. In general, the trust value decreases proportionally to the social distance. The second value in the trust table is the infectious value $I_v(i,j)$ and it represents the information that peer i has about the possibility of having files infected at peer j. A large infectious value means a high possibility that files are infected in peer j due to the presence of an infected file in the past. An infected download is defined as a download of a file containing a detected virus. A clean download is defined as a download of a file with no (detected) virus. When a peer downloads an infected file, other existing files in this peer can get infected with probability P_I. A peer has a virus-detection software that detects a virus with probability P_d.

The file reputation table holds the reputation value of file f_l, as $F(i,f_l)$, that has been reported to be virus carrier (or a virus itself) at peer i as reported by other peers. This value increases with each warning message.

A. Management Scheme

The trust management scheme works as follows. When peer i searches for file f_l, it checks the local file's reputation in the file record. If the file's reputation value is found at the database and is above the acceptable reputation threshold, Th_R, then the peer proceeds to find the file source.

The values held by a peer are updated after different actions take place. These are described as follows.

File Search. A peer i sends a request for f_l to all trustees whose trust value is above the admissible threshold value Th_T (i.e., trustable trustees). Peer i chooses the peer with the largest T_v and the lowest infectious value among those who have a copy of the requested file. If the file is not available from peer i's trustable trustees, the peer sends a recursive query for f_l to all trustees. In this query, the receiving trustee searches for the requested file among its own trustees. This process is performed recursively until either a fruitful search is achieved or no more trustees are available. After a recursive query, if peer k is introduced to i, new values are calculated: $T_v(i,k) = T_v(i,j)T(j,k)$, and $I_v(i,k) = I_v(i,j) + I_v(j,k)$, then peer i proceeds to the selection of a downloading source.

Post-download update. If the download of f_l is determined clean, $T_v(i,j) = \alpha T_v(i,j)$, where α is the rate of the trust value growth, $\alpha > 1$, while $I_v(i,j)$ remains unchanged. If the download of f_l is determined infected:

$$\begin{aligned} T_v(i,j) &= \delta T_v(i,j) \\ I_v(i,j) &= I_v(i,j) + 1 \\ F(i,f_l) &= F(i,f_l) + 1 \end{aligned}$$

where δ is the rate of the trust value degradation and $1 > \delta > 0$. During this phase, if $T_v(i,j) < th_w$, where th_w is the threshold to trigger a warning process, peer i issues warning messages to all its trusters. A warning message has the following format: $\{ID, v_j, f_m, \Delta, d\}$, where ID is the warning identification number, v_j is the identification of the peer that served as the source of a threatening file, f_m is the file's name, Δ is the decrement of the trust value at peer i, and d is the maximum number of truster hops the warning message is allowed to propagate.

Post-warning updates. After receiving a warning message from peer k about peer j, peer i updates the trust values. If $T_v(i,k) > Th_T$:

$$\begin{aligned} T_v(i,j) &= T_v(i,j) - \Delta T_v(i,j) \\ I_v(i,j) &= I_v(i,j) + \frac{(d-1)}{d} \\ F(i,f_l) &= F(i,f_l) + \frac{(d-1)}{d} \\ \Delta &= \Delta\frac{d-1}{d}. \end{aligned}$$

After each warning message is received by a peer, if $\Delta T_v(k,i) > th_w$, the peer sends a warning message to its trusters with updated values.

Figure 1 shows an example of the truster-trustee relationship between Peers A to F in a P2P network. The tail of an arrow indicates the trustee peer and the head indicates the truster. This example shows the social distance between Peers A and F as $d = 2$. In this example, Peer F seeks File 2, available in Peers A, B, and C. However, Peer F has no T_v and I_v values for those nodes because Peer F has not been a truster of them yet. Therefore, Peer F estimates these values via Peer E. Peer E's T_v value about Peer A is $T_v(E,A) = 0.8$ and about Peer B is $T_v(E,B) = 0.8$, but it has no T_v value about Peer C. Furthermore, Peer E has Iv values about Peer A as $I_v(E,A) = 0.7$, about Peer B as $I_v(E,B) = 0.1$, and about Peer C as $I_v(E,C) = 0.6$ (as Peer E could have received a warning about File 4 but it has not been a truster of Peer C). Peer F then selects Peer B to download File 2, as B has the lowest I_v value. In a different case, if Peer F seeks File 4, it would notice that this file is considered viral, and it would desist.

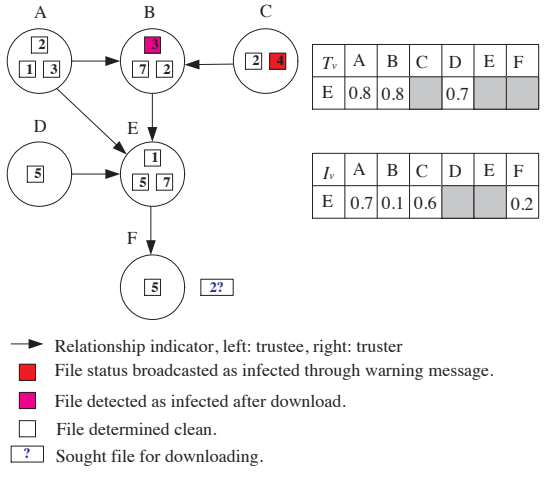

Fig. 1. Example of relationship of peers in P2P network and files.

III. Performance Analysis

A P2P network using a mesh topology with 100 nodes selected randomly as peers was simulated. This mesh represents a general topology and it can also be applied to specific P2P networks [12]. In this network, there are 150 different files, with an average of three copies each, uniformly distributed

randomly among the 100 peers. Out of these files, 60% are popular files (i.e., requested with 20% higher frequency than the other files) and 10 files are infected. The selection of popular and infected files is random (uniform distribution). The minimum time an event (e.g., a download or the trip of warning messages from one peer to another) takes is a fixed amount or a time slot. The the total number of infected peers is evaluated after each time slot, under $P_d = 0.5$.

Figure 2 shows the performance of both DTM and the proposed scheme, measured as the number of infected nodes per number of downloads in the network, under different P_I values in an infectious environment. The proposed trust management scheme uses file reputation, labeled FR in the figure, with and without I_v. Curves a) to d) show that infectious viruses inhibit the efficiency of DTM and all peers may get infected, eventually. This occurs because nodes are isolated but after viruses have infected new (and undetected) peers. Curves e) to h) show that when file reputation is used, without recurring to I_v, the number of infected peers is bounded as the number of infected files is smaller than the total number of peers in the network. The proliferation is bounded because a peer can now identify a file coming from a node with a record of no infections, in a proactive way. Curves i) to l) show how file reputation, in combination with I_v, limits virus proliferation as the number of infected peers decreases to an average of 10 for these curves. The warning messages are also used to identify nodes with trustable values but those peers may contain infected files. This case is shown under the highest considered P_I value, $P_I = 0.9$, as the number of infected peers of l) is smaller than those of h).

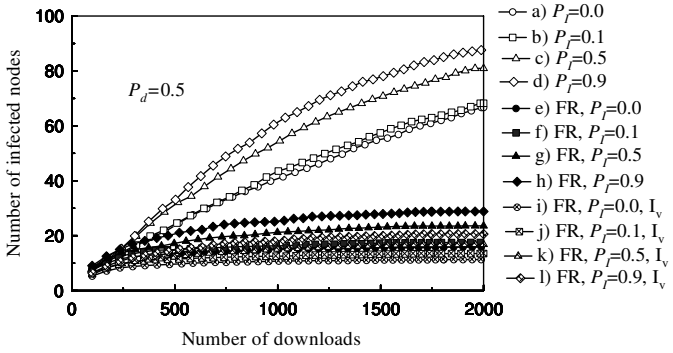

Fig. 2. Comparison of proliferation of viruses using T_v only and with the proposed scheme.

An increase on the number of trust parameters in the management scheme may discourage the download activity between peers. The number of downloads was then evaluated to observe the impact of the proposed scheme. Figure 3 shows the number downloads when using a single parameter, T_v, and our proposed scheme, in downloads per time slot. These results show that the number of trust parameters has no significant impact on the number of downloads.

IV. Conclusions

It was shown that infectious viruses underscore management schemes using single peer trust values combined with a warning notification system. This letter proposed a trust management to bound the proliferation of viruses in P2P networks. The proposed scheme uses infectious peer values and file reputation values in combination with trust values and a warning system. It was shown that the proposed scheme can bound virus proliferation as the number of infected peers is limited.

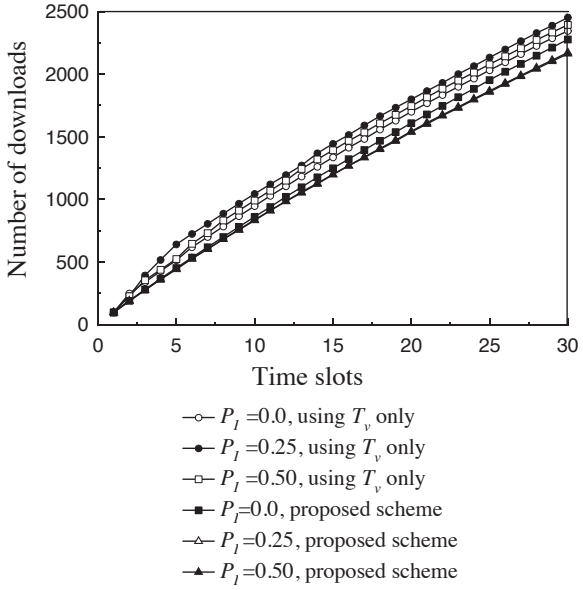

Fig. 3. Comparison of the number of downloads in the P2P network.

References

[1] X. Xu, Y. Wang, S. P. Panwar, and K.W. Ross, "A Peer-to-Peer Video-on-Demand System using Multiple Description Coding and Server Diversity," *Proc. IEEE ICIP*, pp. 1759-1762, October 2004

[2] X. Hei, C. Liang, J. Liang, Y. Liu and K.W. Ross, "A Measurement Study of a Large-Scale P2P IPTV System," *IEEE Trans. on Multimedia*, pp. 1672-1787, December 2007.

[3] M. Macedonian, "Distributed File Sharing: Barbarians at the Gate?" *IEEE Computer*, Vol. 33, Issue 8, pp. 99-101, Aug. 2000.

[4] Y. Wang, X. Yun, Y. Li, "Analyzing the Characteristics of Gnutella Overlays," *Proc. IEEE IV International Conference in Information Technology*, 2007, pp. 1095-1100, 2-4 April, 2007.

[5] X. Zhang, S. D., and H-H. Chen, "Analysis of Virus and Antivirus Spreading Dynamics," Proc. *IEEE Globecom*, Vol. 3, 5 pages, Nov. 28-Dec 2, 2005.

[6] P. Li, Z. Wang, X. Tan, "Characteristic Analysis of Virus Spreading in Ad Hoc Networks," Proc. *IEEE WCIS*, pp. 538-541, December 16-19, 2007.

[7] L-C. Chen and K.M. Carley, "The Impact of Countermeasure Propagation on the Prevalence of Computer Viruses," *IEEE Trans. on System, Man, and Cybernetics*, Vol. 34, issue 2, pp. 823-833, April 2004.

[8] E. Damiani, S. De Capitani Di Vimercati, S. Paraboschi, P. Samarati, and F. Violante, "A Reputation-based Approach for choosing Reliable Resources in Peer-to-Peer Networks," Proc. of the 9th *ACM conference on Computer and communications security (CCS)*, pp. 207-216, Washington, DC, November 2002.

[9] S. Marti and H. Garcia-Molina, "Limited Reputation Sharing in P2P Systems," Proc. of the 5th *ACM EC*, pp. 91-101, New York, NY, May 2004.

[10] J. Shin, T. Kim, and S. Tak, "A Reputation Management Scheme Improving the Trustworthiness of P2P Networks," Proc. *IEEE ICCHIT 2008*, pp. 92-97, August 28-30, 2008.

[11] X. Dong, W. Yu, and Y. Pan, "A Dynamic Trust Management Scheme to Mitigate Malware Proliferation in P2P network," Proc. *IEEE ICC 2008*, 5 pages, Beijing, China, May 2008.

[12] E.K. Lua, J. Crowcroft, M. Pias, R. Sharma, and S. Lim., "A Survey and Comparison of Peer-to-Peer Overlay Network Schemes," *IEEE Comm. Survey and Tutorial*, Vol. 7, Issue 2, 2^{nd} Quarter 2005, pp. 72-93, 2005.