Efficient Immunization Algorithm for Peer-to-Peer Networks[*]

Hao Chen, Hai Jin, Jianhua Sun, and Zongfen Han

Cluster and Grid Computing Lab
Huazhong University of Science and Technology, Wuhan, 430074, China
{haochen,hjin,jhsun,zfhan}@hust.edu.cn

Abstract. In this paper, we present a detail study about the immunization of viruses in Peer-to-Peer networks with power-law degree distributions. By comparing two different immunization strategies, we conclude that it is efficient to immunize the highly connected nodes in order to eradicate viruses from the network. Furthermore, we propose an efficient updating algorithm of global virus database according to the degree-based immunization strategy.

1 Introduction

Recently, a large proportion of research effort has been devoted to the study and modeling of a wide range of natural systems that can be regarded as networks, focusing on large scale statistical properties of networks other than single small networks. Some reviews on complex networks can be found [6]. From biology to social science to computer science, systems such as the Internet [5], the World-Wide-Web [2], social communities and biological networks can be represented as graphs, where nodes represent individuals and links represent interactions among them. Like these complex networks, one important characteristic of P2P networks is that they often show high degree of tolerance against random failures, while they are vulnerable under intentional attacks [3]. Such property has motivated us to carry out a study about the virus spreading phenomenon and some hacker's behaviors in P2P networks from a topological point of view. In our study, we choose Gnutella as our testbed. The main contributions of this paper are: first, an optimal immunization strategy is given; second, we propose an efficient information updating algorithm for P2P networks based on the immunization strategy.

The rest of this paper is organized as follows. Section 2 describes the immunization model of P2P networks. In Section 3, we propose an information updating algorithm for P2P networks. In Section 4, we give our conclusions and point some directions for future work.

2 Immunization Model of P2P Networks

2.1 Modeling Immunization of P2P Network

One widely used model of virus spreading is called SIS (*susceptible-infective-susceptible*) model [4]. This model assumes that the nodes in the network can be in

---

two states: *susceptible* (one node is healthy but could be infected by others), *infective* (one node has the virus, and can spread it to others). Each susceptible node is infected with rate ν if it is connected to one or more infected nodes. At the same time, an infected node is cured with rate δ, defining an effective spreading rate $\lambda = \nu/\delta$ for the virus.

A widely used theoretical model for power-law networks [3] is the Barabasi and Albert (BA) model [1]. In the following, we will use the BA model to deduce a theoretical framework of the prevalence of virus, and then compare with the real data obtained from Gnutella network [3].

In order to take into account the different connectivity of all the nodes, we denote the density of infected nodes with degree k by $\rho_k(t)$, where the parameter t indicates the time, and the average density of all infected nodes in the network by $\rho = \Sigma_k p(k)\rho_k$. According to the results in [7], we have the following equation:

$$\rho \simeq \frac{2e^{-1/m\lambda}}{1 - e^{-1/m\lambda}}. \tag{1}$$

The ρ is the stationary density of all infected nodes after time evolution of the stochastic cycle of SIS model.

## 2.2   Immunization Strategies of P2P Networks

The power-law networks exhibit different behaviors under random failures and intentional attacks [3], from which two intuitive immunization strategies are randomized and degree-based immunizations. In the randomized immunization strategy, a proportion of nodes randomly chosen in the network are immunized. Accordingly, in the degree-based strategy, nodes are chosen for immunization if their degrees are greater than a predefined value.

In the randomized case, for a fixed spreading rate λ, defining the fraction of immunized nodes in the network as f, we can get the effective spreading rate $\lambda(1 - f)$, and substituting it into equation (1) we obtain

$$\rho_f = \frac{2e^{-1/m\lambda(1-f)}}{1 - e^{-1/m\lambda(1-f)}}. \tag{2}$$

Evidently, in the case of degree-based immunization, we can not use equation (2) to deduce an explicit formula as in the randomized case, but we will use simulations to compare the difference between the theoretical BA model and the real data of Gnutella network.

Our simulations are performed with a fixed spreading rate $\lambda = 0.15$, the smallest node degree $m = 3$ and the number of nodes $N = 34206$ the same as the real data of the topology collected from Gnutella network [3]. Initially we infect a proportion of healthy nodes in the network, and iterate the rules of SIS model. In Fig.1 (a), we plot the simulation results of degree-based immunization for BA network (line) and Gnutella network (square-line). With the increasing of f, ρ_f decays much faster in Gnutella network than in BA model, and the linear regression from the largest values of f yields the estimated thresholds $f_c \simeq 0.03$ in Gnutella network, $f_c \simeq 0.2$ in BA network. The value of f_c in Gnutella network indicates that the Gnutella network is
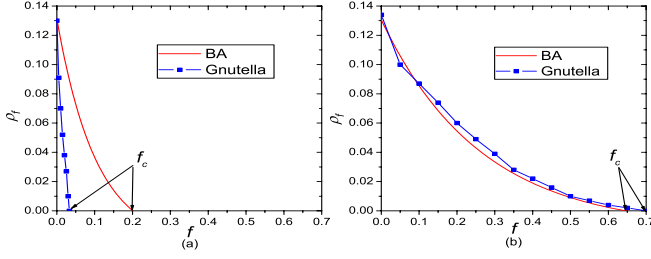
Fig. 1. Results for randomized and degree-based immunization measured by the density of infected nodes ρ_f as a function of the fraction of immunized nodes f.

very sensitive to the degree-based immunization, and the immunization of just a very small fraction (3%) of nodes will eradicate the spreading of virus. On the other hand, in Fig.1 (b), the simulation results of randomized immunization are plotted for Gnutella Network (square-line), which is in good agreement with the theoretical prediction (line) by equation (2), except for a larger value of $f_c \simeq 0.7$ compared with the value $f_c \simeq 0.64$ of BA network.

3   Efficient Immunization Algorithm for P2P Networks

Based on the analysis of immunization strategies, we use high degree nodes to transfer immunization information (when an intrusion or a virus is detected) to other nodes. First, we formulate the highest degree k_{max} in the network as a function of the network size. Given a specific degree distribution p_k, as stated in [6], we have $\frac{dp_k}{dk} \simeq -np_k^2$ [1]. For BA model, the probability distribution of degree is $p_k = 2m^2 k^{-3}$. Substituting it into above equation, we have $k_{max} \simeq \sqrt{2m^2 n/3}$. For simplicity, suppose that the degrees of the nodes in the transferring sequence, through which we update immunization information, are all approximate to k_{max}, then the number of steps needed to transfer the information in the network of size n is $s = n/k_{max} \simeq \sqrt{3n/2m^2}$.

We perform simulations of the real data of Gnutella network with a power-law exponent $\gamma = 2.0$ [3], and compare the simulation results with the theoretical prediction of BA network . The number of nodes range from $N = 10^3$ to $N = 10^4$. Fig.2 shows that the algorithm of transferring update information based on high degrees in Gnutella network is as efficient as the prediction of the theoretical BA model. We need only $s = 11$ steps to update all high degree nodes in Gnutella network with $N = 1000$ nodes, and $s = 36$ steps in Gnutella network with same number of nodes.

4   Conclusions

In this paper, based on the simple SIS model, we analyze the influence of virus spreading on P2P networks with two different immunization strategies, namely randomized and
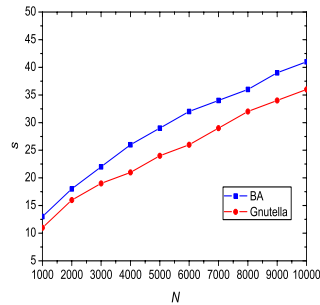
Fig. 2. The number of steps s needed to transfer information through high degree nodes as a function of the network size N.

degree-based immunization, and performe theoretical modeling and real data simulations. The results show that the degree-based strategy is more efficient than the randomized strategy, which also motivate us to design an effective immunization information transferring algorithm.

References

1. A. L. Barabasi and R. Albert, "Emergence of scaling in random networks", *Science*, Vol.286, pp.509, 1999.
2. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, and R. Stata, "Graph structure in the web", *Computer Networks*, Vol.33, pp.309-320, 2000.
3. H. Chen, H. Jin, and J. H. Sun, "Analysis of Large-Scale Topological Properties for Peer-to-Peer Networks", *Proceedings of International Symposium on Cluster Computing and the Grid*, 2004.
4. O. Diekmann and J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, JohnWiley & Sons, New York, 2000.
5. M. Faloutsos, P. Faloutos, and C. Faloutsos, "On Power-law Relationships of the Internet Topology", *Computer Communications Review*, Vol.29, pp.251-262, 1999.
6. M. E. J. Newman, "The structure and function of complex networks", *SIAM Review*, Vol.45, pp.167-256, 2003.
7. R. P. Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks", *Phys. Rev. Lett*, Vol.86, pp.3200-3203, 2001.