Worm Propagation Models

By Zesheng Chen

ABOUT THE AUTHOR. Zesheng Chen is currently a Ph.D. Candidate in the Communication Networks and Machine Learning Group at the School of Electrical and Computer Engineering, Georgia Institute of Technology, advised by Professor Chuanyi Ji. Chen's research interests focus on network security, especially modeling the spread of malware on networks, the performance of distributed detection systems, the effectiveness of defense systems, and the performance evaluation of communication networks.

In this article, we introduce worm propagation models that have been used to describe worm spreading dynamics using different scanning methods. We begin with random-scanning worms that randomly select the targets in IPv4 address space, and then consider propagation models for worms using other scanning strategies.

When the Code Red v2 worm surged in July of 2001, Stuart Staniford presented a simple model to explain the Random Constant Spread (or RCS) theory of the worm. Such a model is essentially identical to a biological epidemic model, which is widely used in epidemiology. For a simple epidemic model, each host has only two states: susceptible and infected. A susceptible host can be infected by other infectious hosts, while an infected host can be recovered and become susceptible. Combining infection and recovery provides one of the simplest models, the *susceptible->infected->susceptible* (SIS) model.

Traditionally, epidemic model describes the SIS model using a nonlinear differential equation to measure the infected-population dynamics:

$$\frac{dn}{dt} = \beta n(1-n) - dn\,,$$

where $n(t)$ is the fraction of infected hosts among all vulnerable hosts, β is the birth rate (the rate at which an infected host infects other susceptible hosts), and d is the death rate (the rate at which an infected host becomes susceptible). The solution to the above equation is

$$n(t) = \frac{n_0(1-\rho)}{n_0 + (1-\rho-n_0)e^{-(\beta-d)t}}\,,$$

where $\rho=d/\beta$ and $n_0=n(t=0)$. When the random-scanning worm propagation is concerned, the birth rate β becomes $sN/2^{32}$, where N is the total number of vulnerable hosts and s is the scanning rate (the number of scans that an infected host sends per unit time).

Another discrete-time and continuous state deterministic approximation model, called Analytical Active Worm Propagation (AAWP) model, has been proposed by Chen et al. to model the spread of active worms that employ random scanning. A nonlinear difference equation is used to model the worm propagation dynamics:

$$n_{i+1} = (1-d)n_i + (N - n_i)[1-(1-\frac{1}{2^{32}})^{sn_i}],$$

where n_i is the expected number of infected host at time step i. Such model considers the time that it takes a worm to infect a host.

The differences between the AAWP model and the epidemic model are:

(1) The epidemic model uses a continuous-time differential equation, while the AAWP model is based on a discrete-time model. We believe that the AAWP model is more accurate. Because in the AAWP model, a computer cannot infect other hosts before it is infected completely. But in the epidemic model, a computer begins devoting itself to infecting other hosts even though only a "small part" of it is infected. Therefore, the speed that the worm can achieve and the number of hosts that can be infected may be very different.

(2) The epidemic model does not consider the time that it takes the worm to infect a host, while the AAWP model does. Different worms have different infection abilities that are reflected by the scanning rate (or the birth rate) and the time spent to infect a host. The time required to infect a host always depends on the size of the worm's copy, the degree of network congestion, the distance between source and destination, and the vulnerability that the worm exploits. It can be shown that the time to infect a host is an important factor for the spread of active worms.

(3) In the AAWP model, we consider the case that the worm can infect the same destination at the same time, while the epidemic model ignores the case. In fact, it is not uncommon for a vulnerable host to be hit by two (or more) scans at the same time.

Both models, however, are deterministic and try to get the expected number of infected hosts, given the size of the initially-infected hosts, the total number of vulnerable hosts, the scanning rate/birth rate, and the death rate. The epidemic model can easily deduce the closed form, while the AAWP model predicts the spread of random-scanning worms more accurately.

To account for the stochastic property of worm propagation, Rohloff et al. introduced a stochastic density-dependent Markov jump process propagation model for a random-scanning worm drawn from the field of epidemiology. Their analysis indicates that, excluding the early and late growth stages of an epidemic and the "time-shifting" effects, simulations of a worm's propagation using the deterministic and stochastic models are effectively equivalent when the number of vulnerable hosts is sufficiently large. Therefore, if these effects can be ignored, the possible variability of random-scanning-worm epidemics is surprisingly minor. This finding suggests that the deterministic models can be applied to study the performance of worm detection and defense systems.

Among these applications, Zou et al. use an epidemic model to estimate the propagation speed of worm epidemics at the early stage through observing traffic arriving at the unused IP addresses. Since a random-scanning worm randomly selects target IP

addresses, some worm scans may hit the address space where no hosts exist. If we monitor on these unused IP addresses, we can detect scans from worms that employ random scanning. These monitored unused IP addresses are called “network telescope” or “Darknet”. Some “background noise”, however, may also reach Darknet, such as hostile reconnaissance scans, old worm scans, responses of victims of Denial of Service (DoS) attack, and visits of mis-configuration hosts. Thus, we need to distinguish worm scans traffic from other sources. Zou et al. designed a Kalman filter to estimate the infection speed of worms at the early stage. Such filter detects the trend of traffic, rather than the burst, and thus is robust to background noise.

Other scanning methods, such as localized scanning, importance scanning, and sequential scanning, have been studied by extending the epidemic model and the AAWP model. The key observation is that vulnerable hosts are not uniformly distributed in the Internet, and thus the models are based on subnets. That is, the deterministic models are applied in each subnet. Different subnets are related by counting the average number of scans falling into each subnet. It turns out that uneven distribution of vulnerable hosts fosters the spread of worms using these scanning methods, comparing with random scanning.

Finally, what about the worm propagation model for topological scanning? Such a scanning method is quite different from other scanning methods. Topological-scanning worms rely on the information contained in the victim host in order to locate new targets. The information may include routing tables, email addresses, a list of peers, and Uniform Resource Locations (URLs). Thus, topological-scanning worms spread analogously to biological viruses. Boguna et al. extended epidemic models to model topological scanning dynamics, taking into consideration the nodal degree distribution of the underlying topology. Ganesh et al. applied contact process to analyze the ease of topological-scanning worm propagation on different topologies. Garetto et al. analyzed e-mail spreading in small-world topologies using a variation of the influence model, where the influence of neighbors is constrained to take a multilinear form. Chen et al. used a spatial-temporal random process to describe the statistical dependence of worm propagation in arbitrary topologies.

In summary, modeling the spread of worms that employ different scanning methods is an active research area. Mathematical tools play an important role for understanding how worms spread and how we can detect and defend against them.

Reference

[1] M. Boguna, R. Pastor-Satorras, and A. Vespignani, “Epidemic Spreading in Complex Networks with Degree Correlations,” in *Statistical Mechanics of Complex Networks, Edited by R. Pastor-Satorras, M. Rubi and A. Diaz-Guilera*, Lecture Notes in Physics, vol. 625, p.127-147, 2003.

[2] Z. Chen, L. Gao, and K. Kwiat, “Modeling the Spread of Active Worms,” in *Proc. of INFOCOM 2003*, San Francisco, April, 2003.

[3] Z. Chen and C. Ji, “Importance-Scanning Worm Using Vulnerable-Host Distribution,” in *Proc. of IEEE Globecom 2005*, St. Louis, MO, 2005,

[4] Z. Chen and C. Ji, “Spatial-Temporal Modeling of Malware Propagation in Networks,” in *IEEE Transactions on Neural Networks: Special Issue on Adaptive Learning Systems in Communication Networks*, vol. 16, no. 5, Sept. 2005 .

[5] A. Ganesh, L. Massoulie, and D. Towsley, “The Effect of Network Topology on the Spread of Epidemics,” in *Proc. of INFOCOM 2005*, Miami, March 2005.

[6] M. Garetto, W. Gong, D. Towsley, “Modeling Malware Spreading Dynamics,” in *Proc. of INFOCOM 2003*, San Francisco, April, 2003.

[7] M. A. Rajab, F. Monrose, and A. Terzis, “On the Effectiveness of Distributed Worm Monitoring,” in *Usenix Security 2005*.

[8] K. Rohloff and T. Basar, “Stochastic Behavior of Random Constant Scanning Worms,” in *Proc. of IEEE Conference on Computer Communications and Networks 2005 (ICCCN 2005)*, San Diego, CA, Oct., 2005.

[9] S. Staniford, V. Paxson, and N. Weaver, “How to 0wn the Internet in Your Spare Time,” in *Proc. of the 11th USENIX Security Symposium (Security '02)*, 2002.

[10] C. C. Zou, W. Gong, D. Towsley, and L. Gao. “The Monitoring and Early Detection of Internet Worms,” *IEEE/ACM Transactions on Networking*, 13(5), 961- 974, October 2005.

[11] C. C. Zou, D. Towsley, and W. Gong. “On the Performance of Internet Worm Scanning Strategies,” to appear in Journal of Performance Evaluation.